Introducción al Análisis de Regresión Lineal

Montgomery • Peck • Vining

3a. edición

CECSA

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

Introducción al Análisis de Regresión Lineal

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

INTRODUCCIÓN AL ANÁLISIS DE REGRESIÓN LINEAL

TERCERA EDICIÓN EN INGLÉS
(PRIMERA EDICIÓN EN ESPAÑOL)

DOUGLAS C. MONTGOMERY
Arizona State University

ELIZABETH A. PECK
The Coca Cola Company

G. GEOFFREY VINING
Virginia Tech

TERCERA REIMPRESIÓN
MÉXICO, 2006

COMPAÑÍA EDITORIAL CONTINENTAL

http://carlos2524.jimdo.com/

Para establecer comunicación con nosotros puede hacerlo por:

correo:
Renacimiento 180, Col. San Juan Tlihuaca, Azcapotzalco, 02400, México, D.F.

fax pedidos:
(01 55) 5561 4063 • 5561 5231

e-mail:
info@patriacultural.com.mx

home page:
www.patriacultural.com.mx

Título original de la obra:
Introduction to linear regression analysis / Douglas C. Montgomery
Copyright @ 2001 John Wiley & Sons, Inc. All rights reserved
ISBN 0-471-31565-6

Dirección editorial: Javier Enrique Callejas
Coordinadora de edición: Elisa Pecina Rosas
Diseño de interiores: Guillermo Rodríguez Luna
Diseño de portada: Enrique Román Esnaurrízar

Traducción:
Ing. Virgilio González Pozo
Perito traductor

Revisión técnica:
Ph. D. César Acosta Mejía
Profesor titular
Departamento de Ingeniería Industrial, ITAM

Introducción al análisis de regresión lineal
Derechos reservados respecto a la primera edición en español:
© 2002, Douglas C. Montgomery / Elizabeth A. Peck / G. Geoffrey Vining
© 2002, GRUPO PATRIA CULTURAL, S.A. DE C.V.
bajo el sello de Compañía Editorial Continental
Renacimiento 180, Colonia San Juan Tlihuaca,
Delegación Azcapotzalco, Código Postal 02400, México, D.F.

Miembro de la Cámara Nacional de la Industria Editorial
Registro núm. 43

ISBN 970-24-0327-8

Queda prohibida la reproducción o transmisión total o parcial del contenido de la presente obra en cualesquiera formas, sean electrónicas o mecánicas, sin el consentimiento previo y por escrito del editor.

Impreso en México
Printed in Mexico

Primera edición: 2002
Segunda reimpresión: 2005
Tercera reimpresión: 2006

http://carlos2524.jimdo.com/

CONTENIDO

Prefacio xiii

1. Introducción 1

1.1 Regresión y formación de modelos 1
1.2 Recolección de datos 6
1.3 Usos de la regresión 9
1.4 El papel de la computadora 10

2. Regresión lineal simple 13

2.1 Modelo de regresión lineal simple 13
2.2 Estimación de los parámetros por mínimos cuadrados 13
2.2.1 Estimación de β_0 y β_1 14
2.2.2 Propiedades de los estimadores por mínimos cuadrados y el modelo ajustado de regresión 19
2.2.3 Estimación de σ^2 21
2.2.4 Una forma alterna del modelo 23
2.3 Prueba de hipótesis de la pendiente y de la ordenada al origen 23
2.3.1 Uso de pruebas t 24
2.3.2 Prueba del significado de la regresión 25
2.3.3 Análisis de varianza 27
2.4 Estimación de intervalo en la regresión lineal simple 30
2.4.1 Intervalos de confianza de β_0, β_1 y σ^2 30
2.4.2 Estimación de intervalos de la respuesta media 32
2.5 Predicción de nuevas observaciones 35
2.6 Coeficiente de determinación 37
2.7 Algunas consideraciones sobre el uso de la regresión 38
2.8 Regresión por el origen 41
2.9 Estimación por máxima verosimilitud 46
2.10 Caso en el que el regresor x es aleatorio 48
2.10.1 Distribución conjunta de x y y 48
2.10.2 Distribución normal conjunta de x y y: el modelo de correlación 49
Problemas 54

http://carlos2524.jimdo.com/

3. Regresión lineal múltiple **61**

3.1 Modelo de regresión múltiple 61
3.2 Estimación de los modelos del parámetro 65
3.2.1 Estimación de los coeficientes de regresión por mínimos cuadrados 65
3.2.2 Interpretación geométrica de mínimos cuadrados 72
3.2.3 Propiedades de los estimadores de mínimos cuadrados 74
3.2.4 Estimación de σ^2 74
3.2.5 Inadecuación de los diagramas de dispersión en regresión múltiple 76
3.2.6 Estimación de máxima posibilidad 77
3.3 Prueba de hipótesis en regresión lineal múltiple 78
3.3.1 Pruebas de la significancia de la regresión 79
3.3.2 Pruebas sobre coeficientes individuales de regresión 82
3.3.3 Caso especial de columnas ortogonales en **X** 87
3.3.4 Prueba de la hipótesis general lineal 89
3.4 Intervalos de confianza en regresión múltiple 92
3.4.1 Intervalos simultáneos de confianza para coeficientes de regresión 92
3.4.2 Estimación del intervalo de confianza de la respuesta media 93
3.4.3 Intervalos de confianza simultáneos de coeficientes de regresión 94
3.5 Predicción de nuevas observaciones 98
3.6 Extrapolación oculta en la regresión múltiple 99
3.7 Coeficientes normalizados de regresión 102
3.8 Multicolinealidad 106
3.9 ¿Por qué los coeficientes de regresión tienen signo incorrecto? 108
Problemas 111

4. Comprobación de la adecuación del modelo **117**

4.1 Introducción 117
4.2 Análisis de residuales 117
4.2.1 Definición de residuales 117
4.2.2 Métodos para escalar residuales 118
4.2.3 Gráficas de residuales 122
4.2.4 Gráficas de regresión parcial y de residuales parciales 131
4.2.5 Otras gráficas de residuales y métodos de análisis 133
4.3 La estadística PRESS 136
4.4 Detección y tratamiento de puntos atípicos 137
4.5 Falta de ajuste del modelo de regresión 141
4.5.1 Una prueba formal de falta de ajuste 141
4.5.2 Estimación del error puro a partir de vecinos próximos 145
Problemas 148

http://carlos2524.jimdo.com/

5. Transformaciones y ponderación para corregir inadecuaciones del modelo 155

5.1 Introducción 155
5.2 Transformaciones estabilizadoras de la varianza 156
5.3 Transformaciones para linealizar el modelo 159
5.4 Métodos analíticos para seleccionar una transformación 166
5.4.1 Transformaciones de *y*: el método de Box-Cox 166
5.4.2 Transformaciones de las variables regresoras 170
5.5 Mínimos cuadrados generalizados y ponderados 172
5.5.1 Mínimos cuadrados generalizados 173
5.5.2 Mínimos cuadrados ponderados 174
5.5.3 Algunos asuntos prácticos 175
Problemas 179

6. Diagnóstico para balanceo e influencia 185

6.1 Importancia de detectar observaciones influyentes 185
6.2 Balanceo 186
6.3 Medidas de influencia: La *D* de Cook 188
6.4 Medidas de influencia: *DFFITS* y *DFBETAS* 191
6.5 Una medida de eficiencia del modelo 193
6.6 Detección de grupos de observaciones influyentes 194
6.7 Tratamiento de observaciones influyentes 195
Problemas 195

7. Modelos polinomiales de regresión 197

7.1 Introducción 197
7.2 Modelos polinomiales en una variable 197
7.2.1 Principios básicos 197
7.2.2 Ajuste polinomial por segmentos (*SPLINE*) 203
7.2.3 Términos polinomiales y trigonométricos 211
7.3 Regresión no paramétrica 211
7.3.1 Regresión Kernel 212
7.3.2 Regresión ponderada localmente (Loess) 213
7.3.3 Precauciones finales 217
7.4 Modelos polinomiales con dos o más variables 217
7.5 Polinomios ortogonales 226
Problemas 230

8. Variables indicadoras 237

8.1 El concepto general de las variables indicadoras 237

http://carlos2524.jimdo.com/

8.2 Comentarios sobre el uso de variables indicadoras 249
8.2.1 Variables indicadoras en función de la regresión con códigos asignados 249
8.2.2 Variables indicadoras como sustitutos de un regresor cuantitativo 250
8.3 Método de regresión para análisis de varianza 251
Problemas 256

9. Selección de variable y formación el modelo 261

9.1 Introducción 261
9.1.1 El problema de la construcción del modelo 261
9.1.2 Consecuencias de la mala especificación del modelo 262
9.1.3 Criterios para evaluar modelos de regresión con subconjuntos de variables 265
9.2 Técnicas computacionales para seleccionar variables 271
9.2.1 Todas las regresiones posibles 271
9.2.2 Métodos de regresión por segmentos 277
9.3 Algunas recomendaciones finales para la práctica 284
Problemas 285

10. Multicclinealidad 291

10.1 Introducción 291
10.2 Fuentes de multicolinealidad 291
10.3 Efectos de la multicolinealidad 294
10.4 Diagnóstico de multicolinealidad 299
10.4.1 Examen de la matriz de correlación 299
10.4.2 Factores de inflación de varianza 302
10.4.3 Análisis del eigensistema de $\mathbf{X}'\mathbf{X}$ 303
10.4.4 Otros diagnósticos 307
10.5 Métodos de manejar la multicolinealidad 308
10.5.1 Recolección de datos adicionales 309
10.5.2 Reespecificación del modelo 309
10.5.3 Regresión Ridge 311
10.5.4 Otros métodos 327
10.5.5 Comparación y evaluación de estimadores sesgados 337
Problemas 339

11. Regresión robusta 343

11.1 La necesidad de una regresión robusta 343
11.2 Estimadores *M* 346
11.3 Propiedades de los estimadores robustos 359

http://carlos2524.jimdo.com/

11.3.1 Punto de quiebre 359
11.3.2 Eficiencia 359
11.4 Reseña de otros estimadores robustos de regresión 360
11.4.1 Estimadores de punto de quiebre alto 360
11.4.2 Estimadores de influencia limitada 364
11.4.3 Otros procedimientos 365
11.4.4 Cálculo de estimadores de regresión robusta 367
Problemas 368

12. Introducción a la regresión no lineal 373

12.1 Modelos de regresión lineales y no lineales 373
12.1.1 Modelos de regresión lineales 373
12.1.2 Modelos de regresión no lineales 374
12.2 Mínimos cuadrados no lineales 376
12.3 Transformación a un modelo lineal 378
12.4 Estimación de parámetros en un sistema no lineal 382
12.4.1 Linealización 382
12.4.2 Otros métodos de estimación de parámetros 389
12.4.3 Valores iniciales 390
12.4.4 Programas de cómputo 391
12.5 Inferencia estadística en regresión no lineal 392
12.6 Ejemplos de modelos de regresión no lineales 394
Problemas 395

13. Modelos lineales generalizados 399

13.1 Introducción 399
13.2 Modelos de regresión logística 399
13.2.1 Modelos con una variable de respuesta binaria 399
13.2.2 Estimación de parámetros en un modelo de regresión logística 403
13.2.3 Interpretación de parámetros en un modelo de regresión logística 405
13.2.4 Pruebas de hipótesis sobre los parámetros del modelo 408
13.3 Regresión de Poisson 414
13.4 El modelo lineal generalizado 420
13.4.1 Funciones de cadena y predictores lineales 421
13.4.2 Estimación de parámetros e inferencia en el modelo lineal general 422
13.4.3 Predicción y estimación con el modelo lineal general 426
13.4.4 Análisis de residuales en el modelo lineal generalizado 428

http://carlos2524.jimdo.com/

13.4.5 Sobredispersión 429
Problemas 431

14. Otros temas acerca del uso del análisis de regresión 441

14.1 Modelos de regresión con errores autocorrelacionados 441
14.1.1 Causas y efectos de la autocorrelación 441
14.1.2 Detección de la presencia de autocorrelación 442
14.1.3 Métodos de estimación de parámetros 446
14.2 Efecto de los errores de medición sobre los regresores 452
14.2.1 Regresión lineal simple 452
14.2.2 El modelo de Berkson 454
14.3 Estimación inversa – El problema de la calibración 454
14.4 Bootstrap en regresión 459
14.4.1 Muestreo bootstrap en regresión 460
14.4.2 Intervalos de confianza de bootstrap 461
14.5 Árboles de clasificación y regresión (CART) 466
14.6 Redes neurales 468
14.7 Experimentos diseñados para regresión 470
Problemas 473

15. Validación de los modelos de regresión 477

15.1 Introducción 477
15.2 Técnicas de validación 478
15.2.1 Análisis de los coeficientes y valores predichos del modelo 478
15.2.2 Recolección de nuevos datos – Corridas de confirmación 479
15.2.3 División de los datos 482
15.3 Datos obtenidos con experimentos planeados 491
Problemas 492

APÉNDICE A. Tablas estadísticas 495

APÉNDICE B. Conjuntos de datos para los ejercicios 513

APÉNDICE C. Material suplementario 529

C.1 Resumen de pruebas básicas en estadística prueba 529
C.2 Resumen de la teoría de modelos lineales 531
C.3 Resultados importantes acerca de SS_R y SS_{Res} 535
C.4 El teorema de Gauss-Markov, Var $(\varepsilon) = \sigma^2 \mathbf{I}$ 540
C.5 Aspectos computacionales de la regresión múltiple 541
C.6 Un resultado acerca de la inversa de una matriz 543

http://carlos2524.jimdo.com/

C.7 Desarrollo del estadístico 544
C.8 Desarrollo de $S_{(i)}^2$ 546
C.9 Una prueba para valores atípicos basada en R de Student 547
C.10 El teorema de Gauss-Markov, Var $(\varepsilon) = \mathbf{V}$ 550
C.11 El sesgo en MS_{Res} cuando el modelo está subespecificado 552
C.12 Cálculo de diagnóstico de influencia 553
C.13 Modelos lineales generalizados 555

Referencias **567**

Índice

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

PREFACIO

El análisis de regresión es una de las técnicas de uso más frecuente para analizar datos multifactoriales. Su atractivo y utilidad general son el resultado del proceso conceptualmente lógico de usar una ecuación para expresar la relación entre una variable de interés (la respuesta) y un conjunto de variables predictoras relacionadas. El análisis de regresión tiene también interés teórico, por las elegantes matemáticas que se usan y por una teoría estadística bien desarrollada. Para usar bien la regresión se requiere apreciar tanto la teoría como los problemas prácticos que se suelen presentar cuando se emplea esta técnica con datos del mundo real.

Pretendemos que este texto sea el de un curso básico de análisis de regresión, ya que contiene los temas acostumbrados en estos cursos, así como muchos de los más nuevos. Combina la teoría y la práctica, por lo que el lector conocerá los principios básicos para aplicar las técnicas de formación de modelos de regresión, en una gran variedad de campos de aplicación. Este libro comenzó como una evolución de notas para un curso de análisis de regresión, dictado para estudiantes de licenciatura de último año, y graduados de primer año, en diversos campos de ingeniería, ciencias químicas y físicas, estadística, matemáticas y administración. También se usó el material de muchos seminarios y cursos cortos en la industria, para auditorios profesionales. Suponemos que el lector ha tomado un primer curso de análisis y que está familiarizado con pruebas de hipótesis e intervalos de confianza, así como con las distribuciones normal, t, x^2 y F, además es necesario que tenga algunos conocimientos de álgebra de matrices.

La computadora desempeña un papel importante en la aplicación moderna de la regresión. Hoy, hasta los programas de hoja de cálculo tienen la capacidad de ajustar ecuaciones de regresión por mínimos cuadrados. En consecuencia hemos integrado muchos aspectos del uso de computadoras en el texto, incluyendo pantallas de resultados (tanto tabulares como gráficas) y descripciones generales de algunos paquetes de programas. Hemos usado SAS, S-PLUS, MINITAB y STATGRAPHICS en diversos problemas y ejemplos en el texto. Muchos de los problemas de tarea requieren programas de cómputo para su solución. Todos los conjuntos de datos del libro se pueden conseguir en forma electrónica, tanto en la guía de los profesores como en la página de Internet de este libro.

CAMBIOS EN LA TERCERA EDICIÓN

Nuestro objetivo fue hacer más útil el libro, por lo que hicimos cambios en esta edición, tanto para texto como para referencia, y actualizamos nuestra exposición de ciertos temas. La segunda edición tenía 10 capítulos; ésta tiene 15. Esto refleja nuestra estrategia de acortar los capítulos que se enfocan con más claridad sobre temas específicos y con ello se mejora mucho el flujo de material, desde el punto de vista docente, además que facilita la localización de la información sobre un tema específico.

http://carlos2524.jimdo.com/

El capítulo 1 es una introducción general al modelado de regresión, y describe algunas aplicaciones típicas de la regresión. Los capítulos 2 y 3 presentan los resultados generales del ajuste por mínimos cuadrados, en la regresión simple y múltiple, junto con los procedimientos básicos de inferencia (pruebas de hipótesis, intervalos de confianza y de predicción). En el capítulo 4 se describen algunos aspectos preliminares de comprobación de adecuación de modelo, incluyendo el análisis de residuales y un gran énfasis sobre gráficas de residuales, detección y manejo de datos atípicos, el estadístico PRESS y la prueba de falta de ajuste. El capítulo 5 describe cómo se pueden usar las transformaciones y los cuadrados mínimos ponderados para resolver problemas de inadecuación de modelos, o para manejar violaciones de las hipótesis básicas de la regresión. Se presentan las técnicas de Box-Cox y de Box-Tidwell, para especificar analíticamente la forma de una transformación. Los diagnósticos de influencia se presentan en el capítulo 6, junto con una descripción preliminar de cómo manejar observaciones influyentes. Los modelos polinomiales de regresión, y sus variaciones, se describen en el capítulo 7. Entre los temas está el de los procedimientos básicos de ajuste e inferencia de polinomiales, y una discusión del centrado de polinomios, la jerarquía, los polinomios por segmentos, modelos con términos tanto polinomiales como trigonométricos, polinomios ortogonales, una revisión de superficies de respuesta y una introducción a las técnicas de regresión no paramétrica y de suavizamiento. El capítulo 8 presenta las variables indicadoras y también establece la relación entre los modelos de regresión y los de análisis de varianza. La selección de variable y las técnicas de construcción de modelos se explican en el capítulo 9, incluyendo los procedimientos por pasos y todas las regresiones posibles; también se describen y se ilustran varios criterios para evaluar modelos de regresión de subconjuntos.

Los nueve primeros capítulos son el núcleo del libro. Muchos de los conceptos y ejemplos se desarrollan entre éstos. Los seis capítulos restantes abarcan una variedad de temas que son importantes para el practicante de la regresión, y se pueden leer en forma independiente. El capítulo 10 se concentra en el problema de la multicolinealidad. Se describen sus causas, sus efectos perjudiciales, su diagnóstico y diversas medidas correctivas. Se presenta la estimación sesgada, incluyendo la regresión ridge y algunas de sus variantes, así como la regresión de componentes principales. El capítulo 11 presenta varias técnicas de regresión robusta, que incluyen los estimadores M, la mínima mediana de cuadrados, la mínima suma recortada de cuadrados y la estimación S. Igualmente se presentan los conceptos de punto de quiebre y eficiencia de un estimador. El capítulo 12 introduce la regresión no lineal, y el capítulo 13 es una presentación básica de los modelos lineales generalizados. Si bien todos éstos sean quizá temas no acostumbrados en un libro de texto sobre regresión lineal, son tan importantes para los alumnos y los profesionistas de ingeniería y ciencias, que hubiéramos faltado gravemente si no presentáramos una introducción a ellos. El capítulo 14 abarca varios temas en forma de perspectiva, incluyendo los modelos de regresión con errores autocorrelacionados, el efecto de los errores de medición sobre los regresores, el problema de estimación inversa o calibración, estimados bootstrap de regresión, árboles de clasificación y regresión, redes neurales y diseño de experimentos para regresión. En el capítulo 15 se describe la validación del modelo de regresión.

En apoyo al material de este libro, hemos incluido en el apéndice C presentaciones breves de algunos temas adicionales, de naturaleza más técnica o teórica. Algunos de esos temas serán interesantes para los especialistas en regresión, o para los profesores que enseñen cursos superiores con este libro.

http://carlos2524.jimdo.com/

USO DEL LIBRO COMO TEXTO

Debido al amplio alcance de los temas, este libro tiene una gran flexibilidad cuando se usa como texto. Para un primer curso en regresión, recomendaríamos estudiar los capítulos 1 al 9, con detalle, para después seleccionar temas que sean de interés específico para la audiencia. Por ejemplo, uno de los autores (Douglas C. Montgomery) enseña con regularidad un curso de regresión a alumnos de ingeniería. Los temas para esa audiencia incluyen la regresión no lineal (porque los modelos mecanísticos que casi siempre son no lineales se presentan con frecuencia en ingeniería), una descripción de las redes neurales y la validación del modelo de regresión. Otros temas que recomendaríamos tener en cuenta son la multicolinealidad (ya que el problema se presenta tan a menudo) y una introducción a la regresión robusta. G. Geoffrey Vining ha dictado un curso de regresión para alumnos de posgrado en estadística, donde se usa mucho el material del apéndice C.

Creemos que se debe integrar directamente el uso de la computadora al curso. En años recientes hemos llevado a nuestras clases una computadora de cuaderno de notas y una pantalla de cristal líquido, e ilustramos con ellos las técnicas a medida que se presentan en clase. Hemos encontrado que así se facilita mucho la comprensión y el aprecio de las técnicas por parte del alumno. También pedimos a éstos que usen programas de regresión para resolver los problemas de tarea. En la mayor parte de los casos, se usan datos reales, o basados en ambientes del mundo real que representan aplicaciones características de la regresión.

Hay un manual de profesores que contiene las soluciones de todos los ejercicios; también hay versiones electrónicas de todos los conjuntos de datos, así como preguntas y problemas que se podrían adecuar para usarlos en los exámenes, además se ofrece una guía de soluciones para el alumno, que muestra las soluciones completas para algunos problemas.

RECONOCIMIENTOS

Agradecemos a todas las personas que nos retroalimentaron y ayudaron para preparar este libro. El doctor Ronald G. Askin, doctor Mary Sue Younger, doctor Russel G. Heikes, doctor John A. Cornell, doctor André I. Khuri, doctor George doctor Runger, doctor Marie Gaudard, doctor James W. Wisnowski y doctor James R. Simpson, que formularon muchas sugerencias con las que mejoraron tanto las ediciones anteriores como esta tercera edición. En particular agradecemos a los muchos alumnos de posgrado y a los practicantes profesionistas quienes dieron retroalimentación, con frecuencia en forma de preguntas muy agudas, que condujeron a la reformulación o a la ampliación del material en el libro. También estamos en deuda con John Wiley & Sons, con la American Statistical Association y con Biometrika Trustees, por el permiso de usar su material con derecho de autor.

DOUGLAS C. MONTGOMERY
ELIZABETH A. PECK
G. GEOFFREY VINING

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

CAPÍTULO 1

INTRODUCCIÓN

1.1 REGRESIÓN Y FORMACIÓN DE MODELOS

El análisis de regresión es una **técnica estadística** para investigar y **modelar la relación entre variables**. Son numerosas las aplicaciones de la regresión, y las hay en casi cualquier campo, incluyendo en ingeniería, ciencias físicas y químicas, economía, administración, ciencias biológicas y de la vida y en las ciencias sociales. De hecho, puede ser que el análisis de regresión sea la técnica estadística más usada.

Como ejemplo de un problema en el que puede ser de utilidad el análisis de regresión, supongamos que un ingeniero industrial, empleado por un embotellador de refrescos, analiza las operaciones de entrega y servicio de producto en máquinas tragamonedas. Cree que el tiempo utilizado por un repartidor, en cargar y dar servicio a una máquina, se relaciona con la cantidad de cajas de producto entregadas. El ingeniero visita 25 tiendas de menudeo, escogidas al azar, con máquinas tragamonedas, y anota el tiempo de entrega en la tienda (en minutos) y el volumen del producto entregado (en cajas), para cada una. Las 25 observaciones se grafican en la figura 1.1a. Esta gráfica se llama **diagrama de dispersión**, y en este caso parece indicar con claridad que hay una relación entre el tiempo y el volumen de entrega. De hecho, da la impresión que los datos caen, en general, pero no exactamente, en una línea recta. La figura 1.1b ilustra esta relación de línea recta.

Si y representa el tiempo de entrega y x representa el volumen entregado, la ecuación de una recta que relaciona esas dos variables es

$$y = \beta_0 + \beta_1 x \tag{1.1}$$

donde β_0 es la ordenada al origen y β_1 es la pendiente. Ahora bien, los datos no caen exactamente sobre una recta, por lo que se debe modificar la ecuación (1.1) para tomar en cuenta esto. Sea la diferencia entre el valor observado de y y el de la línea recta $(\beta_0 + \beta_1 x)$ un **error** ε. Conviene imaginar que ε es un error estadístico, esto es, que es una variable aleatoria que explica por qué el modelo no ajusta exactamente los datos. Este error puede estar formado por los efectos de otras variables sobre el tiempo de entrega, por errores de medición, etc. Así, un modelo más plausible para los datos del tiempo de entrega es

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1.2}$$

La ecuación (1.2) se llama **modelo de regresión lineal**. Por costumbre se dice que x es la variable independiente y y la variable dependiente. Sin embargo, eso causa confusión con el concepto de la independencia estadística, así que aquí llamaremos a x la variable **predictora** o **regresora** y y la variable de **respuesta**. Como la ecuación (1.2) sólo tiene una variable regresora, se llama **modelo de regresión lineal simple**.

http://carlos2524.jimdo.com/

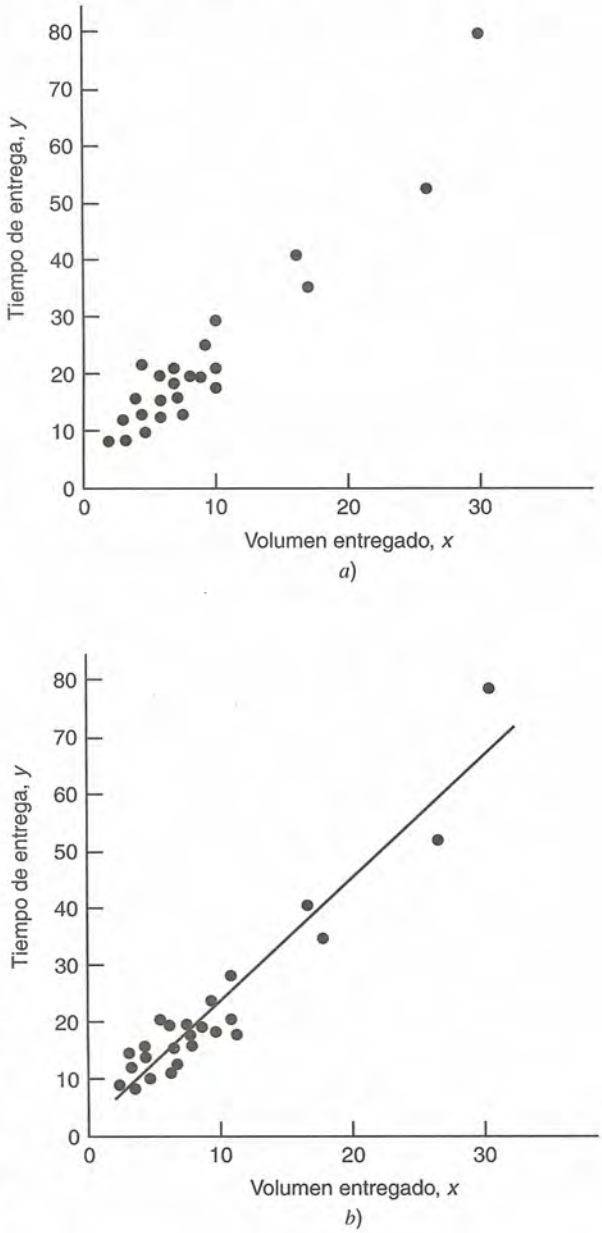


Figura 1.1
a) Diagrama de dispersión para el volumen entregado. *b*) Relación rectilínea entre el tiempo de entrega y el volumen entregado.

Para comprender mejor el modelo de regresión lineal, supongamos que se puede fijar el valor de la variable regresora x para observar el valor correspondiente de la respuesta y. Ahora, si x está fija, el componente aleatorio ε del lado derecho de la ecuación (1.2) determina las propiedades de y. Supongamos que el promedio y la varianza de ε son 0 y σ^2, respectivamente. Entonces, la respuesta media en cualquier valor de la variable regresora es

$$E(y|x) = \mu_{y|x} = E(\beta_0 + \beta x + \varepsilon) = \beta_0 + \beta_1 x$$

http://carlos2524.jimdo.com/

Obsérvese que es la misma relación que se escribió al principio, al examinar el diagrama de dispersión de la figura 1.1a. La varianza de y para cualquier valor dado de x es

$$\text{Var}(y|x) = \sigma_{y|x}^2 = \text{Var}(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2$$

Así, el verdadero modelo de regresión $\mu_{y|x} = \beta_0 + \beta_1 x$ es una línea recta de valores promedios, esto es, la altura de la línea de regresión en cualquier valor de x no es más que el valor esperado de y para esa x. Se puede interpretar que la pendiente β_1 es el cambio de la media de y para un cambio unitario de x. Además, la variabilidad de y en algún valor particular de x queda determinada por la varianza del componente de error en el modelo, σ^2. Esto implica que hay una distribución de valores de y en cada x, y que la varianza de esta distribución es igual en cada x.

Por ejemplo, supongamos que el modelo verdadero de regresión, que relaciona al tiempo de entrega con el volumen entregado es $\mu_{y|x} = 3.5 + 2x$, y que la varianza es $\sigma^2 = 2$. La figura 1.2 ilustra este caso. Obsérvese que se ha usado una distribución normal para describir la variación aleatoria de ε. Como y es la suma de la constante $\beta_0 + \beta_1 x$ (la media) más una variable aleatoria con distribución normal, y será una variable aleatoria con distribución normal. Por ejemplo, si $x = 10$ cajas, el tiempo de entrega y tiene una distribución normal con promedio $3.5 + 2(10) = 23.5$ minutos, y varianza 2. La varianza σ^2 determina la cantidad de variabilidad, o ruido en las observaciones y del tiempo de entrega. Cuando σ^2 es pequeña, los valores observados del tiempo de entrega serán cercanos a la recta, y cuando σ^2 es grande, se pueden desviar bastante de la línea.

En casi todas las aplicaciones de regresión, la ecuación de regresión sólo es una aproximación a la verdadera relación funcional entre las variables de interés. Esas relaciones funcionales se basan, con frecuencia, en una teoría física, química o de otra disciplina científica o técnica; esto es, en el conocimiento del mecanismo básico. En consecuencia, a esta clase de modelos se le llama con frecuencia **modelos mecanísticos**; a su vez, los modelos de regresión se suponen **modelos empíricos**. La figura 1.3 muestra un caso en el que la relación verdadera entre y y x es relativamente complicada, pero se puede aproximar bastante bien con una ecuación de regresión lineal. A veces, el mecanismo básico es más

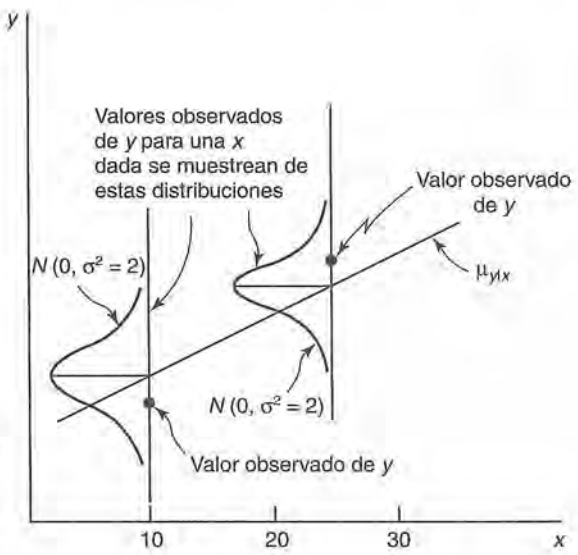


Figura 1.2 Cómo se generan las observaciones en la regresión lineal.

http://carlos2524.jimdo.com/

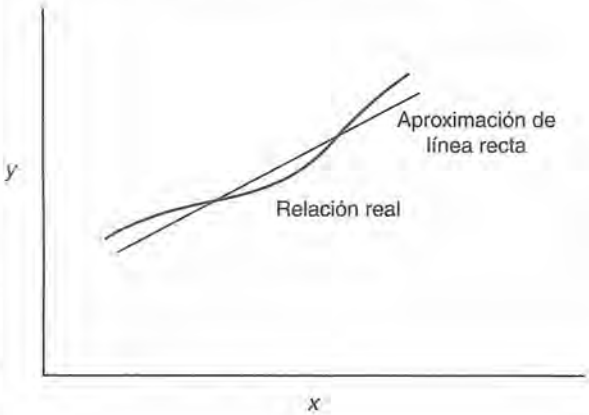


Figura 1.3 Aproximación a una relación compleja por regresión lineal.

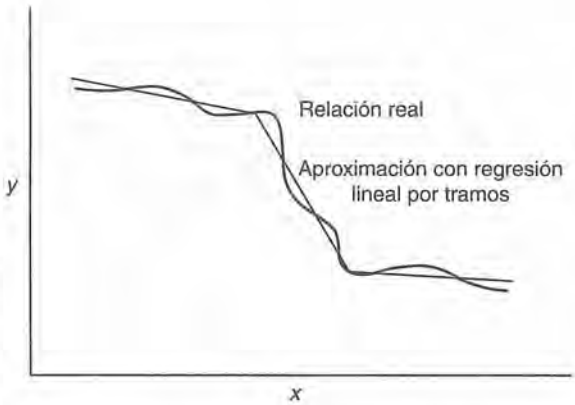


Figura 1.4 Aproximación lineal por tramos de una relación compleja.

complejo, y en consecuencia se necesita una función de aproximación más complicada, como en la figura 1.4, donde se usa una función "lineal por segmentos" de regresión para aproximar la verdadera relación entre y y x.

En general, las ecuaciones de regresión sólo son válidas dentro del rango de las variables regresoras contenidas en los datos observados. Por ejemplo, se tiene la figura 1.5, supongamos que se reunieron datos de y y x en el intervalo $x_1 \leq x \leq x_2$. En este intervalo, la ecuación de regresión lineal de la figura 1.5 es una buena aproximación de la verdadera relación. Sin embargo, supongamos que se usará esta ecuación para calcular valores de y con valores de la variable regresora en la región $x_2 \leq x \leq x_3$. Es claro que no funciona bien el modelo de regresión lineal dentro de este intervalo de x, porque hay error de modelo o error de ecuación.

En general, la variable de respuesta y se puede relacionar con k regresores $x_1, x_2, \ldots, x_k$, de modo que

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{1.3}$$

A esto se le llama **modelo de regresión lineal múltiple**, ya que implica a más de un regresor. El adjetivo lineal es para indicar que el modelo es lineal respecto a los parámetros $\beta_0, \beta_1, \ldots, \beta_k$, y no porque y sea una función lineal de las x. Después veremos que muchos modelos en los que y se relaciona con las x en forma no lineal, se pueden seguir manejando como modelos de regresión lineal, siempre y cuando la ecuación sea lineal en las β.

http://carlos2524.jimdo.com/

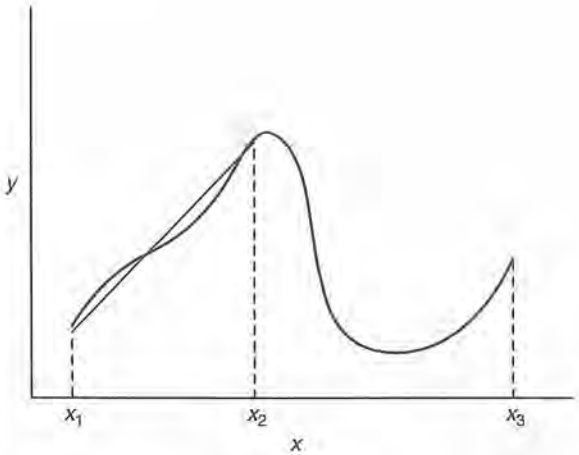


Figura 1.5 El riesgo de la extrapolación en la regresión.

Un objetivo importante del análisis de regresión es **estimar los parámetros desconocidos** en el modelo de regresión. También se le llama a este proceso ajuste del modelo a los datos. Estudiaremos varias técnicas de estimación de parámetros en este libro. Una de ellas es el método de mínimos cuadrados (véase el Cap. 2). Por ejemplo, el ajuste de los datos de tiempo de entrega por cuadrados mínimos es

$$\hat{y} = 3.3208 + 2.1762x$$

donde $\hat{y}$ es el valor ajustado, estimado o calculado, del tiempo de entrega que corresponde a un volumen de entrega de x cajas. Esta ecuación ajustada es la que se graficó en la figura 1.1b.

La siguiente fase del análisis de regresión se llama **comprobación de la adecuación del modelo** en donde se estudia lo apropiado del modelo y la calidad del ajuste determinado. Mediante esos análisis se puede determinar la utilidad del modelo de regresión. El resultado de la comprobación de adecuación puede indicar que el modelo es razonable, o que debe modificarse el ajuste original. Por lo anterior, el análisis de regresión es un procedimiento **iterativo**, en el que los datos conducen a un modelo, y se produce un ajuste del modelo a los datos. A continuación se investiga la calidad del ajuste y se pasa a modificar el modelo, o el ajuste, o a adoptar el modelo. Este proceso se ilustrará varias veces en los capítulos que siguen.

Un modelo de regresión no implica que haya una relación de causa efecto entre las variables. Aunque pueda existir una marcada relación empírica entre dos o más variables, no puede considerarse como prueba de que las variables regresoras y la respuesta estén relacionadas en forma de causa-efecto. Para establecer la causalidad, la relación entre los regresores y la respuesta debe tener una base ajena a los datos de la muestra, por ejemplo, la relación puede indicarse a través de consideraciones teóricas. El análisis de regresión ayudará a confirmar la relación de causa-efecto, pero no puede ser la base única para ésta.

Por último, es esencial recordar que el análisis de regresión es una parte de un método más amplio, de análisis de datos para resolver problemas. Esto es, la ecuación misma de regresión puede no ser el objetivo principal del estudio, ya que es más importante comprender y apreciar el sistema que genera los datos.

http://carlos2524.jimdo.com/

1.2 RECOLECCIÓN DE DATOS

Un aspecto esencial del análisis de regresión es la recolección, recopilación o adquisición de datos. Todo análisis de regresión es tan bueno como lo son los datos sobre los que se basa. Hay tres métodos básicos de recolectar datos:

- Un estudio retrospectivo basado en datos históricos.
- Un estudio observacional.
- Un experimento diseñado.

Un buen esquema de recolección de datos puede asegurar un análisis simplificado y un modelo de aplicación más general. Un mal esquema de recolección de datos induce a graves problemas en el análisis y su interpretación. El ejemplo que sigue ilustra estos tres métodos.

Ejemplo

Se tiene una columna de destilación de acetona-alcohol butílico. Los factores posibles que pueden influir sobre la concentración de acetona en la corriente de destilado (el producto) son: la temperatura del hervidor, la temperatura del condensado y la tasa de reflujo. El departamento de producción mantiene y archiva los siguientes registros:

- La concentración de acetona en una muestra de la corriente de producto, tomada cada hora.
- La gráfica del controlador de temperatura del hervidor.
- La gráfica del controlador de temperatura del condensador.
- La tasa nominal de reflujo, cada hora.

Se supone que para este proceso la tasa de reflujo debe ser constante, sólo en raras ocasiones cambia esta tasa.

Estudio retrospectivo

Se podría hacer un estudio retrospectivo que utilice todos los datos históricos del proceso, o una muestra de ellos, dentro de algún periodo, para determinar las relaciones entre las dos temperaturas y la tasa de reflujo, con la concentración de acetona en la corriente de producto. Al hacerlo se aprovecha la ventaja de contar con datos previamente reunidos, y minimizar el costo del estudio. Sin embargo, se debe hacer notar que hay varios problemas.

1. En realidad, no se puede ver el efecto de la relación de reflujo sobre la concentración, porque se debe suponer que no varió mucho durante el periodo histórico.
2. Los datos que relacionan las dos temperaturas con la concentración de acetona no se corresponden en forma directa. Para establecer una correspondencia aproximada se presentan, por lo general, grandes dificultades.

http://carlos2524.jimdo.com/

3. El departamento de producción mantiene ambas temperaturas lo más cercanas posible a los valores específicos deseados, usando controladores automáticos. Como las dos temperaturas varían tan poco a través del tiempo, dificultará poder apreciar su impacto real sobre la concentración.
4. Dentro de los límites estrechos, entre los que varía, la temperatura del condensado tiende a aumentar con la del hervidor. En consecuencia, habrá dificultad para separar los efectos individuales de las dos temperaturas. Esto conduce al problema de **colinealidad** o de **multicolinealidad**, que describiremos en el capítulo 10.

Los estudios restrospectivos ofrecen, con frecuencia, cantidades limitadas de informa ción útil. En general, sus principales desventajas son:

- Con frecuencia faltan algunos de los datos importantes.
- La fiabilidad y la calidad de los datos suelen ser muy dudosas.
- La naturaleza de los datos con frecuencia pueden no permitir atacar el problema a la mano.
- El analista trata, con frecuencia, de usar los datos en formas que nunca se pretendió que se usaran.
- Los registros, cuadernos de notas y memorias pueden no explicar fenómenos interesantes que identifica el análisis de datos.

El uso de datos históricos siempre corre el riesgo de que, por cualquier razón, algunos de los datos se perdieron o no se anotaron. Normalmente éstos consisten en información que se considera crítica, o que conviene registrar. Cuando ésta es esencial se reúne con gran cuidado y exactitud. La información conveniente no se reúne así, muchas veces. En consecuencia, los datos históricos suelen sufrir de errores de transcripción y otros problemas con la calidad de datos. Esos errores hacen que los datos históricos sean propensos a tener **datos atípicos**. Un análisis de regresión sólo es tan fiable como los datos sobre los que se basa.

El que sea conveniente reunir ciertos datos no significa que sean de utilidad particular. Con frecuencia hay datos que no se consideran esenciales para vigilar el proceso rutinario, y que no conviene reunir, pero que sí tienen un impacto apreciable sobre el proceso. Los datos históricos no pueden proporcionar esta información, porque en este caso nunca se coleccionaron. Por ejemplo, puede ser que la temperatura ambiente influya sobre las pérdidas de calor en el ejemplo de la columna de destilación. En los días fríos, la columna pierde más calor que durante los días muy cálidos. Los registros de producción de esta columna de acetona y alcohol butílico no registran la temperatura ambiente, por lo tanto, los datos históricos no permiten al analista incluir este factor en el análisis, aunque tenga cierta importancia.

En algunos casos se trata de usar datos que se coleccionaron como derivados de los que en realidad se necesita coleccionar. El análisis que resulta sólo es informativo hasta el grado en que esos datos derivados realmente reflejen lo que representan. Por ejemplo, la naturaleza de la mezcla de acetona y alcohol butílico en la entrada puede afectar mucho el funcionamiento de la columna, ya que se diseñó para que la entrada fuera un líquido saturado (que estuviera al punto de ebullición de la mezcla). Los registros de producción muestran la temperatura de alimentación, pero no las concentraciones específicas de acetona y alcohol butílico en la corriente de alimentación. Es muy difícil determinar, con regularidad, esas concentraciones. En este caso, la temperatura de entrada es un dato derivado de la naturaleza de la mezcla de entrada. Es perfectamente posible que la alimentación esté a la

http://carlos2524.jimdo.com/

temperatura específica correcta, y que la entrada sea un líquido subenfriado, o una mezcla de líquido y vapor.

En algunos casos, sucede que los datos recolectados más informalmente, y por consiguiente con menos calidad, menos exactitud y fiabilidad mínima, tienen mucha influencia para explicar la respuesta. Dicha influencia puede ser real, o ser un artificio relacionado con las inexactitudes de los datos. Cuando hay demasiados análisis se llegan a conclusiones inválidas, porque se dio demasiado crédito a datos que nunca se pretendió usar estrictamente para el análisis.

Finalmente, el objetivo principal de muchos análisis es aislar las causas básicas de fenómenos interesantes. Éstos pueden haber sucedido meses o años atrás y según los datos históricos, pasar desapercibidos. Los registros y cuadernos de notas con frecuencia no permiten tener mucho acceso a esas causas básicas, y naturalmente, la memoria comienza a fallar con el tiempo.

Estudio observacional

Se podría usar un estudio observacional para recolectar datos para el problema. Como su nombre lo indica, en un estudio observacional sólo se observa el proceso o la población y se interacciona o perturba el proceso lo necesario para obtener datos relevantes. Planteándolo adecuadamente, estos estudios pueden asegurar datos exactos, completos y fiables, a la vez que suelen proporcionar información muy limitada acerca de las relaciones específicas entre los datos.

En el ejemplo de la columna de destilación se establece un formato de recolección de datos que permite que el personal de producción anote las dos temperaturas y la proporción real de reflujo, en horas especificadas, que correspondan a la concentración observada de acetona en la corriente de producto. Dicho formato debe permitir la posibilidad de agregar comentarios para anotar cualquier fenómeno interesante que pueda suceder. Con ese procedimiento se aseguraría una recopilación exacta y fiable de datos, y se resolverían los problemas 1) y 2) anteriores. Este método también reduce al mínimo las probabilidades de observar un dato atípico relacionado con algún error en los datos. Desafortunadamente, un estudio observacional no puede resolver los problemas 3) y 4). En consecuencia, los estudios observacionales se prestan a tener problemas con la colinealidad.

Experimento diseñado

Para este problema, la mejor estrategia de recolección de datos es hacer un experimento diseñado donde se puedan manipular las dos temperaturas y la relación de reflujo, a los que se llama factores, de acuerdo con una estrategia bien definida, llamada diseño de experimentos. Esta estrategia debe asegurar que se puedan separar los efectos de cada factor sobre la concentración de acetona. Los valores especificados de los factores, que se ajustan en el experimento, se llaman niveles. Comúnmente se usa una pequeña cantidad (dos o tres) de niveles para cada factor. Supongamos que, para el ejemplo de la columna de destilación, se usa un nivel "alto" o "+1" y un nivel "bajo" o "-1". De este modo se usarían dos niveles para cada uno de los tres factores. Una combinación de tratamientos es una combinación específica de los niveles de cada factor. Cada vez que se lleva a cabo un tratamiento se tiene una corrida experimental. El plan o diseño del experimento consiste en una serie de corridas.

http://carlos2524.jimdo.com/

TABLA 1.1 El experimento diseñado para la columna de destilación

Temp. de la columna	Temp. condensado	Reflujo
−1	−1	−1
+1	−1	−1
−1	+1	−1
+1	+1	−1
−1	−1	+1
+1	−1	+1
−1	+1	+1
+1	+1	+1

Para el ejemplo de la destilación, una estrategia experimental muy razonable usa cada combinación posible de tratamientos para formar un experimento básico, con ocho corridas diferentes en el proceso. La tabla 1.1 muestra esas combinaciones de niveles altos y bajos.

La figura 1.6 muestra que este diseño forma un cubo, desde el punto de vista de esos niveles altos y bajos. Cada corrida con las condiciones del proceso deja que la columna llegue al equilibrio, se toma una muestra de la corriente del producto y se determina la concentración de acetona. A continuación se podrán sacar inferencias específicas acerca del efecto de esos factores, además que este método permite estudiar una población o proceso en forma proactiva.

1.3 USOS DE LA REGRESIÓN

Los modelos de regresión se usan con varios fines, que incluyen los siguientes:

1. Descripción de datos.
2. Estimación de parámetros.
3. Predicción y estimación.
4. Control.

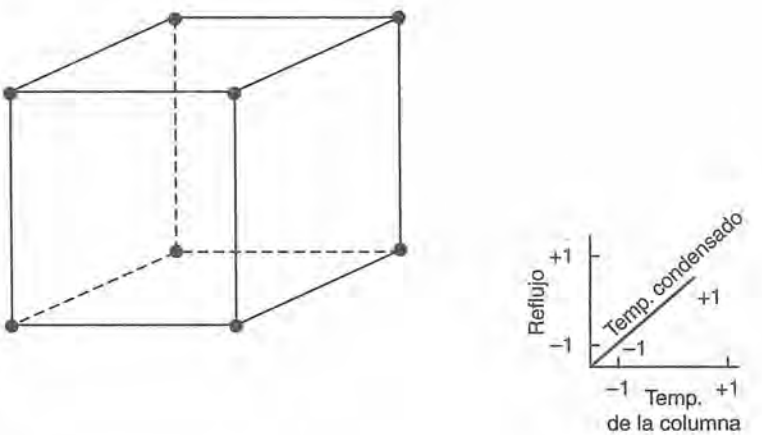


Figura 1.6 El experimento diseñado para la columna de destilación.

http://carlos2524.jimdo.com/

Es común que los ingenieros y los científicos usen ecuaciones para resumir o describir un conjunto de datos. El análisis de regresión es útil para plantear esas ecuaciones. Por ejemplo, se puede reunir una cantidad considerable de tiempo y volumen de entrega, por lo cual un modelo de regresión sería probablemente un resumen mucho más conveniente y útil de esos datos, más que una tabla o una gráfica.

A veces se pueden resolver por métodos de regresión los problemas de estimación de parámetros. Por ejemplo, en ingeniería química se usa la ecuación de Michaelis-Menten, $y = \beta_1 x/(x + \beta_2) + \varepsilon$ para describir la relación entre velocidad de reacción y y concentración, x. Ahora bien, en este modelo β_1 es la velocidad asintótica de reacción, esto es, la velocidad máxima cuando la concentración se hace grande. Si se dispone de una muestra de valores observados de velocidad a distintas concentraciones, el ingeniero puede aplicar el análisis de regresión para ajustar este modelo a los datos, y obtener un estimado de la velocidad máxima. Más adelante indicaremos cómo ajustar los modelos de regresión de este tipo.

Muchas aplicaciones de regresión requieren de la predicción de la variable de respuesta. Por ejemplo, se podría tratar de predecir el tiempo en el que se puede entregar una cantidad específica de cajas de refrescos. Estas predicciones pueden ser útiles para planear actividades de entrega, como por ejemplo, las rutas y los programas, o para evaluar la productividad en las operaciones de entrega. Ya se han discutido los peligros de extrapolar cuando se usa un modelo de regresión para pronosticar, debidos a errores en el modelo o a la ecuación (véase la Fig. 1.5). Sin embargo, aun cuando sea correcta la forma del modelo, las malas estimaciones de los parámetros de éste pueden seguir causando mal desempeño de la predicción.

Los modelos de regresión se pueden usar para fines de control. Por ejemplo, un ingeniero químico podría aplicar el análisis de regresión para establecer un modelo que relacionara la resistencia del papel a la tensión con la concentración de fibra corta (es decir, de madera dura) en la pulpa. Esta ecuación se podría usar después para controlar la resistencia dentro de valores adecuados, variando la concentración de fibra corta. Cuando se usa una ecuación de regresión para fines de control, es importante que las variables estén relacionadas en forma causal. Nótese que podría no necesitarse una relación de causa a efecto si sólo se usara la ecuación para predicción. En este caso sólo es necesario que las relaciones que existían en los datos originales con los que se formuló la ecuación de regresión sigan siendo válidas. Por ejemplo, el consumo diario de electricidad en agosto en Atlanta, Georgia, puede ser un buen predictor de la temperatura diaria máxima en ese mes y en ese lugar. Sin embargo, es claro que cualquier intento de reducir la temperatura máxima restringiendo el consumo de electricidad, está condenado al fracaso.

1.4 EL PAPEL DE LA COMPUTADORA

El establecimiento de un modelo de regresión es un proceso iterativo. Comúnmente, se requieren varios análisis a medida que se descubren mejoras en la estructura del modelo y fallas en los datos. Un buen programa de cómputo para regresión es una herramienta necesaria en este procedimiento.

La aplicación rutinaria de programas estándar de cómputo para regresión no conduce, con frecuencia, a buenos resultados. La computadora **no es** sustituto del razonamiento creativo sobre el problema. En el análisis de regresión se requiere el uso **inteligente** y **diestro** de la computadora. Se debe aprender a interpretar lo que la computadora dice, y cómo

http://carlos2524.jimdo.com/

incorporar esa información en los modelos siguientes. En este libro, describiremos e ilustraremos, el uso de varios programas de cómputo para regresión. En general, esos programas son parte de paquetes generales de estadística, como SAS, Minitab, STATGRAPHICS y S-PLUS. Nuestros objetivos no serán aprender el funcionamiento detallado de esos programas, sino más bien comprender la información que contienen los resultados. Sin estos conocimientos, es virtualmente imposible formular un modelo de regresión.

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

CAPÍTULO 2

REGRESIÓN LINEAL SIMPLE

2.1 MODELO DE REGRESIÓN LINEAL SIMPLE

Este capítulo explica el **modelo de regresión lineal simple**, un modelo con un solo regresor x que tiene una relación con una respuesta y, donde la relación es una línea recta. Este modelo de regresión lineal simple es

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{2.1}$$

donde la ordenada al origen β_0 y la pendiente β_1 son constantes desconocidas, y ε es un componente aleatorio de error. Se supone que los errores tienen promedio cero y varianza σ^2 desconocida. Además, se suele suponer que los errores no están correlacionados. Esto quiere decir que el valor de un error no depende del valor de cualquier otro error.

Conviene considerar que el regresor x está controlado por el analista de datos, y se puede medir con error despreciable, mientras que la respuesta y es una variable aleatoria. Con lo que hay una distribución de probabilidades de y para cada valor posible de x. La media de esta distribución es

$$E(y|x) = \beta_0 + \beta_1 x \tag{2.2a}$$

y la varianza es

$$\text{Var}(\, y|x\,) = \text{Var}(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2 \tag{2.2b}$$

Así, la media de y es una función lineal de x, aunque la varianza de y no depende del valor de x. Además, ya que los errores no están correlacionados, las respuestas tampoco lo están.

A los parámetros β_0 y β_1 se les suele llamar **coeficientes de regresión**. Éstos tienen una interpretación simple y, frecuentemente, útil. La pendiente β_1 es el cambio de la media de la distribución de y producido por un cambio unitario en x. Si el intervalo de los datos incluye a $x = 0$, entonces la ordenada al origen, β_0, es la media de la distribución de la respuesta y cuando $x = 0$. Si no incluye al cero, β_0 no tiene interpretación práctica.

2.2 ESTIMACIÓN DE LOS PARÁMETROS POR MÍNIMOS CUADRADOS

Los parámetros β_0 y β_1 son desconocidos, y se deben estimar con los datos de la muestra. Supongamos que hay n pares de datos: $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$. Como se indicó en el capítulo 1, esos datos pueden obtenerse en un experimento controlado, diseñado en forma específica para recolectarlos, o en un estudio observacional, o a partir de registros históricos existentes (lo que se llama un estudio retrospectivo).

http://carlos2524.jimdo.com/

2.2.1 Estimación de β_0 y β_1

Para estimar β_0 y β_1 se usa el **método de mínimos cuadrados**. Esto es, se estiman β_0 y β_1 tales que la suma de los cuadrados de las diferencias entre las observaciones y_i y la línea recta sea mínima. Según la ecuación (2.1), se puede escribir

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{2.3}$$

Se puede considerar que la ecuación (2.1) es un **modelo poblacional de regresión**, mientras que la ecuación (2.3) es un **modelo muestral de regresión**, escritos en términos de los n pares de datos (y_i, x_i) $(i = 1, 2, \ldots, n)$. Así, el criterio de mínimos cuadrados es

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{2.4}$$

Los estimadores, por mínimos cuadrados, de β_0 y β_1, que se designarán por $\hat{\beta}_0$ y $\hat{\beta}_1$, deben satisfacer

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

y

$$\left.\frac{\partial S}{\partial \beta_1}\right|_{\hat{\beta}_0, \hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)x_i = 0$$

Se simplifican estas dos ecuaciones y se obtiene

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \tag{2.5}$$

Las ecuaciones (2.5) son llamadas **ecuaciones normales de mínimos cuadrados**. Su solución es la siguiente:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{2.6}$$

y

$$\hat{\beta}_1 = \frac{\displaystyle\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\displaystyle\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \tag{2.7}$$

http://carlos2524.jimdo.com/

en donde

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{y} \qquad \bar{x} \quad \frac{1}{n}\sum_{i=1}^{n} x_i$$

son los promedios de y_i y x_i, respectivamente. Por consiguiente, $\hat{\beta}_0$ y $\hat{\beta}_1$ en las ecuaciones (2.6) y (2.7) son los **estimadores por mínimos cuadrados** de la ordenada al origen y la pendiente, respectivamente. El modelo ajustado de regresión lineal simple es, entonces,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{2.8}$$

La ecuación (2.8) produce un estimado puntual, de la media de y para una determinada x.

Como el denominador de la ecuación (2.7) es la suma corregida de cuadrados de las x_i y el numerador es la suma corregida de los productos cruzados de x_i y y_i, estas ecuaciones se pueden escribir en una forma más compacta como sigue:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{2.9}$$

y

$$S_{xy} = \sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n} = \sum_{i=1}^{n} y_i(x_i - \bar{x}) \tag{2.10}$$

Entonces, una forma cómoda de escribir la ecuación (2.7) es

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \tag{2.11}$$

La diferencia entre el valor observado y_i y el valor ajustado correspondiente $\hat{y}_i$ se llama **residual**. Matemáticamente, el i-ésimo residual es

$$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \quad i = 1, 2, \ldots, n \tag{2.12}$$

Los residuales tienen un papel importante para investigar la **adecuación** del modelo de regresión ajustado, y para detectar diferencias respecto a las hipótesis básicas. Este tema se describirá en capítulos posteriores.

Ejemplo 2.1 *Datos del propelente*

Un motor cohete se forma pegando entre sí un propelente de ignición y un propelente de sostenimiento dentro de una caja metálica. La resistencia al corte de la pegadura entre los dos propelentes es una característica importante de la calidad. Se cree que la resistencia al corte se relaciona con la edad, en semanas, del lote del propelente de sostenimiento. Se

http://carlos2524.jimdo.com/

hicieron 20 observaciones de resistencia al corte y la edad del lote correspondiente de propelente, y se ven en la tabla 2.1. El diagrama de dispersión que se ve en la figura 2.1 parece indicar que hay una fuerte relación estadística entre la resistencia al cortante y la edad del propelente, y que parece razonable la hipótesis tentativa del modelo de línea recta, $y = \beta_0 + \beta_1 x + \varepsilon$.

TABLA 2.1 Datos para el ejemplo 2.1

Observación i	Resistencia al corte (psi) y_i	Edad del propelente (semanas) x_i
1	2158.70	15.50
2	1678.15	23.75
3	2316.00	8.00
4	2061.30	17.00
5	2207.50	5.50
6	1708.30	19.00
7	1784.70	24.00
8	2575.00	2.50
9	2357.90	7.50
10	2256.70	11.00
11	2165.20	13.00
12	2399.55	3.75
13	1779.80	25.00
14	2336.75	9.75
15	1765.30	22.00
16	2053.50	18.00
17	2414.40	6.00
18	2200.50	12.50
19	2654.20	2.00
20	1753.70	21.50

Para estimar los parámetros del modelo se calcula primero:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 4\,677.69 - \frac{71\,422.56}{20} = 1\,106.56$$

y

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} = 528\,492.64 - \frac{(267.25)\,(42\,627.15)}{20}$$

$$= -41\,112.65$$

http://carlos2524.jimdo.com/

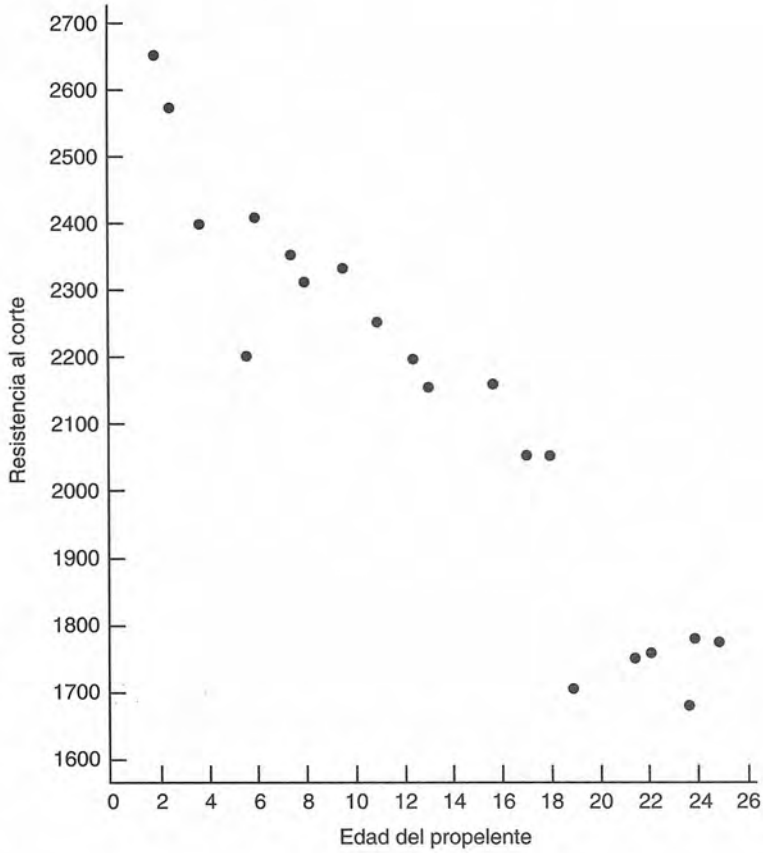


Figura 2.1 Diagrama de dispersión de la resistencia cortante en función de la edad del propelente. Ejemplo 2.1.

Por consiguiente, según las ecuaciones (2.11) y (2.6), se ve que

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-41\,112.65}{1\,106.56} = -37.15$$

y

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 2\,131.3575 - (-37.15)13.3625 = 2\,627.82$$

El ajuste de mínimos cuadrados es

$$\hat{y} = 2\,627.82 - 37.15x$$

Se puede interpretar que la pendiente de –37.15 es la disminución semanal promedio de resistencia del propelente al corte, debido a su edad. Como el límite inferior de las x está cerca del origen, la ordenada al origen de 2 627.82 representa la resistencia al corte de un lote de propelente inmediatamente después de ser fabricado. La tabla 2.2 muestra los valores observados y_i, los valores calculados o "ajustados" $\hat{y}_i$ y los residuales.

http://carlos2524.jimdo.com/

TABLA 2.2 Datos, valores ajustados y residuales para el ejemplo 2.1

Valor observado, y_i	Valor ajustado, $\hat{y}_i$	Residual, e_i
2158.70	2051.94	106.76
1678.15	1745.42	−67.27
2316.00	2330.59	−14.59
2061.30	1996.21	65.09
2207.50	2423.48	−215.98
1708.30	1921.90	−213.98
1784.70	1736.14	48.56
2575.00	2534.94	40.06
2357.90	2349.17	8.73
2256.70	2219.13	37.57
2165.20	2144.83	20.37
2399.55	2488.50	−88.95
1779.80	1698.98	80.82
2336.75	2265.58	71.17
1765.30	1810.44	−45.14
2053.50	1959.06	94.44
2414.40	2404.90	9.50
2200.50	2163.40	37.10
2654.20	2553.52	100.68
1753.70	1829.02	−75.32
$\Sigma y_i = 42627.15$	$\Sigma \hat{y}_i = 42627.15$	$\Sigma e_i = 0.00$

Después de obtener el ajuste por mínimos cuadrados, surgen varias preguntas interesantes:

1. ¿Qué tan bien se ajusta esta ecuación a los datos?
2. ¿Es probable que el modelo sea útil como predictor?
3. ¿Se viola alguna de las hipótesis básicas (como la de varianza constante y la de errores no correlacionados)? y en caso afirmativo, ¿qué tan grave es eso?

Se deben investigar todos estos asuntos antes de adoptar al modelo en forma definitiva y usarlo. Como se dijo anteriormente, los residuales juegan un papel clave para evaluar la adecuación del modelo. Se puede considerar que los residuales son realizaciones de los errores ε_i del modelo. Así, para comprobar la constancia de la varianza y la hipótesis de errores no correlacionados, uno se debe preguntar si los residuales parecen ser realmente una muestra aleatoria de una distribución con esas propiedades. Regresaremos a estas cuestiones en el capítulo 4, al investigar el uso de los residuales en la comprobación de la adecuación del modelo.

Resultado en computadora

Los paquetes de programas de cómputo se usan mucho para ajustar modelos de regresión. Las rutinas de regresión se encuentran tanto en programas estadísticos para computadora central, como para PC, y también en muchos de los paquetes más usados de hojas de cálculo. La tabla 2.3 representa el resultado obtenido con Minitab, un paquete estadístico para PC de uso

http://carlos2524.jimdo.com/

muy frecuente, con los datos del propelente de cohetes del ejemplo 2.1. La parte superior de la tabla contiene el modelo ajustado de regresión. Obsérvese que antes de redondear, los coeficientes de regresión concuerdan con los que se calcularon manualmente. También, la tabla 2.3 contiene más información sobre el modelo de regresión. Regresaremos a este resultado y explicaremos esas cantidades más adelante.

TABLA 2.3 Resultados de regresión para el ejemplo 2.1 con Minitab

Regression Analysis

The regression equation is
Strength = 2628 – 37.2 Age

Predictor	Coef	StDev	T	P
Constant	2627.82	44.18	59.47	0.000
Age	–37.154	2.889	–12.86	0.000
S = 96.11	R-Sq = 90.2%		R-Sq(adj) = 89.6%	

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1527483	1527483	165.38	0.000
Error	18	166255	9236		
Total	19	1693738			

2.2.2 Propiedades de los estimadores por mínimos cuadrados y el modelo ajustado de regresión

Los estimadores por cuadrados mínimos $\hat{\beta}_0$ y $\hat{\beta}_1$ tienen algunas propiedades importantes. Primero, obsérvese que, según las ecuaciones (2.6) y (2.7), $\hat{\beta}_0$ y $\hat{\beta}_1$ son **combinaciones lineales** de las observaciones y_i. Por ejemplo

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^{n} c_i y_i$$

donde $c_i = (x_i - \bar{x})/S_{xx}$, para $i = 1, 2, \ldots, n$.

Los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ por mínimos cuadrados son **estimadores insesgados** de los parámetros β_0 y β_1 del modelo. Para demostrarlo con $\hat{\beta}_1$, considérese

$$E(\hat{\beta}_1) = E\left(\sum_{i=1}^{n} c_i y_i\right) = \sum_{i=1}^{n} c_i E(y_i)$$

$$= \sum_{i=1}^{n} c_i(\beta_0 + \beta_1 x_i) = \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i$$

ya que, se supuso que, $E(\varepsilon_i) = 0$. Ahora se puede demostrar en forma directa que $\sum_{i=1}^{n} c_i = 0$ y que $\sum_{i=1}^{n} c_i x_i = 1$, y entonces

$$E(\hat{\beta}_1) = \beta_1$$

http://carlos2524.jimdo.com/

Esto es, si se supone que el modelo es correcto [que $E(y_i) = \beta_0 + \beta_1 x_1$], entonces $\hat{\beta}_1$ es un estimador insesgado de β_1. De igual manera se puede demostrar que $\hat{\beta}_0$ es un estimador insesgado de β_0, es decir,

$$E(\hat{\beta}_0) = \beta_0$$

La varianza de $\hat{\beta}_1$ se calcula como sigue:

$$\begin{aligned} \operatorname{Var}(\hat{\beta}_1) &= \operatorname{Var}\left(\sum_{i=1}^{n} c_i y_i\right) \\ &= \sum_{i=1}^{n} c_i^2 \operatorname{Var}(y_i) \end{aligned} \tag{2.13}$$

ya que las observaciones y_i son no correlacionadas, por lo que la varianza de la suma es igual a la suma de las varianzas. La varianza de cada término en la suma es $c_i^2 \operatorname{Var}(y_i)$ y hemos supuesto que $\operatorname{Var}(y_i) = \sigma^2$; en consecuencia,

$$\begin{aligned} \operatorname{Var}(\hat{\beta}_1) &= \sigma^2 \sum_{i=1}^{n} c_i^2 = \frac{\sigma^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{S_{xx}^2} \\ &= \frac{\sigma^2}{S_{xx}} \end{aligned} \tag{2.14}$$

La varianza de $\hat{\beta}_0$ es

$$\begin{aligned} \operatorname{Var}(\hat{\beta}_1) &= \operatorname{Var}(\bar{y} - \hat{\beta}_1 \bar{x}) \\ &= \operatorname{Var}(\bar{y}) + \bar{x}^2 \operatorname{Var}(\hat{\beta}_1) - 2\bar{x}\operatorname{Cov}(\bar{y}, \hat{\beta}_1) \end{aligned}$$

Ahora bien, la varianza de $\bar{y}$ no es más que $\operatorname{Var}(\bar{y}) = \sigma^2/n$, y se puede demostrar que la covarianza entre $\bar{y}$ y $\hat{\beta}_1$ es cero (véase el Prob. 2.19). Así,

$$\begin{aligned} \operatorname{Var}(\hat{\beta}_0) &= \operatorname{Var}(\bar{y}) + \bar{x}^2 \operatorname{Var}(\hat{\beta}_1) \\ &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \end{aligned} \tag{2.15}$$

Otro resultado importante acerca de la calidad de los estimadores por mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$ es el **teorema de Gauss-Markov**, que establece que para el modelo de regresión (2.1) con las hipótesis $E(\varepsilon) = 0$, $\operatorname{Var}(\varepsilon) = \sigma^2$ y con errores no correlacionados, los estimadores por mínimos cuadrados son insesgados y tienen varianza mínima en comparación con todos los demás estimadores insesgados que sean combinaciones lineales de las y_i. Con frecuencia se dice que los estimadores por mínimos cuadrados son los **estimadores lineales insesgados óptimos**, donde "óptimos" implica que son de varianza mínima.

http://carlos2524.jimdo.com/

En el apéndice C.4 se demuestra el teorema de Gauss-Markov para el caso más general de regresión lineal múltiple, del cual la regresión lineal simple es un caso especial.

Hay varias otras propiedades útiles del ajuste por mínimos cuadrados:

1. La suma de los residuales en cualquier modelo de regresión que contenga una ordenada al origen β_0 siempre es igual a cero, esto es,

$$\sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} e_i = 0$$

   Esta propiedad es consecuencia directa de la primera de las ecuaciones normales (2.5), y se demuestra en la tabla 2.2 para los residuales del ejemplo 2.1. Redondear los errores puede afectar la suma.

2. La suma de los valores observados y_i es igual a la suma de los valores ajustados $\hat{y}_i$:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$$

   La tabla 2.2 demuestra este resultado para el ejemplo 2.1.

3. La línea de regresión de mínimos cuadrados siempre pasa por el **centroide** de los datos, que es el punto $(\bar{y}, \bar{x})$.

4. La suma de los residuales, ponderados por el valor correspondiente de la variable regresora, siempre es igual a cero:

$$\sum_{i=1}^{n} x_i e_i = 0$$

5. La suma de los residuales, ponderados por el valor ajustado correspondiente, siempre es igual a cero:

$$\sum_{i=1}^{n} \hat{y}_i e_i = 0$$

2.2.3 Estimación de σ^2

Además de estimar β_0 y β_1, se requiere un estimado de σ^2 para probar hipótesis y formar estimados de intervalo pertinentes al modelo de regresión. En el caso ideal este estimado no debería depender de la adecuación del modelo ajustado. Eso sólo es posible cuando hay varias observaciones de y para cuando menos un valor de x (véase la Sec. 3.4) o cuando se dispone de información anterior acerca de σ^2. Cuando no se puede usar este método, el estimado de σ^2 se obtiene de la suma de cuadrados de residuales, o suma de cuadrados de error:

$$SSr_{\text{Res}} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.16}$$

Se puede deducir una fórmula cómoda para calcular SS_{Res} sustituyendo $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ en la ecuación (2.16), y simplificando.

http://carlos2524.jimdo.com/

Así se llega a

$$SS_{\text{Res}} = \sum_{i=1}^{n} y_i^2 = n\bar{y}^2 - \hat{\beta}_1 S_{xy} \tag{2.17}$$

Pero

$$\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 \equiv SS_{\text{T}}$$

es justo la suma de cuadrados corregida, de las observaciones de la respuesta, por lo que

$$SS_{\text{Res}} = SS_{\text{T}} - \hat{\beta}_1 S_{xy} \tag{2.18}$$

La suma de cuadrados de residuales tiene $n - 2$ grados de libertad, porque dos grados de libertad se asocian con los estimados $\hat{\beta}_0$ y $\hat{\beta}_1$ que se usan para obtener $\hat{y}_i$. El apéndice C.3 demuestra que el valor esperado de SS_{Res} es $E(SS_{\text{Res}}) = (n-2)\sigma^2$, por lo que un **estimador insesgado de** σ^2 es

$$\hat{\sigma}^2 = \frac{SS_{\text{Res}}}{n-2} = MS_{\text{Res}} \tag{2.19}$$

La cantidad MS_{Res} se llama **cuadrado medio residual**. La raíz cuadrada de $\hat{\sigma}^2$ se llama, a veces, el **error estándar de la regresión** y tiene las mismas unidades que la variable de respuesta y.

Ya que $\hat{\sigma}^2$ depende de la suma de cuadrados de residuales, cualquier violación de las hipótesis sobre los errores del modelo, o cualquier especificación equivocada de la forma del modelo pueden dañar gravemente la utilidad de $\hat{\sigma}^2$ como estimado de σ^2. Como σ^2 se calcula con los residuales del modelo de regresión, se dice que es un estimado de σ^2 **dependiente del modelo.**

Ejemplo 2.2 *Datos del propelente de reacción*

Para estimar σ^2 de los datos del propelente de cohetes en el ejemplo 2.1, primero se calcula

$$SS_{\text{T}} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$= 92\,547{,}433.45 - \frac{(42\,627.15)^2}{20} = 1\,693{,}737.60$$

De acuerdo con la ecuación (2.19), la suma de cuadrados de residuales es

$$SS_{\text{Res}} = SS_{\text{T}} - \hat{\beta}_1 S_{xy}$$

$$= 1\,693\,737.60 - (-37.15)(-41\,112.65) = 166\,402.65$$

http://carlos2524.jimdo.com/

Por consiguiente el estimado de σ^2 se calcula con la ecuación (2.19) como sigue:

$$\hat{\sigma}^2 = \frac{SS_{\text{Res}}}{n-2} = \frac{166\,402.65}{18} = 9\,244{,}59$$

Recuérdese que este estimado de σ^2 **depende del modelo.**

2.2.4 Una forma alterna del modelo

Hay una forma alterna del modelo de regresión lineal simple, que a veces es útil. Supongamos que se redefine la variable regresora x_i como la desviación respecto a su propio promedio, esto es, $x_i - \bar{x}$. Entonces, el modelo de regresión se transforma en

$$\begin{aligned} y_i &= \beta_0 + \beta_1(x_i - \bar{x}) + \beta_1\bar{x} + \varepsilon_i \\ &= (\beta_0 + \beta_1\bar{x}) + \beta_1(x_i - \bar{x}) + \varepsilon_i \\ &= \beta_0' + \beta_1(x_i - \bar{x}) + \varepsilon_i \end{aligned} \tag{2.20}$$

Nótese que al redefinir la variable regresora en la ecuación (2.20), el origen de las x se desplazó de cero a $\bar{x}$. Para mantener iguales los valores ajustados tanto en el modelo original como en el transformado, es necesario modificar la ordenada al origen que se tenía. La relación entre la ordenada al origen original y la transformada es

$$\beta_0' = \beta_0 + \beta_1\bar{x} \tag{2.21}$$

Es fácil demostrar que el estimador de mínimos cuadrados de la ordenada al origen transformada es $\hat{\beta}_0' = \bar{y}$. El estimador de la pendiente no se afecta con la transformación. Esta forma alterna del modelo tiene ciertas ventajas. La primera es que los estimadores $\hat{\beta}_0' = \bar{y}$ y $\hat{\beta}_1 = S_{xy}/S_{xx}$, de mínimos cuadrados, son **no correlacionados**, esto es, $\text{Cov}(\hat{\beta}_0', \hat{\beta}_1) = 0$. Esto facilita algunas aplicaciones del modelo, como por ejemplo la determinación de los intervalos de confianza del promedio de y (véase la Sec. 2.4.2). Por último, el modelo ajustado es

$$\hat{y} = \bar{y} + \hat{\beta}_1(x - \bar{x}) \tag{2.22}$$

Aunque las ecuaciones (2.22) y (2.8) son equivalentes (las dos dan como resultado el mismo valor de $\bar{y}$ para el mismo valor de x), la ecuación 2.22 recuerda al analista, en forma directa, que el modelo de regresión sólo es válido dentro del intervalo de x en los **datos originales**. Esta región está centrada en $\bar{x}$.

2.3 PRUEBA DE HIPÓTESIS DE LA PENDIENTE Y DE LA ORDENADA AL ORIGEN

Con frecuencia interesa probar hipótesis y establecer intervalos de confianza de los parámetros del modelo. La prueba de hipótesis se explicará en esta sección, y en la sección 2.4 veremos los intervalos de confianza. Estos procedimientos requieren hacer la hipótesis adicional de que los errores ε_i del modelo estén distribuidos normalmente. Así, las hipótesis completas son: que los errores estén distribuidos en forma normal e independiente,

http://carlos2524.jimdo.com/

con media 0 y varianza σ^2, lo cual se abrevia "NID(0, σ^2)". NID viene de *normally and independently distributed* (distribuido normal e independientemente). En el capítulo 4 se describirá cómo se pueden comprobar esas hipótesis a través del **análisis de residuales**.

2.3.1 Uso de las pruebas *t*

Supongamos que se desea probar la hipótesis que la pendiente es igual a una constante, por ejemplo, a β_{10}. Las hipótesis correspondientes son

$$H_0: \beta_1 = \beta_{10}$$
$$H_1: \beta_1 \neq \beta_{10} \tag{2.23}$$

en donde se ha especificado una alternativa bilateral. Como los errores ε_i son NID(0, σ^2), las observaciones y_i son NID($\beta_0 + \beta_1 x_i$, σ^2). Ahora, $\hat{\beta}_1$ es una combinación lineal de las observaciones, de modo que $\hat{\beta}_1$ está distribuido normalmente con promedio β_1 y varianza σ^2/S_{xx}, usando la media y la varianza de $\hat{\beta}_1$ que se determinó en la sección 2.2.2. Por consiguiente, el estadístico

$$Z_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\sigma^2/S_{xx}}}$$

está distribuido $N(0, 1)$ si es cierta la hipótesis nula H_0: $\beta_1 = \beta_{10}$. Si se conociera σ^2, se podría usar Z_0 para probar la hipótesis (2.23). Comúnmente se desconoce σ^2. Ya se ha visto que MS_{Res} es un estimador insesgado de σ^2. En el apéndice C.3 se establece que $(n-2)\, MS_{\text{Res}}/\sigma^2$ tiene una distribución χ^2_{n-2} y que MS_{Res} y $\hat{\beta}_1$ son independientes. De acuerdo con la definición del estadístico t que se presenta en el apéndice C.1,

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{MS_{\text{Res}}/S_{xx}}} \tag{2.24}$$

sigue una distribución t_{n-2} si es cierta la hipótesis nula H_0: $\beta_1 = \beta_{10}$. La cantidad de grados de libertad asociados con t_0 es igual a la cantidad de grados de libertad asociados con MS_{Res}. Así, la razón t_0 es el estadístico con que se prueba H_0: $\beta_1 = \beta_{10}$. El procedimiento de prueba calcula t_0 y compara su valor observado de acuerdo con la ecuación (2.24) con el punto porcentual $\alpha/2$ superior de t_{n-2} de la distribución ($t_{\alpha/2,\, n-2}$). Este procedimiento rechaza la hipótesis nula si

$$|t_0| > t_{\alpha/2,\, n-2} \tag{2.25}$$

También se podría usar el método del valor P para tomar la decisión.

El denominador del estadístico t_0 en la ecuación (2.24) se llama con frecuencia el **error estándar estimado**, o más sencillamente el **error estándar** de la pendiente. Esto es,

$$\text{se}(\hat{\beta}_1) = \sqrt{\frac{MS_{\text{Res}}}{S_{xx}}} \tag{2.26}$$

http://carlos2524.jimdo.com/

Por lo anterior, se ve con frecuencia a t_0 escrito en la forma

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\text{se}(\hat{\beta}_1)} \tag{2.27}$$

Se puede usar un procedimiento parecido para probar hipótesis acerca de la ordenada al origen. Para probar

$$\begin{aligned} H_0 &: \beta_0 = \beta_{00} \\ H_1 &: \beta_0 \neq \beta_{00} \end{aligned} \tag{2.28}$$

se podría usar el **estadístico de prueba**

$$t_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{MS_{\text{Res}}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} = \frac{\hat{\beta}_0 - \beta_{00}}{\text{se}(\hat{\beta}_0)} \tag{2.29}$$

en donde $\text{se}(\hat{\beta}_0) = \sqrt{MS_{\text{Res}}(1/n + \bar{x}^2/S_{xx})}$ es el **error estándar de la ordenada al origen**. La hipótesis nula H_0: $\beta_0 = \beta_{00}$ se rechaza si $|t_0| > t_{\alpha/2,\, n-2}$.

2.3.2 Prueba de significancia de la regresión

Un caso especial muy importante de la hipótesis en la ecuación (2.24) es el siguiente:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned} \tag{2.30}$$

Estas hipótesis se relacionan con la **significancia de la regresión**. El no rechazar H_0: $\beta_1 = 0$ implica que no hay relación lineal entre x y y. Este caso se ilustra en la figura 2.2. Nótese que eso puede implicar que x tiene muy poco valor para explicar la variación de y y que el mejor estimador para cualquier x es $\hat{y} = \bar{y}$ (Fig. 2.2a), o que la verdadera relación entre x y y no es lineal (Fig. 2.2b). Por consiguiente, si no se rechaza H_0: $\beta_1 = 0$, equivale a decir que **no hay relación lineal entre y y x**.

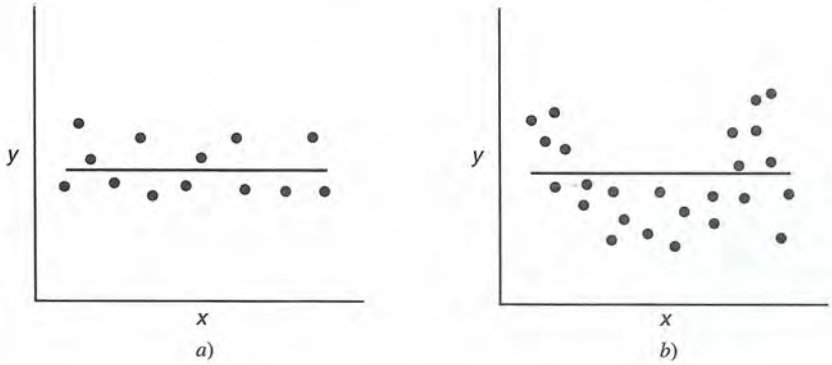


Figura 2.2 Casos en los que no se rechaza la hipótesis H_0: $\beta_1 = 0$.

http://carlos2524.jimdo.com/

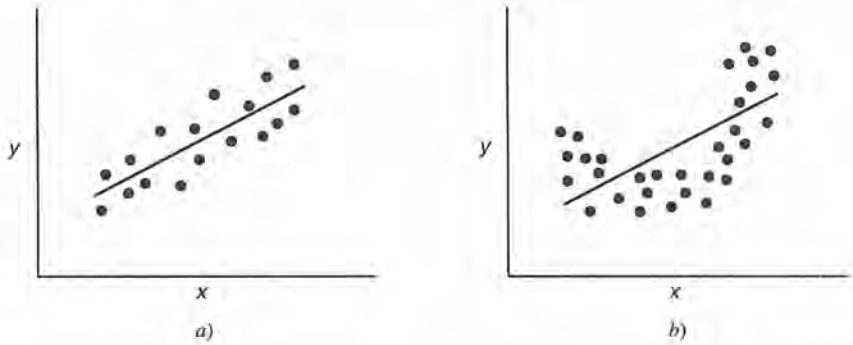


Figura 2.3 Casos en los que no se rechaza la hipótesis H_0: $\beta_1 = 0$.

También, si se rechaza H_0: $\beta_1 = 0$, eso implica que x sí tiene valor para explicar la variabilidad de y. Esto se ilustra en la figura 2.3. Sin embargo, rechazar H_0: $\beta_1 = 0$ podría equivaler a que el modelo de línea recta es adecuado (Fig. 2.3a), o que aunque hay un efecto lineal de x se podrían obtener mejores resultados agregando términos polinomiales en x (Fig. 2.3b).

El procedimiento de prueba para H_0: $\beta_1 = 0$ se puede establecer con dos métodos. El primero tan sólo usa el estadístico t en la ecuación (2.26), con $\beta_{10} = 0$, es decir,

$$t_0 = \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)}$$

La hipótesis de la significancia de la regresión se rechazaría si $|t_0| > t_{\alpha/2,\, n-2}$.

Ejemplo 2.3 *Datos del propelente de reacción*

A continuación se probará la significancia de la regresión en el modelo del propelente de reacción del ejemplo 2.1. El estimado de la pendiente es $\hat{\beta}_1 = -37.15$, y en el ejemplo 2.2 se calculó el estimado de σ^2, que resultó $MS_{\text{Res}} = \hat{\sigma}^2 = 9\,244.59$. El error estándar de la pendiente es

$$\text{se}(\hat{\beta}_1) = \sqrt{\frac{MS_{\text{Res}}}{S_{xx}}} = \sqrt{\frac{9\,244.59}{1\,106.56}} = 2.89$$

Por consiguiente, el estadístico de prueba es

$$t_0 = \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} = \frac{-37.15}{2.89} = -12.85$$

Si se escoge $\alpha = 0.05$, el valor crítico de t es $t_{0.025,\,18} = 2.101$. Así, se rechazaría H_0: $\beta_1 = 0$ y se llegaría a la conclusión que hay una relación lineal entre la resistencia al corte y la edad del propelente.

http://carlos2524.jimdo.com/

Resultados de Minitab

El resultado de Minitab de la tabla 2.3 muestra los errores estándar de la pendiente (en la tabla aparecen como "StDev"), junto con el estadístico t para probar H_0: $\beta_1 = 0$ y H_0: $\beta_0 = 0$. Obsérvese que los resultados de la pendiente, en esta tabla, concuerdan en esencia con los cálculos manuales del ejemplo 2.3. Al igual que en la mayor parte de los programas de cómputo, en Minitab se usa el método del valor P para pruebas de hipótesis. El valor P para probar el significado de la regresión aparece como $P = 0.000$ (es un valor redondeado; el valor real de P es $P = 1.64\text{E-}10$). Es claro que hay fuertes evidencias de que la resistencia se relaciona en forma lineal con la edad del propelente. El estadístico de prueba para H_0: $\beta_0 = 0$ aparece como $t_0 = 59.47$, con $P = 0.000$. Cabría sentirse muy cómodo al afirmar que la ordenada al origen no es cero en este modelo.

2.3.3 Análisis de varianza

También se puede usar un método de **análisis de varianza** para probar el significado de la regresión. Este análisis se basa en una partición de la variabilidad total de la variable y de respuesta. Para obtener esta partición se comienza con la identidad

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \tag{2.31}$$

Se elevan al cuadrado ambos lados de la ecuación (2.30) y se suma para todas las n observaciones. Así se obtiene

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)(y_i - \hat{y}_i)$$

Nótese que el tercer término del lado derecho de esta ecuación se puede escribir de la siguiente forma:

$$\begin{aligned} 2\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)(y_i - \hat{y}_i) &= 2\sum_{i=1}^{n} \hat{y}_i(y_i - \hat{y}_i) - 2\bar{y}\sum_{i=1}^{n} (y_i - \hat{y}_i) \\ &= 2\sum_{i=1}^{n} \hat{y}_i e_i - 2\hat{y}\sum_{i=1}^{n} e_i = 0 \end{aligned}$$

ya que la suma de los residuales siempre es igual a cero (propiedad 1, Sec. 2.2.2) y la suma de los residuales ponderados por el valor ajustado $\hat{y}_i$ correspondiente también es igual a cero (propiedad 5, Sec. 2.2.2). Por lo anterior,

$$\sum_{i=1}^{n} (y_i - \bar{y}_i)^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.32}$$

El lado izquierdo de la ecuación (2.32) es la suma corregida de cuadrados de las observaciones, SS_T, que mide la variabilidad total en las observaciones. Los dos componentes de SS_T miden, respectivamente, la cantidad de variabilidad en las observaciones y_i explicada por la línea de regresión, y la variación residual que queda sin explicar por la línea de regresión. Se ve que $SS_{Res} = \Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2$ es la **suma de cuadrados de los residuales o**

http://carlos2524.jimdo.com/

la suma de cuadrados de error de la ecuación (2.16). Se acostumbra llamar a $SS_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ la **suma de cuadrados de regresión, o del modelo.**

La ecuación (2.32) es la **identidad fundamental del análisis de varianza para un modelo de regresión**. En forma simbólica, se acostumbra escribir

$$SS_T = SS_R + SS_{Res} \tag{2.33}$$

Si se comparan las ecuaciones (2.33) y (2.18), se ve que la suma de cuadrados de regresión se puede calcular como sigue:

$$SS_R = \hat{\beta}_1 S_{xy} \tag{2.34}$$

La cantidad de **grados de libertad** se determina como sigue. La suma total de cuadrados, SS_T, tiene $df_T = n - 1$ grados de libertad, porque se perdió un grado de libertad como resultado de la restricción $\sum_{i=1}^{n}(y_i - \bar{y})$ para las desviaciones $y_i - \bar{y}$. La suma de cuadrados del modelo, o de la regresión es SS_R y tiene $df_R = 1$ grado de libertad, porque SS_R queda completamente determinado por un parámetro, que es $\hat{\beta}_1$ [véase la Ec. (2.34)]. Por último, antes se dijo que SS_R tiene $df_{Res} = n - 2$ grados de libertad, porque se imponen dos restricciones a las desviaciones $y_i - \hat{y}_i$ como resultado de estimar $\hat{\beta}_0$ y $\hat{\beta}_1$. Obsérvese que los grados de libertad tienen una propiedad aditiva:

$$\begin{aligned} df_T &= df_R + df_{Res} \\ n - 1 &= 1 + (n - 2) \end{aligned} \tag{2.35}$$

Se puede aplicar la prueba F normal del **análisis de varianza** para probar la hipótesis H_0: $\beta_1 = 0$. En el apéndice C.3 se demuestra que: 1) $SS_{Res} = (n - 2)MS_{Res}$ sigue una distribución χ^2_{n-2}; 2) si es cierta la hipótesis nula H_0: $\beta_1 = 0$, entonces SS_R tiene una distribución χ^2_{n-2}; y 3) SS_{Res} y SS_R son independientes. De acuerdo con la definición de una estadística F presentada en el apéndice C.1,

$$F_0 = \frac{SS_R/df_R}{SS_{Res}/df_{Res}} = \frac{SS_R/1}{SS_{Res}/(n-2)} = \frac{MS_R}{MS_{Res}} \tag{2.36}$$

sigue la distribución $F_{1,n-2}$. También se indica en el apéndice C.3 que los valores esperados de estos cuadrados medios son

$$\begin{aligned} E(MS_{Res}) &= \sigma^2 \\ E(MS_R) &= \sigma^2 + \beta_1^2 S_{xx} \end{aligned}$$

Estos cuadrados medios esperados indican que si es grande el valor observado de F_0, es probable que la pendiente $\beta_1 \neq 0$. También se menciona en el apéndice C.3 que si $\beta_1 \neq 0$, entonces F_0 sigue una distribución F no central, con 1 y $n - 2$ grados de libertad, y un parámetro de **no centralidad** λ:

$$\lambda = \frac{\beta_1^2 S_{xx}}{\sigma^2}$$

http://carlos2524.jimdo.com/

TABLA 2.4 Análisis de varianza para probar el significado de la regresión

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0
Regresión	$SS_R = \hat{\beta}_1 S_{xy}$	1	MS_R	MS_R/MS_{Res}
Residual	$SS_{Res} = SS_T - \hat{\beta}_1 S_{xy}$	$n-2$	MS_{Res}	
Total	SS_T	$n-1$		

Este parámetro de no centralidad también indica que el valor observado de F_0 debe ser grande si $\beta_1 \neq 0$. Por consiguiente, para probar la hipótesis H_0: $\beta_1 = 0$, se calcula el estadístico F_0 de prueba y se rechaza H_0 si

$$F_0 > F_{\alpha, 1, n-2}$$

El procedimiento de prueba se resume en la tabla 2.4.

Ejemplo 2.4 *Datos del propelente de reacción*

Se probará el significado de la regresión para el modelo desarrollado en el ejemplo 2.1, de los datos de propelente de cohete. El modelo ajustado es $\hat{y} = 2\ 627.82 - 37.15x$, $SS_T = 1\ 693\ 737.60$, y $S_{xy} = -41\ 112.65$. La suma de cuadrados de regresión se calcula con la ecuación (2.34), como sigue:

$$SS_R = \hat{\beta}_1 S_{xy} = (-37.15)(-41\ 112.65) = 1\ 527\ 334.95$$

El análisis de varianza se resume en la tabla 2.5. El valor calculado de $F_0 = 165.21$, y de acuerdo con la tabla A.4 del apéndice, $F_{0.01, 1, 18} = 8.29$. El valor P para esta prueba es $1.66E{-}10$. Por consiguiente, se rechaza H_0: $\beta_1 = 0$.

TABLA 2.5 Tabla de análisis de varianza para el modelo de regresión del propelente de reacción

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Regresión	1 527 334.95	1	1 527 334.95	165.21	1.66×10^{-10}
Residual	166 402.65	18	9 244.59		
Total	1 693 737.60	19			

Resultados con Minitab

Los resultados de Minitab de la tabla 2.3 también presentan la prueba de análisis de varianza de la significancia de la regresión. Al comparar las tablas 2.3 y 2.5 se ve que hay unas ligeras diferencias entre los cálculos manuales y los que hace la computadora para las sumas de cuadrados. Esto se debe al redondeo de los cálculos manuales a dos cifras decimales. En esencia los valores calculados de las estadísticas de prueba concuerdan.

http://carlos2524.jimdo.com/

Más acerca de la prueba t

Se dijo, en la sección 2.3.2, que el estadístico de prueba

$$t_0 = \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\sqrt{MS_{\text{Res}}/S_{xx}}} \tag{2.37}$$

se podría usar para probar la significancia de la regresión. Sin embargo, nótese que al elevar al cuadrado ambos lados de la ecuación (2.37) se obtiene

$$t_0^2 = \frac{\hat{\beta}_1^2 S_{xx}}{MS_{\text{Res}}} = \frac{\hat{\beta}_1 S_{xy}}{MS_{\text{Res}}} = \frac{MS_{\text{R}}}{MS_{\text{Res}}} \tag{2.38}$$

Así, t_0^2 en la ecuación (2.38) es idéntica a F_0 del método de análisis de varianza en la ecuación (2.36). Una muestra es el ejemplo del propelente de reacción, $t_0 = -12.5$, así que $(t_0^2 = (-12.5)^2 = 165.12 = F_0 = 165.21$. En general, el cuadrado de una variable aleatoria t con f grados de libertad es una variable aleatoria F con uno y f grados de libertad en el numerador y el denominador, respectivamente. Aunque la prueba t para H_0: $\beta_1 = 0$ equivale a la prueba F en la regresión lineal simple, la prueba t es algo más adaptable, porque se podría usar para probar hipótesis alternativas unilaterales (sea H_1: $\beta_1 < 0$ o H_1: $\beta_1 > 0$), mientras que la prueba F sólo considera la alternativa bilateral. Los programas de cómputo para regresión producen, en forma rutinaria, tanto el análisis de varianza de la tabla 2.4, como el estadístico t. Véase los resultados de Minitab en la tabla 2.3.

La utilidad real del análisis de varianza está en los **modelos de regresión múltiple**. En el siguiente capítulo explicaremos esos modelos.

Por último, recuérdese que decidir que $\beta_1 = 0$ es una conclusión muy importante que sólo es **apoyada** por la prueba t o la prueba F. La incapacidad de demostrar que la pendiente no es estadísticamente distinta de cero no necesariamente quiere decir que y y x no están relacionadas. Puede indicar que la capacidad de detectar esta relación se ha confundido por la varianza del proceso de medición, o que el intervalo de valores de x es inadecuado. Se requiere una gran cantidad de evidencia no estadística y conocimiento del problema, para llegar a la conclusión que $\beta_1 = 0$.

2.4 ESTIMACIÓN DE INTERVALO EN LA REGRESIÓN LINEAL SIMPLE

En esta sección se describirá la estimación del intervalo de confianza de los parámetros del modelo de regresión. También se explicará la estimación de intervalo de la respuesta media $E(y)$ para valores dados de x. Se continúan aplicando las hipótesis sobre normalidad que se introdujeron en la sección 2.3.

2.4.1 Intervalos de confianza de β_0, β_1 y σ^2

Además de los estimados puntuales de β_0, β_1 y σ^2, también se pueden obtener estimados de intervalo de confianza para esos parámetros. El ancho de dichos intervalos es una medida de la calidad general de la recta de regresión. Si los errores se distribuyen en forma normal e independiente, entonces la distribución de muestreo tanto de

http://carlos2524.jimdo.com/

$$\frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \quad \text{y} \quad \frac{\hat{\beta}_0 - \beta_0}{\text{se}(\hat{\beta}_0)}$$

es t, con $n - 2$ grados de libertad. Así, un intervalo de confianza de $100(1 - \alpha)$ por ciento para la pendiente β_1 se determina con

$$\hat{\beta}_1 - t_{\alpha/2,\, n-2}\,\text{se}(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2,\, n-2}\,\text{se}(\hat{\beta}_1) \tag{2.39}$$

y un intervalo de confianza de $100(1 - \alpha)$ por ciento para la ordenada al origen β_0 es

$$\hat{\beta}_0 - t_{\alpha/2,\, n-2}\,\text{se}(\hat{\beta}_0) \le \beta_0 \le \hat{\beta}_0 + t_{\alpha/2,\, n-2}\,\text{se}(\hat{\beta}_0) \tag{2.40}$$

Estos intervalos de confianza tienen la interpretación usual de frecuencia, por lo tanto, si hubiera que tomar muestras repetidas del mismo tamaño a los mismos valores de x, y formar, por ejemplo, intervalos de confianza de 95% de la pendiente para cada muestra, entonces el 95% de esos intervalos contendrán el verdadero valor de β_1.

Si los errores están distribuidos en forma normal e independiente, el apéndice C.3 demuestra que la distribución de muestreo de

$$(n - 2)MS_{\text{Res}}/\sigma^2$$

es ji cuadrada, con $n - 2$ grados de libertad. Así,

$$P\left\{\chi^2_{1-\alpha/2,\, n-2} \le \frac{(n-2)MS_{\text{Res}}}{\sigma^2} \le \chi^2_{\alpha/2,\, n-2}\right\} = 1 - \alpha$$

y en consecuencia, un intervalo de confianza de $100(1 - \alpha)$ por ciento para σ^2 es

$$\frac{(n-2)\,MS_{\text{Res}}}{\chi^2_{\alpha/2,\, n-2}} \le \sigma^2 \le \frac{(n-2)\,MS_{\text{Res}}}{\chi^2_{1-\alpha/2,\, n-2}} \tag{2.41}$$

Ejemplo 2.5 *Datos del propelente de reacción*

Se establecerán los intervalos de confianza de 95% para β_1 y σ^2 con los datos del propelente de cohetes del ejemplo 2.1. El error estándar de $\hat{\beta}_1$ es $\text{se}(\hat{\beta}_1) = 2.89$, y $t_{0.025,\,18} = 2.101$. Por consiguiente, de acuerdo con la ecuación (2.35), el intervalo de confianza de 95% para la pendiente es

$$\hat{\beta}_1 - t_{0.025,\,18}\,\text{se}(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{0.025,\,18}\,\text{se}(\hat{\beta}_1)$$

$$-37.15 - (2.101)(2.89) \le \beta_1 \le -37.15 + (2.101)(2.89)$$

o sea

$$-43.22 \le \beta_1 \le -31.08$$

En otras palabras, el 95% de esos intervalos incluirán el valor verdadero de la pendiente.

http://carlos2524.jimdo.com/

Si se hubiera escogido un valor distinto de α, el ancho del intervalo de confianza resultante hubiera sido distinto. Por ejemplo, el intervalo de confianza de 90% para β_1 es $-42.16 \leq \beta_1 \leq 32.14$, más angosto que el intervalo de confianza de 95%. El intervalo de confianza de 99% es $-45.49 \leq \beta_1 \leq 28.81$, que es más ancho que el intervalo de confianza de 95%. En general, cuando más grande sea el coeficiente de confianza $(1 - \alpha)$, el intervalo de confianza será mayor.

El intervalo de confianza de 95% para σ^2 se determina a partir de la ecuación (2.41), como sigue:

$$\frac{(n-2)MS_{\text{Res}}}{\chi^2_{0.025,\,n-2}} \leq \sigma^2 \leq \frac{(n-2)MS_{\text{Res}}}{\chi^2_{0.975,\,n-2}}$$

$$\frac{18(9\,244.59)}{\chi^2_{0.025,\,18}} \leq \sigma^2 \leq \frac{18(9\,244.59)}{\chi^2_{0.975,\,18}}$$

En la tabla A.2 del apéndice, $\chi^2_{0.025,\,18} = 31.5$, y $\chi^2_{0.975,\,18} = 8.23$. En consecuencia, el intervalo deseado de confianza resulta

$$\frac{18(9\,244.59)}{31.5} \leq \sigma^2 \leq \frac{18(9\,244.59)}{8.23}$$

es decir,

$$5282.62 \leq \sigma^2 \leq 20{,}219.03$$

2.4.2 Estimación de intervalos de la respuesta media

Una aplicación importante de un modelo de regresión es estimar la respuesta media, $E(y)$, para determinado valor de la variable regresora x. Por ejemplo, se podría tratar de estimar la resistencia promedio al corte de la adhesión de propelente en un motor de reacción hecho con un lote de propelente de sustentación que tiene 10 semanas de edad. Sea x_0 el valor, o "nivel", de la variable regresora para el que se desea estimar la respuesta media, es decir, $E(y|x_0)$. Se supone que x_0 es cualquier valor de la variable regresora dentro del intervalo de los datos originales de x que se usaron para ajustar el modelo. Un estimador insesgado de $E(y|x_0)$ se determina a partir del modelo ajustado como sigue:

$$\widehat{E(y|x_0)} = \hat{\mu}_{y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0 \tag{2.42}$$

Para obtener un intervalo de confianza de $100(1 - \alpha)$ por ciento para $E(y|x_0)$, se debe notar primero que $\hat{\mu}_{y|x_0}$ es una variable aleatoria normalmente distribuida, porque es una combinación lineal de las observaciones y_i. La varianza de $\hat{\mu}_{y|x_0}$ es

$$\begin{aligned}\text{Var}(\hat{\mu}_{y|x_0}) &= \text{Var}(\hat{\beta}_0 + \hat{\beta}_1 x_0) = \text{Var}[\bar{y} + \hat{\beta}_1(x_0 - x)] \\ &= \frac{\sigma^2}{n} + \frac{\sigma^2(x_0 - \bar{x})^2}{S_{xx}} = \sigma^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]\end{aligned}$$

http://carlos2524.jimdo.com/

porque, como se hizo notar en la sección 2.2.4, $\text{Cov}(\bar{y}, \hat{\beta}_1) = 0$. Así, la distribución de muestreo de

$$\frac{\hat{\mu}_{y|x_0} - E(y|x_0)}{\sqrt{MS_{\text{Res}}\left(1/n + (x_0 - \bar{x})^2 / S_{xx}\right)}}$$

es t, con $n - 2$ grados de libertad. Por lo anterior, **un intervalo de confianza de 100(1 – α) por ciento para la respuesta media en el punto $x = x_0$** es

$$\hat{\mu}_{y|x_0} - t_{\alpha/2,\, n-2}\sqrt{MS_{\text{Res}}\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

$$\leq E(y|x_0) \leq \hat{\mu}_{y|x_0} + t_{\alpha/2,\, n-2}\sqrt{MS_{\text{Res}}\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

Nótese que el ancho del intervalo de confianza para $E(y|x_0)$ es una función de x_0. El ancho del intervalo es un mínimo para $x_0 = \bar{x}$, y crece a medida que aumenta $|x_0 - \bar{x}|$. En forma intuitiva eso es razonable, porque cabría esperar que las mejores estimaciones de y se hacen con valores de x cerca del centro de los datos, y que la precisión de la estimación se redujera al moverse hacia la frontera del espacio de x.

Ejemplo 2.6 *Datos del propelente de reacción*

Se trata de determinar un intervalo de confianza de 95% para $E(y|x_0)$ con los datos del propelente de reacción del ejemplo 2.1. Ese intervalo de confianza se calcula con la ecuación (2.43) como sigue:

$$\hat{\mu}_{y|x_0} - t_{\alpha/2,\, n-2}\sqrt{MS_{\text{Res}}\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)} \leq E(y|x_0)$$

$$\leq \hat{\mu}_{y|x_0} + t_{\alpha/2,\, n-2}\sqrt{MS_{\text{Res}}\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

$$\hat{\mu}_{y|x_0} - (2.101)\sqrt{9\,244.59\left(\frac{1}{20} + \frac{(x_0 - 13.3625)^2}{1\,106.56}\right)} \leq E(y|x_0)$$

$$\leq \hat{\mu}_{y|x_0} + (2.101)\sqrt{9\,244.59\left(\frac{1}{20} + \frac{(x_0 - 13.3625)^2}{1\,106.56}\right)}$$

Si se sustituyen los valores de x_0 y el valor ajustado $\hat{y}_0 = \hat{\mu}_{y|x_0}$ en el valor de x_0 en esta última ecuación, se obtendrá el intervalo de confianza de 95% para la respuesta media en $x = x_0$. Por ejemplo, si $x_0 = \bar{x} = 13.3625$, entonces $\hat{\mu}_{y|x_0} = \hat{y}_0 = 2\,131.40$, y el intervalo de confianza se vuelve

$$2\,086.230 \leq E(y \mid 13.3625) \leq 2\,176.571$$

http://carlos2524.jimdo.com/

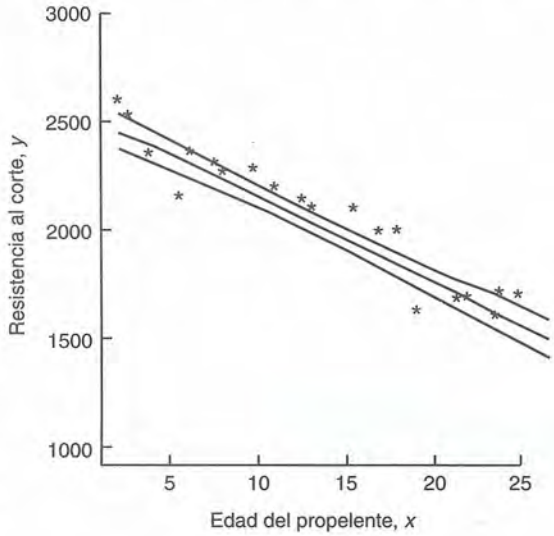


Figura 2.4 Límites superior e inferior de 95% de confianza para los datos del propelente.

La tabla 2.6 contiene los límites de confianza de $E(y \mid x_0)$ para otros valores de x_0. Esos límites de confianza se muestran gráficamente en la figura 2.4. Nótese que el ancho del intervalo de confianza aumenta a medida que aumenta $|x_0 - \bar{x}|$.

TABLA 2.6 Límites de confianza de $E(y \mid x_0)$ para algunos valores de x_0

Límite inferior de confianza	x_0	Límite superior de confianza
2438.919	3	2593.821
2341.360	6	2468.481
2241.104	9	2345.836
2136.098	12	2227.942
2086.230	$\bar{x} = 13.3625$	2176.571
2024.318	15	2116.822
1905.890	18	2012.351
1782.928	21	1912.412
1657.395	24	1815.045

En muchos textos de regresión se dice que nunca se debería usar un modelo de regresión para **extrapolar** fuera del intervalo de los datos originales. Se entiende por extrapolación el uso de la ecuación de predicción fuera de los límites del espacio de x. La figura 1.5 muestra con claridad los peligros inherentes a la extrapolación; el error del modelo o la ecuación pueden dañar mucho la predicción.

La ecuación (2.43) hace destacar que el asunto de la extrapolación es mucho más sutil; mientras más alejado está el valor de x del centro de los datos, el estimado $E(y \mid x_0)$ es más variable. Sin embargo, téngase en cuenta que no hay nada "mágico" en la frontera del espacio de x. No es irrazonable suponer que la predicción es espléndida en el valor de datos observados más alejado del centro, y que es totalmente mala justo fuera del límite. Es claro que la ecuación (2.43) indica que uno se debe ocupar de la calidad de predicción al acercarse al límite, y que al salir de ese límite la predicción puede empeorar con rapidez. Además, mientras más se aleje uno de la región original del espacio x, es más probable que el error de la ecuación o del modelo juegue un papel mayor en el proceso.

http://carlos2524.jimdo.com/

Eso no es lo mismo que decir "nunca extrapolar". Los ingenieros y los economistas usan en forma rutinaria las ecuaciones de predicción para pronosticar una variable de interés en uno o más periodos en el futuro. Hablando estrictamente, este pronóstico es una extrapolación. La ecuación (2.43) respalda ese uso de la ecuación de predicción. Sin embargo, esa ecuación (2.43) no respalda el uso del modelo de regresión para pronosticar muchos periodos hacia el futuro. En general, mientras más grande sea la extrapolación, es mayor la probabilidad de que un error de ecuación o un error de modelo tengan gran impacto sobre los resultados.

La declaración de la probabilidad asociada con el intervalo de confianza (2.43) sólo es válida cuando se debe construir un solo intervalo de confianza para la respuesta media. Un procedimiento para formar varios intervalos de confianza que, al considerarlos en conjunto, tengan un nivel especificado de confianza, es un problema de **inferencia estadística simultánea**. Esos problemas se discutirán en el capítulo 3.

2.5 PREDICCIÓN DE NUEVAS OBSERVACIONES

Una aplicación importante del modelo de regresión es predecir nuevas observaciones y que correspondan a un nivel especificado de la variable regresora x. Si x_0 es el valor de interés de la variable regresora, entonces

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \tag{2.44}$$

es el estimado puntual del nuevo valor de la respuesta y_0.

A continuación se obtendrá un estimado de intervalo para esta observación futura y_0. El intervalo de confianza para la respuesta media en $x = x_0$ [Ec. (2.43)] es inadecuado para este problema, porque es un estimado de intervalo para la **media** de y, un parámetro, y no es una declaración de probabilidad sobre futuras observaciones a partir de esa distribución. Se desarrollará un **intervalo de predicción para la observación futura y_0**.

Nótese que la variable aleatoria

$$\psi = y_0 - \hat{y}_0$$

tiene distribución normal, con media 0 y varianza

$$\mathrm{Var}(\psi) = \mathrm{Var}(y_0 - \hat{y}_0) = \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]$$

porque la observación futura y_0 es independiente de $\hat{y}_0$. Si se usa $\hat{y}_0$ para predecir y_0, entonces el error estándar de $\psi = y_0 - \hat{y}_0$ es el estadístico adecuado sobre el cual basar un intervalo de predicción. Así, el intervalo de predicción de $100(1-\alpha)\%$ de confianza para una observación futura en x_0 es

$$\hat{y}_0 - t_{\alpha/2,\,n-2}\sqrt{MS_{\mathrm{Res}}\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)} \le y_0$$

$$\le \hat{y}_0 + t_{\alpha/2,\,n-2}\sqrt{MS_{\mathrm{Res}}\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)} \tag{2.45}$$

http://carlos2524.jimdo.com/

El intervalo de predicción de la ecuación (2.45) es de ancho mínimo en $x_0 = -x$ y se ensancha a medida que aumenta $|x_0 - \bar{x}|$. Al comparar las ecuaciones (2.45) y (2.43) se observa que el intervalo de predicción en x_0 siempre es más ancho que el intervalo de confianza en x_0, porque el intervalo de predicción depende tanto del error del modelo ajustado como del error asociado con observaciones futuras.

Ejemplo 2.7 *Datos del propelente de reacción*

Se determinará un intervalo de predicción de 95% para el valor futuro de la resistencia al corte del propelente, en un motor fabricado con un lote de propelente que tiene 10 semanas de edad. Al aplicar la ecuación (2.45) se encuentra que el intervalo de predicción es

$$\hat{y}_0 - t_{\alpha/2,\,n-2}\sqrt{MS_{\text{Res}}\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)} \le y_0$$

$$\le \hat{y}_0 + t_{\alpha/2,\,n-2}\sqrt{MS_{\text{Res}}\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

$$2\,256.32 - (2.101)\sqrt{9\,244.59\left(1 + \frac{1}{20} + \frac{(10 - 13.3625)^2}{1\,106.56}\right)} \le y_0$$

$$\le 2\,256.32 + (2.101)\sqrt{9\,244.59\left(1 + \frac{1}{20} + \frac{(10 - 13.3625)^2}{1\,106.56}\right)}$$

que se simplifica a

$$2\,048.32 \le y_0 \le 2\,464.32$$

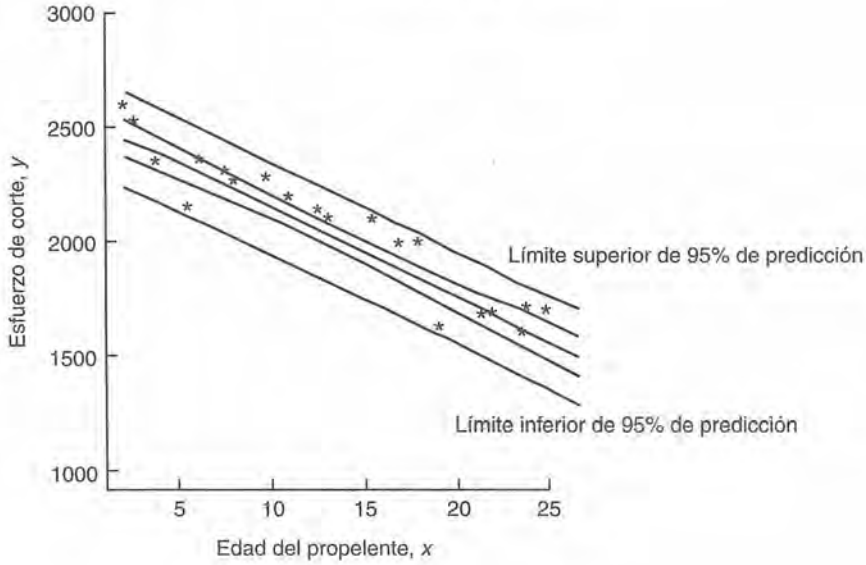


Figura 2.5 Intervalos de 95% de confianza y de predicción para los datos del propilente.

http://carlos2524.jimdo.com/

Así, cabría esperar que un motor nuevo hecho con un lote de propelente de sustentación de 10 semanas de antigüedad tuviera una resistencia al corte entre 2 048.32 y 2 464.32 psi.

La figura 2.5 muestra el 95% del intervalo de predicción calculado con la ecuación (2.45), para el modelo de regresión del propelente de reacción. También en esa gráfica se indica el intervalo de confianza de 95% para la media, esto es $E(y \mid x)$, de acuerdo con la ecuación (2.43). Esta gráfica ilustra muy bien que el intervalo de predicción es más ancho que el intervalo correspondiente de confianza.

Se puede generalizar algo la ecuación (2.45) para determinar un intervalo de predicción de $100(1-\alpha)$ por ciento para la **media** de m observaciones futuras de la respuesta en $x = x_0$. Sea $\bar{y}_0$ la media de m observaciones futuras en $x = x_0$. Un estimador puntual de $\bar{y}_0$ es $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$. El intervalo de predicción de $100(1-\alpha)$ por ciento para $\bar{y}_0$ es

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{MS_{\text{Res}}\left(\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)} \le \bar{y}_0$$

$$\le \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{MS_{\text{Res}}\left(\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)} \tag{2.46}$$

2.6 COEFICIENTE DE DETERMINACIÓN

La cantidad

$$R^2 = \frac{SS_{\text{R}}}{SS_{\text{T}}} = 1 - \frac{SS_{\text{Res}}}{SS_{\text{T}}} \tag{2.47}$$

se llama **coeficiente de determinación**. Como SS_{T} es una medida de la variabilidad de y sin considerar el efecto de la variable regresora x y SS_{Res} es una medida de la variabilidad de y que queda después de haber tenido en consideración a x, R^2 se llama, con frecuencia, la proporción de la variación explicada por el regresor x. Ya que $0 \le SS_{\text{Res}} \le SS_{\text{T}}$, entonces $0 \le R^2 \le 1$. Los valores de R^2 cercanos a 1 implican que la mayor parte de la variabilidad de y está explicada por el modelo de regresión. Para el modelo de regresión con los datos del propelente de reacción, en el ejemplo 2.1, sucede que

$$R^2 = \frac{SS_{\text{R}}}{SS_{\text{T}}} = \frac{1\,527\,334.95}{1\,693\,737.60} = 0.9018$$

por tanto, el 90.18% de la variabilidad de la resistencia queda explicada por el modelo de regresión.

El estadístico R^2 se debe usar con precaución, porque siempre es posible conseguir que R^2 sea grande agregando términos suficientes al modelo. Por ejemplo, si no hay puntos repetidos (más de un valor de y con el mismo valor de x), un polinomio de grado $n-1$ producirá un ajuste "perfecto", con $R^2 = 1$, de los n puntos de datos. Cuando hay puntos repetidos, R^2 nunca puede ser exactamente igual a 1, porque el modelo no puede explicar la variabilidad relacionada con el error "puro".

http://carlos2524.jimdo.com/

Aunque R^2 no puede disminuir si se agrega al modelo una nueva variable regresora, eso no quiere decir necesariamente que el nuevo modelo sea mejor que el anterior. A menos que la suma de cuadrados en el nuevo modelo se reduzca en una cantidad igual al error cuadrático medio original, el nuevo modelo tendrá un mayor error cuadrático medio que el anterior, por la pérdida de un grado de libertad en el error. Así, en realidad el nuevo modelo será peor que el anterior.

La magnitud de R^2 también depende del intervalo de variabilidad de la variable regresora. En general, R^2 aumenta a medida que aumenta la dispersión de las x y disminuye cuando disminuye la dispersión de las x, siempre y cuando sea correcta la forma supuesta del modelo. Hahn [1973] observa que el valor esperado de R^2 en una regresión rectilínea es, aproximadamente,

$$E(R^2) = \frac{\hat{\beta}_1^2 S_{xx}}{\hat{\beta}_1^2 S_{xx} + \sigma^2}$$

Es claro que el valor esperado de R^2 aumentará (disminuirá) cuando aumente (o disminuya) S_{xx}, que es una medida de la dispersión de las x. Así, un valor grande de R^2 puede ser tan sólo el resultado de que x se haya variado en forma no realista dentro de un intervalo grande. Por otro lado, R^2 puede ser pequeña porque el intervalo de las x sea demasiado pequeño como para permitir detectar su relación con y.

Hay otras ideas erróneas acerca de R^2. En general, R^2 no mide la magnitud de la pendiente de la línea de regresión. Un valor grande de R^2 no implica que la pendiente sea grande, además, R^2 no mide la adecuación del modelo lineal, porque con frecuencia R^2 es grande aunque x y y no tengan relación lineal. Por ejemplo, R^2 de la ecuación de regresión en la figura 2.3b es relativamente grande, aunque el comportamiento lineal sea pobre. Recuérdese que aunque R^2 es grande, eso no necesariamente implica que el modelo de regresión sea un predictor exacto.

2.7 ALGUNAS CONSIDERACIONES SOBRE EL USO DE LA REGRESIÓN

El análisis de regresión se usa mucho y, desafortunadamente, se usa mal. Hay varios abusos comunes de la regresión, que se deben mencionar:

1. Los modelos de regresión tienen por objeto servir de ecuaciones de interpolación dentro del intervalo de la o las variables regresoras que se usan para ajustarlos. Como se dijo antes, se debe tener cuidado al extrapolar fuera de ese intervalo. Véase la figura 1.5.
2. La disposición de los valores de x juega un papel importante en el ajuste por mínimos cuadrados. Si bien todos los puntos tienen peso igual para determinar la altura de la recta, la pendiente queda más influenciada por los valores remotos de x. Como ejemplo, véanse los datos de la figura 2.6. La pendiente, en el ajuste por mínimos cuadrados, depende mucho de uno o ambos puntos A y B, así como los datos restantes producirían una estimación muy diferente de la pendiente, si se omitieran A y B. Casos como éste requieren con frecuencia acciones correctivas, como por ejemplo, análisis más detallado y omisión posible de los puntos atípicos, estimación de los parámetros del modelo con alguna técnica que sea influida con menos intensidad por esos puntos que los mínimos cuadrados, o reestructurar el modelo, posiblemente introduciendo más regresores.

http://carlos2524.jimdo.com/

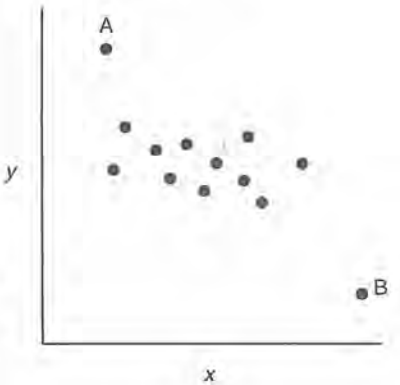


Figura 2.6 Dos observaciones influyentes.

En la figura 2.7 se ve un caso algo diferente, en el que una de las 18 observaciones está muy alejada en el espacio de x. En este ejemplo, la pendiente queda muy determinada por el punto extremo. Si se omite ese punto, es probable que el estimado de la pendiente sea cero. Debido al espacio entre los dos grupos de puntos, en realidad sólo se tienen dos unidades de información con las cuales ajustar el modelo. Así, en realidad hay mucho menos grados de libertad para el error que los 16 aparentes.

Casos como éstos parecen presentarse con bastante frecuencia en la práctica. En general, se debe uno percatar que en algunos conjuntos de datos un solo punto (o un congromerado pequeño de puntos), pueden controlar las propiedades clave del modelo.

3. Los valores atípicos o los valores erróneos pueden afectar el ajuste por mínimos cuadrados. Por ejemplo, se tienen los datos de la figura 2.8. Parece que la observación A es un dato atípico o "malo", porque queda lejos de la recta definida por el resto de los datos. Si en realidad ese punto es atípico, el estimado de la ordenada al origen puede ser incorrecto, y el cuadrado medio residual puede ser un valor inflado de σ^2. Por otro lado, los datos de ese punto pueden ser correctos, y pueden ser una evidencia muy útil acerca del proceso que se investiga. En el capítulo 4 se explican con más detalle los métodos para detectar y manejar los valores atípicos.

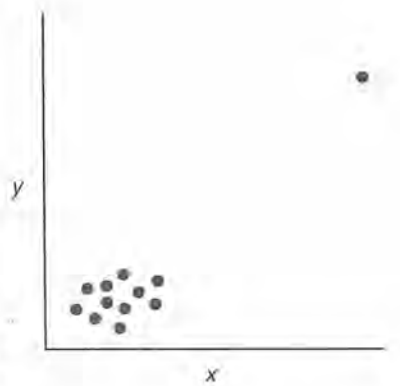


Figura 2.7 Punto remoto en el espacio x

http://carlos2524.jimdo.com/

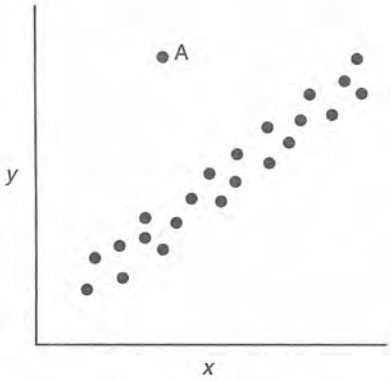


Figura 2.8 Un valor atípico.

4. Como se dijo en el capítulo 1, en el que tan sólo porque un análisis de regresión haya indicado que hay una fuerte relación entre dos variables, no implica que éstas tengan relación alguna en el sentido causal, la causalidad implica, necesariamente, la correlación. El análisis de regresión sólo puede examinar los asuntos de regresión. No puede manejar el asunto de la necesidad. Así, las expectativas de descubrir relaciones de causa-efecto por medio de la regresión no deben ser muchas.

   Como ejemplo de una relación "sin sentido" entre dos variables, se presentan los datos de la tabla 2.7. En ella se ve la cantidad de débiles mentales en el Reino Unido por 10 000 personas de la población estimada (y), la cantidad de licencias de radiorrecepción otorgadas (x_1) y el primer nombre del presidente de Estados Unidos (x_2), durante los años 1924 a 1937. Se puede demostrar que la ecuación de regresión que relaciona a y con x_1 es

$$\hat{y} = 4.582 + 2.204x_1$$

TABLA 2.7 Datos para ilustrar relaciones absurdas entre variables

Año	Cantidad de débiles mentales certificados por 10 000 de población estimada en el Reino Unido (y)	Cantidad de licencias de receptor de radio otorgadas (millones) en el Reino Unido (x_1)	Primer nombre del presidente de Estados Unidos (x_2)
1924	8	1.350	Calvin
1925	8	1.960	Calvin
1926	9	2.270	Calvin
1927	10	2.483	Calvin
1928	11	2.730	Calvin
1929	11	3.091	Calvin
1930	12	3.647	Herbert
1931	16	4.620	Herbert
1932	18	5.497	Herbert
1933	19	6.260	Herbert
1934	20	7.012	Franklin
1935	21	7.618	Franklin
1936	22	8.131	Franklin
1937	23	8.593	Franklin

Fuente: Kendall y Yule [1950] y Tufte [1974].

http://carlos2524.jimdo.com/

El estadístico t para estimar H_0: $\beta_1 = 0$ para este modelo es $t_0 = 27.312$ (el valor P es 3.58 E–12), y el coeficiente de determinación es $R^2 = 0.9842$. Esto es, el 98.42% de la variabilidad de los datos queda explicado por la cantidad de licencias de radiorrecepción otorgadas. Es claro que se trata de una relación absurda, y es muy improbable que la cantidad de débiles mentales tenga una relación funcional con la cantidad de licencias de radiorrecepción otorgadas. La razón de esta fuerte relación estadística es que y y x_1 tienen relación monótona (dos series de números tienen relación monótona si, por ejemplo, al aumentar una siempre aumenta o disminuye la otra). En este ejemplo, y es creciente porque se refinan más los procedimientos para diagnosticar las afecciones mentales, al paso de los años durante el estudio, y x_1 aumenta por la disponibilidad y bajo costo de la tecnología de la radio, también durante esos años.

Dos secuencias cualesquiera de números, relacionadas monótonamente, tendrán propiedades semejantes. Para ilustrar más este ejemplo, supóngase que se hace la regresión y de la cantidad de letras en el nombre del presidente de Estados Unidos en el año correspondiente. El modelo es

$$\hat{y} = -26.442 + 5.900x_2,$$

con $t_0 = 8.996$ (el valor de P es 1.11 E-6) y $R^2 = 0.8709$. Es claro que también se trata de una relación absurda.

5. En algunas aplicaciones de la regresión, se desconoce el valor de la variable regresora x necesario para determinar y. Por ejemplo, supóngase que se trata de predecir la carga diaria máxima en un sistema de generación eléctrica a partir de un modelo de regresión que relacione la carga con la temperatura diaria máxima. Para pronosticar la carga máxima de mañana primero habrá que predecir la temperatura máxima de mañana. En consecuencia, la predicción de la carga máxima es **condicional** respecto al pronóstico de temperatura. La exactitud del pronóstico de carga máxima depende de la exactitud del pronóstico de temperatura. Esto debe tenerse en cuenta al evaluar el desempeño del modelo.

En los capítulos siguientes se describirán otros abusos con la regresión. Para conocer más acerca de este tema, véase el artículo de Box (1966).

2.8 REGRESIÓN POR EL ORIGEN

Algunos casos de regresión parecen implicar que una recta que pase por el origen debe ajustarse a los datos. Con frecuencia, parece adecuado un **modelo de regresión sin ordenada al origen** para analizar datos de procesos químicos y de manufactura. Por ejemplo, el rendimiento de un proceso químico es cero cuando la temperatura de operación del proceso es cero.

El modelo sin ordenada al origen es

$$y = \beta_1 x + \varepsilon \tag{2.48}$$

Dadas n observaciones (y_i, x_i), $i = 1, 2, \ldots, n$, la función de mínimos cuadrados es

$$S(\beta_1) = \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2$$

http://carlos2524.jimdo.com/

La única ecuación normal es

$$\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \tag{2.49}$$

y el **estimador de la pendiente por mínimos cuadrados** es

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \tag{2.50}$$

El estimador $\hat{\beta}_1$ es insesgado para β_1, y el **modelo de regresión ajustado** es

$$\hat{y} = \hat{\beta}_1 x \tag{2.51}$$

El estimador de σ^2 es

$$\hat{\sigma}^2 \equiv MS_{\text{Res}} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - \hat{\beta}_1 \sum_{i=1}^{n} y_i x_i}{n-1} \tag{2.52}$$

con $n - 1$ grados de libertad.

Con la hipótesis de normalidad de los errores, se pueden probar hipótesis y establecer intervalos de confianza y de predicción para el modelo sin ordenada al origen. El **intervalo de confianza de 100(1 – α) por ciento para β_1** es

$$\hat{\beta}_1 - t_{\alpha/2,n-1}\sqrt{\frac{MS_{\text{Res}}}{\sum_{i=1}^{n} x_i^2}} \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2,n-1}\sqrt{\frac{MS_{\text{Res}}}{\sum_{i=1}^{n} x_i^2}} \tag{2.53}$$

Un **intervalo de confianza de 100(1 – α) por ciento** para $E(y|x_0)$, la respuesta media en $x = x_0$, es

$$\hat{\mu}_{y|x_0} - t_{\alpha/2,n-1}\sqrt{\frac{x_0^2\, MS_{\text{Res}}}{\sum_{i=1}^{n} x_i^2}} \le E(y|x_0) \le \hat{\mu}_{y|x_0} + t_{\alpha/2,n-1}\sqrt{\frac{x_0^2\, MS_{\text{Res}}}{\sum_{i=1}^{n} x_i^2}} \tag{2.54}$$

El **intervalo de predicción de 100(1 – α) por ciento para una observación futura en $x = x_0$**, por ejemplo y_0, es

$$\hat{y}_0 - t_{\alpha/2,n-1}\sqrt{MS_{\text{Res}}\left(1 + \frac{x_0^2}{\sum_{i=1}^{n} x_i^2}\right)}$$

$$\le y_0 \le \hat{y}_0 + t_{\alpha/2,n-1}\sqrt{MS_{\text{Res}}\left(1 + \frac{x_0^2}{\sum_{i=1}^{n} x_i^2}\right)} \tag{2.55}$$

http://carlos2524.jimdo.com/

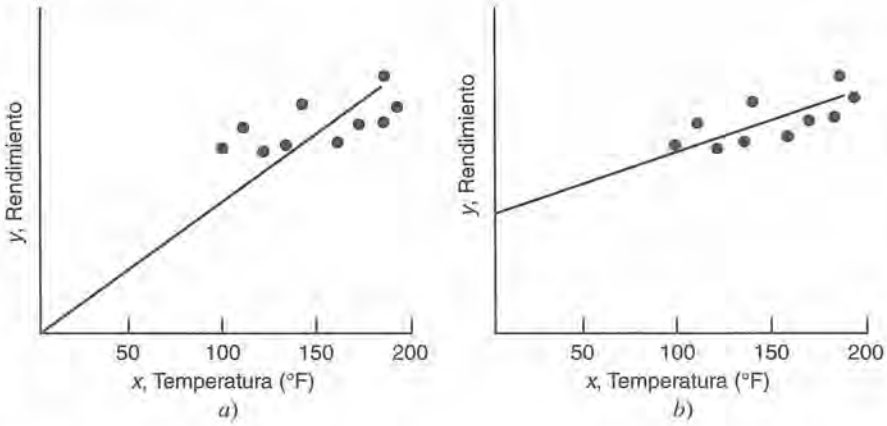


Figura 2.9 Diagramas de dispensión y lineas de regresión para el rendimiento y la temperatura de operación en un proceso químico: *a*) modelo ordenada al origen; *b*) modelo con ordenada al origen.

El intervalo de confianza (2.54) y el intervalo de predicción (2.55) se ensanchan al aumentar x_0. Además, la longitud del intervalo de confianza (2.54) en $x = 0$ es cero, porque el modelo supone conocido con certidumbre que la media de y es cero cuando $x = 0$. Este comportamiento es muy distinto del observado en el modelo con ordenada al origen. El intervalo de predicción (2.55) tiene longitud distinta de cero en $x_0 = 0$, porque se debe tener en cuenta el error aleatorio en observaciones futuras.

Es relativamente fácil usar mal el modelo sin ordenada al origen, en especial en casos donde los datos están en una región del espacio x alejada del origen. Por ejemplo, examinemos el ajuste sin ordenada al origen para el diagrama de dispersión de rendimiento de un proceso químico (y) y la temperatura de operación (x), en la figura 2.9a. Aunque dentro del intervalo de $100°F \leq x \leq 200°F$ de la variable regresora parece que el rendimiento y la temperatura parecen tener relación lineal, al forzar al modelo a que pase por el origen se produce un mal ajuste apreciable. Un modelo con ordenada al origen, como el de la figura 2.9b, tiene un ajuste mucho mejor, en la región del espacio de x donde se recolectaron los datos.

Con frecuencia, la relación entre y y x es muy distinta cerca del origen que en la región del espacio de x que contiene los datos. Esto se ilustra en la figura 2.10, para los datos del proceso químico. En este caso parecería que se necesita un modelo cuadrático, o uno no lineal complicado, para expresar la relación entre y y x en todo el intervalo de valores de x. Ese modelo sólo se manejaría si el margen de x en los datos está suficientemente cercano al origen.

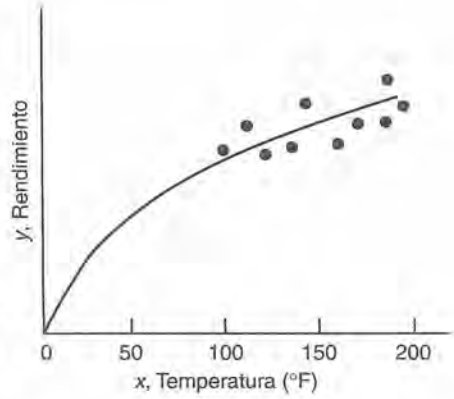


Figura 2.10 Relación real entre el rendimiento y la temperatura.

http://carlos2524.jimdo.com/

A veces, el diagrama de dispersión proporciona una guía para decidir si se ajusta o no el modelo sin ordenada al origen. También, se pueden ajustar ambos modelos, y escoger entre ellos de acuerdo con la calidad del ajuste obtenido. Si no se puede rechazar la hipótesis $\beta_0 = 0$ en el modelo sin ordenada al origen, quiere decir que se puede mejorar el ajuste si se usa ese modelo. El cuadrado medio de residuales es una forma útil de comparar la calidad del ajuste. El modelo que tenga el cuadrado medio residual menor es el mejor ajuste, en el sentido que minimiza el estimado de la varianza de y respecto a la línea de regresión.

En general, R^2 no es un buen estadístico para comparar los dos modelos. Para el modelo con ordenada al origen,

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\text{variación de } y \text{ explicada por la regresión}}{\text{variación total observada en } y}$$

Nótese que R^2 indica la proporción de variabilidad respecto a $\bar{y}$ explicada por la regresión. En el caso sin ordenada al origen, la identidad fundamental del análisis de varianza, ecuación (2.32), se transforma en

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} \hat{y}_i^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

por lo que el análogo de R^2 en el modelo sin ordenada al origen sería

$$R_0^2 = \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2}$$

El estadístico R_0^2 indica la proporción de variabilidad respecto al **origen** (cero) que explica la regresión. A veces se encuentra que R_0^2 es mayor que R^2, aun cuando el cuadrado medio residual (que es una medida razonable de la calidad general del ajuste) para el modelo con ordenada al origen es menor que el cuadrado medio residual para el modelo sin ordenada al origen. Esto se debe a que R_0^2 se calcula con valores de sumas de cuadrados no corregidas.

Hay otras formas de definir R^2 para el modelo sin ordenada al origen. Una posibilidad es

$$R_0^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Sin embargo, en los casos donde $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ es grande, R_0^2 puede ser negativa. Se prefiere usar MS_{Res} como base de comparación entre los modelos con y sin ordenada al origen. Un buen artículo sobre modelos de regresión sin ordenada al origen es el de Hahn [1979].

http://carlos2524.jimdo.com/

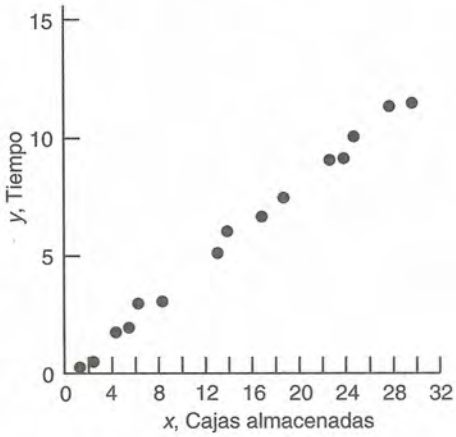


Figura 2.11 Diagrama de dispersión de los datos de almacenamiento.

Ejemplo 2.8 *Datos de almacenamiento en la vitrina*

El tiempo necesario para que un vendedor abastezca de sodas una vitrina en una tienda, y las cajas de producto abastecido se ven en la tabla 2.8. El diagrama de dispersión está en la figura 2.11, y parece indicar que se podría usar una recta que pase por el origen, para expresar la relación entre el tiempo y la cantidad de cajas abastecidas, ya que si la cantidad de cajas $x = 0$, el tiempo de abastecimiento $y = 0$, este modelo parece razonable a la intuición. Nótese también que el margen de x se acerca al origen.

La pendiente, en el modelo sin ordenada al origen, se calcula con la ecuación (2.50):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} = \frac{1\,841.98}{4\,575.00} = 0.4026$$

TABLA 2.8 Datos de abastecimiento de máquina para el ejemplo 2.8

Tiempo y (minutos)	Cajas almacenadas x
10.15	25
2.96	6
3.00	8
6.88	17
0.28	2
5.06	13
9.14	23
11.86	30
11.69	28
6.04	14
7.57	19
1.74	4
9.38	24
0.16	1
1.84	5

http://carlos2524.jimdo.com/

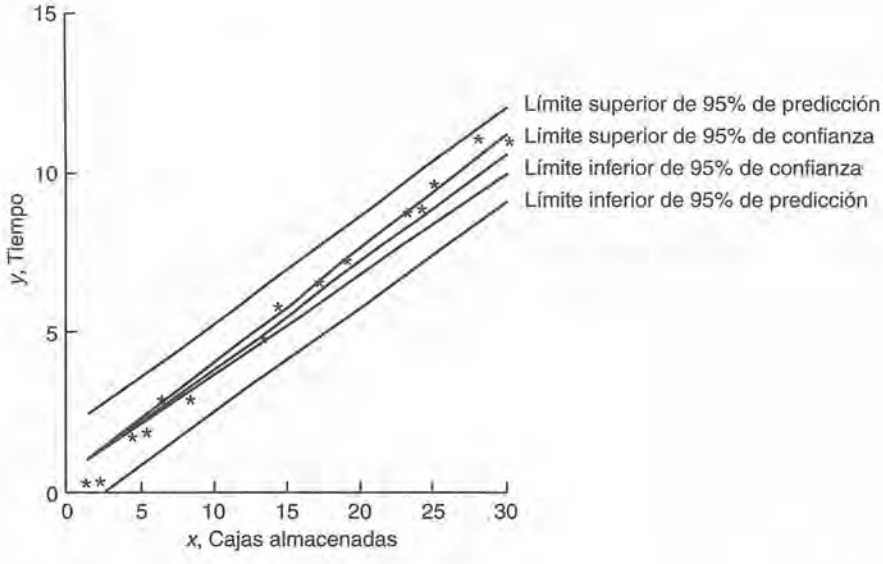


Figura 2.12 Bandas de confianza y de predicción para los datos de almacenamiento.

Por consiguiente, la ecuación ajustada es

$$\hat{y} = 0.4026x$$

Esta recta de regresión se muestra en la figura 2.12. El cuadrado medio residual para este modelo es $MS_{\text{Res}} = 0.0893$ y $R_0^2 = 0.9883$, así como, el estadístico t para probar H_0: $\beta_1 = 0$ es $t_0 = 91.13$, y con ella el valor P es 8.02 E-21. Estos estadísticos de resumen no indican alguna inadecuación notable en el modelo sin ordenada al origen.

Para fines comparativos también se puede intentar el ajuste con el modelo con ordenada al origen. Eso da como resultado

$$\hat{y} = -0.0938 + 0.4071x$$

El estadístico t para probar H_0: $\beta_0 = 0$ es $t_0 = -0.65$, que no es significativa e implica que puede ser que el modelo sin ordenada al origen produzca un ajuste mejor. El cuadrado medio residual para el modelo con ordenada al origen es $MS_{\text{Res}} = 0.0931$ y $R^2 = 0.9997$. Como MS_{Res} para el modelo sin ordenada al origen es menor que MS_{Res} para el modelo con ordenada al origen, se llega a la conclusión que el modelo sin ordenada al origen es mejor. Como se dijo antes, los estadísticos R^2 no son directamente comparables.

También, en la figura 2.12 se ve el intervalo de confianza de 95% para $E(y|x_0)$ calculada con la ecuación (2.54), y el intervalo de predicción de 95% para una sola observación futura y_0 en $x = x_0$, calculado con la ecuación (2.55). Obsérvese que la longitud del intervalo de confianza para $x_0 = 0$ es cero.

2.9 ESTIMACIÓN POR MÁXIMA VEROSIMILITUD

El método de los mínimos cuadrados se puede aplicar para estimar los parámetros en un modelo de regresión lineal, independientemente de la forma de la distribución de los errores ε. Con los mínimos cuadrados se obtienen los mejores estimadores lineales insesgados de β_0 y β_1. Otros procedimientos estadísticos, como prueba de hipótesis y construcción de

http://carlos2524.jimdo.com/

intervalo de confianza, suponen que los errores se distribuyen normalmente. Si se conoce la forma de distribución de los errores, un método alternativo para estimar parámetros es el **método de máxima verosimilitud o máxima posibilidad**.

Se tienen los datos (y_i, x_i), $i = 1, 2, \ldots, n$. Si se supone que los errores en el modelo de regresión son NID$(0, \sigma^2)$, las observaciones y_i en esa muestra son variables aleatorias normal e independientemente distribuidas (NID) con promedio $\beta_0 + \beta_1 x_i$, y varianza σ^2. La función de verosimilitud o de posibilidad se determina con la distribución conjunta de las observaciones. Si se considera esta distribución conjunta con las observaciones dadas y los parámetros β_0, β_1 y σ^2 son constantes desconocidas, se tiene la función de posibilidad. Para el modelo de regresión lineal simple con errores normales, la **función de posibilidad** es

$$L(y_i, x_i, \beta_0, \beta_i, \sigma) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right]$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right] \tag{2.56}$$

Los estimadores de máxima posibilidad son los valores de los parámetros, por ejemplo $\tilde{\beta}_0$, $\tilde{\beta}_1$ y $\tilde{\sigma}^2$, que maximizan a L, o lo que es lo mismo, a $\ln L$. Así,

$$\ln L(y_i, x_i, \beta_0, \beta_1, \sigma^2) = -\left(\frac{n}{2}\right)\ln 2\pi - \left(\frac{n}{2}\right)\ln \sigma^2$$

$$-\left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \tag{2.57}$$

y los estimadores de posibilidad máxima, $\tilde{\beta}_0$, $\tilde{\beta}_1$ y $\tilde{\sigma}^2$ deben satisfacer

$$\left.\frac{\partial \ln L}{\partial \beta_0}\right|_{\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\sigma}^2} = \frac{1}{\tilde{\sigma}^2}\sum_{i=1}^{n}\left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i\right) = 0 \tag{2.58a}$$

$$\left.\frac{\partial \ln L}{\partial \beta_1}\right|_{\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\sigma}^2} = \frac{1}{\tilde{\sigma}^2}\sum_{i=1}^{n}\left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i\right)x_i = 0 \tag{2.58b}$$

y

$$\left.\frac{\partial \ln L}{\partial \sigma^2}\right|_{\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\sigma}^2} = -\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\sum_{i=1}^{n}\left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i\right)^2 = 0 \tag{2.58c}$$

La solución de la ecuación (2.58) determina los **estimadores de máxima posibilidad**:

$$\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{x} \tag{2.59a}$$

$$\tilde{\beta} = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.59b}$$

http://carlos2524.jimdo.com/

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i \right)^2}{n} \tag{2.59c}$$

Obsérvese que los estimadores de máxima posibilidad de la ordenada al origen y la pendiente, $\tilde{\beta}_0$ y $\tilde{\beta}_1$, son idénticos a los obtenidos con los mínimos cuadrados. También, $\tilde{\sigma}^2$ es un estimador sesgado de σ^2. El estimador sesgado se relaciona con el estimador insesgado $\hat{\sigma}^2$ [Ec. (2.19)] mediante $\tilde{\sigma}^2 = [n-1)/n]\hat{\sigma}^2$. El sesgo es pequeño cuando n es moderadamente grande, por lo general se usa el estimador insesgado $\hat{\sigma}^2$.

Casi siempre, los estimadores de máxima verosimilitud tienen mejores **propiedades estadísticas** que los estimadores por mínimos cuadrados. Los de máxima posibilidad son **insesgados** (incluyendo a $\tilde{\sigma}^2$, que es **asintóticamente insesgado** o insesgado cuando n se hace grande) y tienen **varianza mínima** comparados con **todos** los demás estimadores insesgados. También son **estimadores consistentes** (consistencia es una propiedad de una muestra grande que indica que los estimadores difieren del valor verdadero del parámetro) una cantidad muy pequeña, cuando n se hace grande. También, los estimadores de máxima posibilidad son un conjunto de **estadísticos suficientes** (esto implica que contienen toda la "información" de la muestra original, de tamaño n). Por otro lado, la estimación por máxima verosimilitud requiere hipótesis estadísticas más estrictas que las de los estimadores de mínimos cuadrados. Los estimadores de mínimos cuadrados requieren sólo hipótesis de segundo momento (que son hipótesis sobre el valor esperado, las varianzas y las covarianzas entre los errores aleatorios). Los estimadores de máxima posibilidad requieren una hipótesis de distribución total, en este caso, son errores aleatorios que siguen una distribución normal con los mismos segundos momentos con los que se requirieron en las estimaciones por mínimos cuadrados. Para conocer más información acerca de estimación de máxima verosimilitud en modelos de regresión, véase Graybill [1961, 1976], Myers [1990], Searle [1971] y Seber [1977].

2.10 CASO EN EL QUE EL REGRESOR x ES ALEATORIO

El modelo de regresión lineal que presentamos en este capítulo supone que los valores de la variable regresora x son constantes conocidas. Esta premisa hace que los coeficientes de confianza y los errores de tipo I (o de tipo II) se refieran a muestreo repetido de y a los mismos niveles de x. Hay muchos casos en los que es inadecuado suponer que las x son constantes fijas. Por ejemplo, considérense los datos de tiempo de abastecimiento de sodas, del capítulo 1 (Fig. 1.1). Como las tiendas visitadas por el vendedor se seleccionaron al azar, no es realista creer que se puede controlar el volumen entregado x, es más razonable suponer que tanto x como y son variables aleatorias.

Por fortuna y bajo ciertas circunstancias, todos los resultados anteriores acerca de la estimación de parámetros, prueba y predicción, son válidos. Ahora se describirán esos casos.

2.10.1 Distribución conjunta de x y y

Supongamos que x y y son variables aleatorias con distribución conjunta, pero que se desconoce la forma de esa distribución. Se puede demostrar que son válidos todos los resultados de regresión anteriores si se satisfacen las siguientes condiciones:

http://carlos2524.jimdo.com/

1. La distribución condicional de y dada x es normal, con media condicional $\beta_0 + \beta_1 x$ y varianza condicional σ^2.
2. Las x son variables aleatorias independientes, cuya distribución de probabilidades no implica a β_0, β_1 y σ^2.

Si bien todos los procedimientos de regresión no cambian cuando son válidas estas condiciones, los coeficientes de confianza y los errores estadísticos tienen una interpretación diferente. Cuando el regresor es una variable aleatoria, esas cantidades se aplican a valores repetidos de muestreo de (x_i, y_i), y no a muestreo repetido de y_i a niveles fijos de x_i.

2.10.2 Distribución normal conjunta de x y y: el modelo de correlación

Ahora supongamos que y y x tienen distribución conjunta que sigue la **distribución normal bivariada**. Esto es,

$$f(y, x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right) \right] \right\} \tag{2.60}$$

donde μ_1 y σ_1^2 son la media y la varianza de y, μ_2 y σ_2^2 son la media y la varianza de x, y

$$\rho = \frac{E(y-\mu_1)(x-\mu_2)}{\sigma_1\sigma_2} = \frac{\sigma_{12}}{\sigma_1\sigma_2}$$

es el **coeficiente de correlación** entre y y x. El término σ_{12} es la **covarianza** de y y x.

La **distribución condicional** de y para un valor dado de x es

$$f(y|x) = \frac{1}{\sqrt{2\pi}\,\sigma_{1.2}} \exp\left[-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1 x}{\sigma_{1.2}}\right)^2 \right] \tag{2.61}$$

siendo

$$\beta_0 = \mu_1 - \mu_2 \rho \frac{\sigma_1}{\sigma_2} \tag{2.62a}$$

$$\beta_1 = \frac{\sigma_1}{\sigma_2}\rho \tag{2.62b}$$

y

$$\sigma_{1.2}^2 = \sigma_1^2(1-\rho^2) \tag{2.62c}$$

Esto es, la distribución condicional de y dada x es normal, con media condicional

$$E(y|x) = \beta_0 + \beta_1 x \tag{2.63}$$

http://carlos2524.jimdo.com/

y varianza condicional $\sigma_{1.2}^2$. Nótese que la media de la distribución condicional de y dada x es un modelo de regresión de línea recta. Además, hay una relación entre el coeficiente de correlación ρ y la pendiente β_1. Se ve que, de acuerdo con la ecuación (2.62b), si $\rho = 0$, entonces $\beta_1 = 0$, y eso implica que no hay regresión lineal entre y y x, así es que el conocimiento de x no ayuda a predecir y.

El método de máxima verosimilitud se puede aplicar para estimar los parámetros β_0 y β_1. Se puede demostrar que los estimadores de máxima verosimilitud de esos parámetros son

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{2.64a}$$

y

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} \tag{2.64b}$$

Los estimadores de la ordenada al origen y la pendiente de la ecuación (2.64) son idénticos a los que se obtienen con el método de mínimos cuadrados en el caso en que se supuso que x era una variable controlable. En general, el modelo de regresión con distribución conjunta de y y x se puede analizar con los métodos presentados antes para el modelo en que x es una variable controlable. Esto se debe a que la variable aleatoria y dada x está distribuida en forma independiente y normal, con promedio $\beta_0 + \beta_1 x$ y varianza constante $\sigma_{1.2}^2$. Como se dijo en la sección 2.10.1, esos resultados también son válidos para **toda** distribución conjunta de y y x tal que la distribución condicional de y dada x sea normal.

Es posible obtener inferencias acerca del coeficiente de correlación ρ en este modelo. El estimador de ρ es el **coeficiente de correlación muestral**

$$r = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2\right]^{1/2}} = \frac{S_{xy}}{[S_{xx} SS_T]^{1/2}} \tag{2.65}$$

Nótese que

$$\hat{\beta}_1 = \left(\frac{SS_T}{S_{xx}}\right)^{1/2} r \tag{2.66}$$

por lo que la pendiente $\hat{\beta}$ es justamente el coeficiente de correlación muestral r multiplicado por un factor de escala que es la raíz cuadrada de la dispersión de las y, dividido entre la dispersión de las x. Por tanto, $\hat{\beta}_1$ y r se relacionan en forma estrecha, aunque proporcionan una información algo distinta. El coeficiente de correlación muestral r es una medida de la **asociación lineal** entre y y x, mientras que $\hat{\beta}_1$ mide el cambio del promedio de y cuando x cambia una unidad. En el caso de una variable controlable x, el coeficiente r no tiene significado, porque su magnitud depende de la elección de los espacios para x; también se puede escribir, de acuerdo con la ecuación (2.66), lo siguiente:

http://carlos2524.jimdo.com/

$$r^2 = \hat{\beta}_1^2 \frac{S_{xx}}{SS_T} = \frac{\hat{\beta}_1 S_{xy}}{SS_T}$$

$$= \frac{SS_R}{SS_T} = R^2$$

que de acuerdo con la ecuación (2.47) es el coeficiente de determinación, así es que el coeficiente de determinación R^2 no es más que el cuadrado del coeficiente de correlación entre y y x.

Si bien la regresión y la correlación tienen una relación estrecha, la regresión es más poderosa en muchos casos. La correlación sólo es una medida de la asociación, y tiene poco uso en predicción, sin embargo, los métodos de regresión se pueden aplicar para desarrollar relaciones cuantitativas entre las variables, y esas relaciones se pueden usar en predicciones.

Con frecuencia es útil probar la hipótesis que el coeficiente de correlación es cero, esto es:

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0 \tag{2.67}$$

Para esta hipótesis, el estadístico de prueba adecuado es

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{2.68}$$

que sigue la distribución t con $n-2$ grados de libertad si H_0: $\rho = 0$ es cierta. Así, se rechazaría la hipótesis nula si $|t_0| > t_{\alpha/2, n-2}$. Esta prueba equivale a la prueba t de H_0: $\beta_1 = 0$ descrita en la sección 2.3, y la equivalencia es consecuencia directa de la ecuación (2.66).

El procedimiento de prueba de las hipótesis

$$H_0: \rho = \rho_0$$
$$H_1: \rho \neq \rho_0 \tag{2.69}$$

en donde $\rho_0 \neq 0$ es algo más complicado. Para muestras moderadamente grandes, por ejemplo $n \geq 25$,

$$Z = \operatorname{arctanh} r = \frac{1}{2} \ln \frac{1+r}{1-2} \tag{2.70}$$

tiene una distribución aproximadamente normal, con media

$$\mu_Z = \operatorname{arctanh} \rho = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}$$

y varianza

$$\sigma_Z^2 = (n-3)^{-1}$$

http://carlos2524.jimdo.com/

Así, para probar la hipótesis H_0: $\rho = \rho_0$, se calcula el estadístico

$$Z_0 = (\text{arctanh } r - \text{arctanh } \rho_0)(n-3)^{1/2} \tag{2.71}$$

y se rechaza H_0: $\rho = \rho_0$ si $|Z_0| > Z_{\alpha/2}$.

También es posible establecer un intervalo de confianza de $100(1-\alpha)$ por ciento para ρ, usando la transformación (2.70). Este intervalo de confianza de $100(1-\alpha)$ por ciento es

$$\tanh\left(\text{arctanh } r - \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right) \le \rho \le \tanh\left(\text{arctanh } r + \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right) \tag{2.72}$$

en donde $\tanh u = (e^u - e^{-u})/(e^u + e^{-u})$.

Ejemplo 2.9 *Datos de tiempo de entrega*

Considérense los datos de tiempo de entrega de gaseosas que se presentó en el capítulo 1. Las 25 observaciones de tiempo de entrega y y volumen entregado x se ven en la tabla 2.9. El diagrama de dispersión de la figura 1.1 indica una fuerte relación lineal entre tiempo de entrega y volumen entregado. El coeficiente de correlación muestral, entre tiempo de entrega y y volumen entregado x es

$$r = \frac{S_{xy}}{[S_{xx}SS_T]^{1/2}} = \frac{2\,473.3440}{[(1\,136.5600)(5\,784.5426)]^{1/2}} = 0.9646$$

Si se supone que el tiempo de entrega y el volumen entregado tienen distribución normal conjunta, se podrán probar las hipótesis

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

TABLA 2.9 Datos para el ejemplo 2.9

Observación	Tiempo de entrega, y	Cantidad de cajas, x	Observación	Tiempo de entrega, y	Cantidad de cajas, x
1	16.68	7	14	19.75	6
2	11.50	3	15	24.00	9
3	12.03	3	16	29.00	10
4	14.88	4	17	15.35	6
5	13.75	6	18	19.00	7
6	18.11	7	19	9.50	3
7	8.00	2	20	35.10	17
8	17.83	7	21	17.90	10
9	79.24	30	22	52.32	26
10	21.50	5	23	18.75	9
11	40.33	16	24	19.83	8
12	21.00	10	25	10.75	4
13	13.50	4			

http://carlos2524.jimdo.com/

mediante el estadístico de prueba

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.9646\sqrt{23}}{\sqrt{1-0.9305}} = 17.55$$

Como $t_{0.025,\,23} = 2.069$, se rechaza H_0 y se llega a la conclusión que el coeficiente de correlación $\rho \neq 0$. Por último, se puede establecer un intervalo de confianza aproximado de 95% para ρ con la ecuación (2.72). Ya que arctanh r = arctanh 0.9646 = 2.0082, la ecuación (2.72) se transforma en

$$\tanh\left(2.0082 - \frac{1.96}{\sqrt{22}}\right) \leq \rho \leq \tanh\left(2.0082 + \frac{1.96}{\sqrt{22}}\right)$$

que se reduce a

$$0.9202 \leq \rho \leq 0.9845$$

Aunque se conoce que están muy relacionados el tiempo de entrega y el volumen entregado, esta información tiene poco valor para predecir, por ejemplo, el tiempo de entrega en función de la cantidad de cajas de producto entregadas. Para eso se necesitaría un modelo de regresión. El ajuste de una recta por mínimos cuadrados (se ve gráficamente en la figura 1.1b) que relaciona el tiempo de entrega y el volumen entregado es:

$$\hat{y} = 3.3208 + 2.1762x$$

Se necesitaría más análisis para determinar si esta ecuación es un ajuste adecuado a los datos, y si es probable que sea un buen predictor.

Resultados con SAS

La figura 2.13 muestra una parte de los resultados obtenidos con SAS PROC REG (otro paquete de programas muy usado para ajustar modelos de regresión) para los datos de tiempo de entrega en el ejemplo 2.9. El programa muestra el análisis de varianza para probar H_0: $\beta_1 = 0$, incluyendo el valor P de la prueba, refiriéndose a la fuente de variabilidad de

ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB > F
MODEL	1	5382.40880	5382.40880	307.846	0.0001
ERROR	23	402.13380	17.48407839		
C TOTAL	24	5784.54260			
ROOT MSE		4.181397	R-SQUARE	0.9305	
DEP MEAN		22.384	ADJ R-SQ	0.9275	
C.V.		18.68029			

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T for H0: PARAMETER = 0	PROB > |T|
INTERCEPT	1	3.3207790	1.37107367	2.422	0.0237
CASES	1	2.17616668	0.12402956	17.546	0.000

Figura 2.13 Resultados para los datos de tiempo de entrega en el ejemplo 2.9, obtenidos con SAS PROC REG.

http://carlos2524.jimdo.com/

regresión como modelo y a la fuente de variabilidad de residuales como "error". También se ven R^2 y R^2_{Ajus}, el error estándar de la regresión, $\sqrt{MS_{\text{Res}}}$ (que en los resultados se llama ROOT MSE), la media de la variable de respuesta y el coeficiente de variación (llamado C.V. en esta figura). El coeficiente de variación se calcula como $(\sqrt{MS_{\text{Res}}}/\bar{y})100$, y expresa la variabilidad que queda en los datos en relación con la respuesta media.

PROBLEMAS

2.1 En la tabla B.1 del apéndice aparecen datos sobre el desempeño de los 26 equipos de la Liga Nacional de Futbol en 1976. Se cree que la cantidad de yardas ganadas por tierra por los contrarios (x_8) tiene un efecto sobre la cantidad de juegos que gana un equipo (y).

a. Ajustar un modelo de regresión lineal simple que relacione los juegos ganados, y, con las yardas ganadas por tierra por los contrarios, x_8.

b. Formar la tabla de análisis de varianza y probar el significado de la regresión.

c. Determinar un intervalo de confianza de 95% para la pendiente.

d. ¿Qué porcentaje de variabilidad total da y, y explica este modelo?

e. Determinar un intervalo de confianza de 95% para la cantidad promedio de juegos ganados, si la distancia ganada por tierra por los contrarios se limita a 2 000 yardas.

2.2 Supóngase que se quiere usar el modelo desarrollado en el problema 2.1 para pronosticar la cantidad de juegos que ganará un equipo si puede limitar los avances por tierra de sus contrarios a 1 800 yardas. Determinar un estimado de punto de la cantidad de juegos ganados cuando $x_8 = 1\ 800$. Determinar un intervalo de predicción de 90% para la cantidad de juegos ganados.

2.3 La tabla B.2 del apéndice contiene datos reunidos durante un proyecto de energía solar en el Tecnológico de Georgia.

a. Ajustar un modelo de regresión lineal simple que relacione el flujo total de calor y (kilowatts) con la deflexión radial de los rayos desviados x_4 (milirradianes).

b. Formar la tabla de análisis de varianza y probar la significancia de la regresión.

c. Determinar un intervalo de confianza de 99% para la pendiente.

d. Calcular R^2.

e. Determinar un intervalo de confianza de 95% para el flujo promedio de calor, cuando la deflexión radial es 16.5 milirradianes.

2.4 La tabla B.3 del apéndice contiene datos sobre el rendimiento de la gasolina, en millas, de 32 automóviles diferentes.

a. Ajustar un modelo de regresión lineal simple que relacione el rendimiento de la gasolina y (millas por galón) y la cilindrada del motor x_1 (pulgadas cúbicas).

b. Formar la tabla de análisis de varianza y prueba de significancia de la regresión.

c. ¿Qué porcentaje de la variabilidad total del rendimiento de la gasolina explica la relación lineal con la cilindrada del motor?

d. Determinar un intervalo de confianza de 95% para el rendimiento promedio de gasolina, si el desplazamiento del motor es 275 pulg^3.

e. Suponer que se desea pronosticar el rendimiento de gasolina que tiene un coche con motor de 275 pulg^3. Determine un estimado puntual para el rendimiento. Determinar un intervalo de predicción de 95% para el rendimiento.

f. Comparar los dos intervalos obtenidos en las partes d y e. Explicar la diferencia entre ellos. ¿Cuál es más amplio y por qué?

http://carlos2524.jimdo.com/

2.5 Acerca de los datos sobre rendimiento de gasolina en la tabla B.3 del apéndice, repetir el problema 2.4 (partes a, b y c) usando el peso del vehículo, x_{10}, como la variable regresora. Con base en una comparación entre los dos modelos, ¿se puede llegar a la conclusión de que x_1 es mejor opción como regresor que x_{10}?

2.6 La tabla B.4 del apéndice presenta datos de 27 casas vendidas en Erie, Pennsylvania.

a. Ajustar un modelo de regresión lineal simple que relacione el precio de venta de la casa con los impuestos actuales (x_1).

b. Probar la significancia de la regresión.

c. ¿Qué porcentaje de la variabilidad total del precio de venta queda explicado con este modelo?

d. Determinar un intervalo de confianza de 95% para β_1.

e. Determinar un intervalo de confianza de 95% para el precio promedio de venta de una casa, para la cual los impuestos actuales son $750.

2.7 Se cree que la pureza del oxígeno producido con un proceso de fraccionamiento está relacionada con el porcentaje de hidrocarburos en el condensador principal de la unidad de procesamiento. A continuación se muestran los datos de veinte muestras.

Pureza (%)	Hidrocarburos (%)	Pureza (%)	Hidrocarburos (%)
86.91	1.02	96.73	1.46
89.85	1.11	99.42	1.55
90.28	1.43	98.66	1.55
86.34	1.11	96.07	1.55
92.58	1.01	93.65	1.40
87.33	0.95	87.31	1.15
86.29	1.11	95.00	1.01
91.86	0.87	96.85	0.99
95.61	1.43	85.20	0.95
89.86	1.02	90.56	0.98

a. Ajustar un modelo de regresión lineal simple a los datos.

b. Probar la hipótesis H_0: $\beta_1 = 0$.

c. Calcular R^2.

d. Determinar un intervalo de confianza de 95% para la pendiente.

e. Determinar un intervalo de confianza de 95% para la pureza media, cuando el porcentaje de hidrocarburos es 1.00.

2.8 Para los datos de la planta de oxígeno en el problema 2.7, suponer que la pureza y el porcentaje de hidrocarburos son variables aleatorias con distribución normal conjunta.

a. ¿Cuál es la correlación entre la pureza del oxígeno y el porcentaje de hidrocarburos?

b. Probar la hipótesis que $\rho = 0$.

c. Establecer un intervalo de confianza de 95% para ρ.

2.9 Con los datos de tiempo de entrega de bebidas gaseosas de la tabla 2.9, después de examinar el modelo original de regresión (Ej. 2.9), un analista afirmó que el modelo es inválido, porque la ordenada al origen no es cero. Dijo que si se entregaran cero cajas, el tiempo para abastecer y dar servicio a la máquina sería cero, y por eso el modelo de línea recta debe pasar por el origen. ¿Qué se le contestaría? Ajustar un modelo sin ordenada al origen a esos datos, y determinar qué modelo es el mejor.

http://carlos2524.jimdo.com/

2.10 A continuación se muestran el peso y la presión sistólica sanguínea de 26 hombres seleccionados al azar, en el grupo de edades de 25 a 30. Suponer que el peso y la presión sanguínea (BP) tienen distribución normal conjunta.

a. Determine una recta de regresión que relacione la presión sistólica sanguínea con el peso.
b. Estimar el coeficiente de correlación.
c. Probar la hipótesis que $\rho = 0$.
d. Probar la hipótesis que $\rho = 0.6$.
e. Determinar un coeficiente de confianza de 95% para ρ.

Persona	Peso	BP sistólica	Persona	Peso	BP sistólica
1	165	130	14	172	153
2	167	133	15	159	128
3	180	150	16	168	132
4	155	128	17	174	149
5	212	151	18	183	158
6	175	146	19	215	150
7	190	150	20	195	163
8	210	140	21	180	156
9	200	148	22	143	124
10	149	125	23	240	170
11	158	133	24	235	165
12	169	135	25	192	160
13	170	150	26	187	159

2.11 Para los datos de peso y presión sanguínea del problema 2.10, ajustar a ellos un modelo sin ordenada al origen y compararlo con el obtenido en el problema 2.10. ¿Cuál modelo se diría que es el mejor?

2.12 Se cree que la cantidad de libras de vapor usadas en una planta por mes está relacionada con la temperatura ambiente promedio. A continuación se presentan los consumos y las temperaturas del último año.

Mes	Temperatura	Uso / 1000	Mes	Temperatura	Uso / 1000
Ene.	21	185.79	Jul.	68	621.55
Feb.	24	214.47	Ago.	74	675.06
Mar.	32	288.03	Sep.	62	562.03
Abr.	47	424.84	Oct.	50	452.93
Mayo	50	454.68	Nov.	41	369.95
Jun.	59	539.03	Dic.	30	273.98

a. Ajustar un modelo de regresión lineal simple a los datos.
b. Probar la significancia de la regresión.
c. En la administración de la planta se cree que un aumento de 1 grado en la temperatura ambiente promedio hace aumentar 10 000 libras el consumo mensual de vapor. ¿Estos datos respaldan la afirmación?
d. Determinar un intervalo de predicción de 99% para el uso de vapor en un mes con temperatura ambiente promedio de 58°.

http://carlos2524.jimdo.com/

2.13 Davidson ("Update on Ozone Trends in California's South Coast Air Basin", *Air and Waste*, **43**, 226, 1993) estudió las concentraciones de ozono en la cuenca aérea de la costa sur de California, durante los años 1976 a 1991. Cree que la cantidad de días en que las concentraciones de ozono fueron mayores que 0.20 ppm (la respuesta) depende del índice meteorológico estacional, que es el promedio estacional de la temperatura con 850 milibars (el regresor). La siguiente tabla muestra los datos.

Año	Días	Índice
1976	91	16.7
1977	105	17.1
1978	106	18.2
1979	108	18.1
1980	88	17.2
1981	91	18.2
1982	58	16.0
1983	82	17.2
1984	81	18.0
1985	65	17.2
1986	61	16.9
1987	48	17.1
1988	61	18.2
1989	43	17.3
1990	33	17.5
1991	36	16.6

a. Trazar un diagrama de dispersión con los datos.
b. Estimar la ecuación de predicción.
c. Probar la significancia de la regresión.
d. Calcular y graficar las bandas de 95% de confianza y de predicción.

2.14 Hsuie, Ma y Tsai ("Separación y caracterización de copoliésteres termotrópicos del ácido p-hidroxibenzoico, ácido sebácico e hidroquinona", *Journal of Applied Polymer Science*, **56**, 471-476, 1995) estudian el efecto de la relación molar del ácido sebácico (el regresor) sobre la viscosidad intrínseca de los copoliésteres (la respuesta). La siguiente tabla muestra los datos.

Radio	Viscosidad
1.0	0.45
0.9	0.20
0.8	0.34
0.7	0.58
0.6	0.70
0.5	0.57
0.4	0.55
0.3	0.44

a. Trazar un diagrama de dispersión de los datos.
b. Estimar la ecuación de predicción.
c. Hacer un análisis completo y adecuado (pruebas estadísticas, cálculo de R^2, etcétera).
d. Calcular y graficar las bandas de 95% de confianza y de predicción.

http://carlos2524.jimdo.com/

2.15 Byers y Williams ("Viscosities of Binary and Ternary Mixtures of Polynomatic Hydrocarbons", *Journal of Chemical and Engineering Data,* **32**, 349-354, 1987) estudiaron el impacto de la temperatura sobre la viscosidad de mezclas de tolueno y tetralina. La tabla siguiente muestra los datos para mezclas con fracción molar de tolueno igual a 0.4.

Temperatura (°C)	Viscosidad (mPa · s)
24.9	1.1330
35.0	0.9772
44.9	0.8532
55.1	0.7550
65.2	0.6723
75.2	0.6021
85.2	0.5420
95.2	0.5074

a. Estimar la ecuación de predicción.
b. Hacer un análisis completo del modelo.
c. Calcular y graficar las bandas de 95% de confianza y de predicción.

2.16 Carroll y Spiegelman ("The Effects of Ignoring Small Measurement Errors in Precision Instrument Calibration", *Journal of Quality Technology*, **18**, 170-173, 1986) examinan la relación entre la presión en un tanque y el volumen de líquido. La siguiente tabla muestra los datos. Use un paquete adecuado de programas estadísticos para efectuar un análisis de ellos. Comente los resultados obtenidos con la rutina del programa.

Volumen	Presión	Volumen	Presión	Volumen	Presión
2084	4599	2842	6380	3789	8599
2084	4600	3030	6818	3789	8600
2273	5044	3031	6817	3979	9048
2273	5043	3031	6818	3979	9048
2273	5044	3221	7266	4167	9484
2463	5488	3221	7268	4168	9487
2463	5487	3409	7709	4168	9487
2651	5931	3410	7710	4358	9936
2652	5932	3600	8156	4358	9938
2652	5932	3600	8156	4546	10377
2842	6380	3788	8597	4547	10379

2.17 Para el modelo de regresión lineal simple $y = 50 + 10x + \varepsilon$, donde ε tiene NID (0, 16), suponer que se usan $n = 20$ pares de observaciones para ajustar este modelo. Generar 500 muestras de 20 observaciones, tomando una observación para cada valor de $x = 1, 1.5, 2, \ldots, 10$ para cada muestra.

a. Para cada muestra, calcular los estimados de la pendiente y la ordenada al origen por mínimos cuadrados. Trazar histogramas de los valores muestrales de $\hat{\beta}_0$ y $\hat{\beta}_1$. Comentar la forma de esos histogramas.
b. Para cada muestra, calcular un estimado de $E(y|x = 5)$. Trazar un histograma de los estimados obtenidos. Comentar la forma del histograma.

http://carlos2524.jimdo.com/

c. Determinar un intervalo de confianza de 95% para la pendiente en cada muestra. ¿Cuántos de los intervalos contienen el valor verdadero $\beta_1 = 10$? ¿Es lo que se esperaba?

d. Para cada estimado de $E(y|x = 5)$ en la parte b, calcular el intervalo de confianza de 95%. ¿Cuántos de esos intervalos contienen el valor verdadero de $E(y|x = 5) = 100$? ¿Es lo que se esperaba?

2.18 Repetir el problema 2.17 usando sólo 10 observaciones para cada muestra y tomando una observación de cada nivel $x = 1, 2, 3, \ldots, 10$. ¿Qué impacto tiene usar $n = 10$ sobre las respuestas en el problema 2.17? Comparar las longitudes de los intervalos de confianza y el aspecto de los histogramas.

2.19 Se tiene el modelo de regresión lineal simple $y = \beta_0 + \beta_1 x + \varepsilon$, con $E(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \sigma^2$ y ε no correlacionado.

a. Demostrar que $\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\bar{x}\sigma^2/S_{xx}$.

b. Demostrar que $\mathrm{Cov}(\bar{y}, \hat{\beta}_1) = 0$.

2.20 Se tiene el modelo de regresión lineal simple $y = \beta_0 + \beta_1 x + \varepsilon$, con $E(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \sigma^2$, y ε no correlacionado.

a. Demostrar que $E(MS_R) = \sigma^2 + \beta_1^2 S_{xx}$.

b. Demostrar que $E(MS_{Res}) = \sigma^2$.

2.21 Suponer que se debe ajustar el modelo rectilíneo de regresión $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$, pero que la respuesta está afectada por una segunda variable x_2, de tal modo que la verdadera función de regresión es

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

a. ¿Es insesgado el estimador de mínimos cuadrados para la pendiente, en el modelo original de regresión lineal simple?

b. Demostrar el sesgo en $\hat{\beta}_1$.

2.22 Considérese el estimador $7\sigma^2$ de máxima verosimilitud de σ^2 en el modelo de regresión lineal simple. Se sabe que $\tilde{\sigma}^2$ es un estimador sesgado de σ^2.

a. Demostrar la cantidad de sesgo en $\tilde{\sigma}^2$.

b. ¿Qué sucede con el sesgo a medida que se hace grande el tamaño n de la muestra?

2.23 Suponer que se debe ajustar una recta, y que se desea que el error estándar de la pendiente sea el mínimo posible. Suponer que la "región de interés de x es $-1 \le x \le 1$. ¿Dónde se deberían hacer las observaciones $x_1, x_2, \ldots, x_n$? Comente los aspectos prácticos de este plan de recolección de datos.

2.24 Se tienen los datos del problema 2.12. Supóngase que el consumo de vapor y la temperatura ambiente tienen distribución normal conjunta.

a. Determinar la correlación entre el consumo de vapor y la temperatura ambiente promedio mensual.

b. Probar la hipótesis que $\rho = 0$.

c. Probar la hipótesis que $\rho = 0.5$.

d. Determinar un intervalo de confianza de 99% para ρ.

2.25 Demostrar que el valor máximo de R^2 es menor que 1 si los datos contienen observaciones de y diferentes, para el mismo valor de x.

http://carlos2524.jimdo.com/

2.26 Se tiene el modelo de regresión lineal simple

$$y = \beta_0 + \beta_1 x + \varepsilon$$

donde se conoce la ordenada al origen β_0.

a. Determinar el estimador de β_1 de mínimos cuadrados para este modelo. ¿Parece razonable ese resultado?

b. ¿Cuál es la varianza de la pendiente ($\hat{\beta}_1$) para el estimador de mínimos cuadrados que se determinó en la parte a?

c. Determinar un intervalo de confianza de $100(1 - \alpha)$ por ciento para β_1. ¿Es este intervalo más angosto que el estimador para el caso en el que se desconocen tanto la pendiente como la ordenada al origen?

2.27 Se tienen los residuales de mínimos cuadrados $e_i = y_i - \hat{y}_i$, $i = 1, 2, \ldots, n$, obtenidos con el modelo de regresión lineal simple. Determinar la varianza $\text{Var}(e_i)$ de los residuales. ¿Es constante la varianza de los residuales? Comentar por qué.

http://carlos2524.jimdo.com/

CAPÍTULO 3

REGRESIÓN LINEAL MÚLTIPLE

Un modelo de regresión donde interviene más de una variable regresora se llama **modelo de regresión múltiple**. En este capítulo se explica el ajuste y análisis de esos modelos. Los resultados son extensiones de los que se obtuvieron en el capítulo 2 para la regresión lineal simple.

3.1 MODELO DE REGRESIÓN MÚLTIPLE

Supongamos que el rendimiento de la conversión, en libras, en un proceso químico, depende de la temperatura y de la concentración del catalizador. Un modelo de regresión múltiple que podría describir esta relación es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{3.1}$$

donde y representa el rendimiento, x_1 la temperatura y x_2 la concentración de catalizador. Este es un **modelo de regresión lineal múltiple** con dos variables regresoras. Se usa el término **lineal** porque la ecuación (3.1) es una función lineal de los parámetros desconocidos β_0, β_1 y β_2.

El modelo de regresión de la ecuación (3.1) describe un plano en el espacio tridimensional de y, x_1 y x_2. La figura 3.1a muestra este plano de regresión para el modelo

$$E(y) = 50 + 10x_1 + 7x_2$$

en el que se ha supuesto que el valor esperado del término ε de error en la ecuación (3.1) es cero. El parámetro β_0 es la ordenada al origen del plano de regresión. Si en el intervalo de datos se incluyen $x_1 = x_2 = 0$, entonces β_0 es el promedio de y cuando $x_1 = x_2 = 0$. Si no es así, β_0 no tiene interpretación física. El parámetro β_1 indica el cambio esperado de la respuesta y por cambio unitario en x_1, cuando x_2 se mantiene constante. De igual modo, β_2 mide el cambio esperado de y por unidad de cambio de x_2 cuando se mantiene constante x_1. La figura 3.1b muestra una **gráfica de curvas de nivel** del modelo de regresión, esto es, de líneas de respuesta esperada $E(y)$ constante, en función de x_1 y x_2. Nótese que las curvas de nivel en esta gráfica son rectas paralelas.

En general, se puede relacionar la **respuesta** y con k **regresores**, o **variables predictoras**. El modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{3.2}$$

se llama **modelo de regresión lineal múltiple** con k regresores. Los parámetros β_j, $j = 0, 1, \ldots, k$ se llaman **coeficientes de regresión**. Este modelo describe a un hiperplano en el espacio de k dimensiones de las variables regresoras x_j. El parámetro β_j representa el cambio

http://carlos2524.jimdo.com/

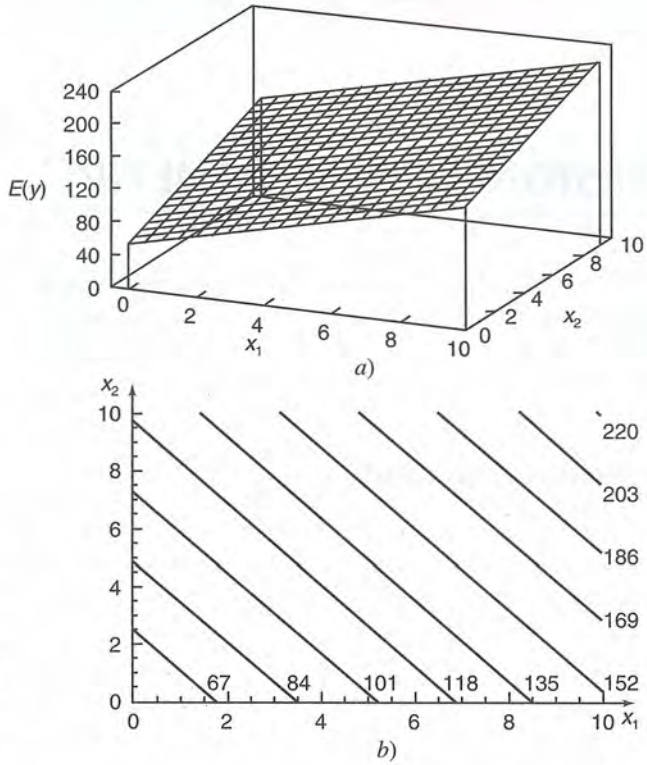


Figura 3.1 a) Plano de regresión para el modelo $E(y) = 50 + 10x_1 + 7x_2$. b) Gráfica de curvas de nivel.

esperado en la respuesta y por cambio unitario en x_j **cuando todas las demás variables regresoras** x_i $(i \neq j)$ **se mantienen constantes**. Por esta razón, a los parámetros β_j, $j = 1, 2, \ldots, k$ se les llama con frecuencia **coeficientes de regresión parcial**.

Los modelos de regresión parcial múltiple se usan con frecuencia como **modelos empíricos** o como funciones de aproximación, ya que se desconoce la relación funcional real entre y y $x_1, x_2, \ldots, x_k$, pero dentro de ciertos márgenes de las variables regresoras, el modelo de regresión lineal es una aproximación adecuada a la función verdadera desconocida.

Los modelos con estructura más complicada que la ecuación (3.2) se pueden analizar muchas veces con técnicas de regresión lineal múltiple. Por ejemplo, se tiene el modelo de polinomio cúbico

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon \tag{3.3}$$

Si se hace que $x_1 = x$, $x_2 = x^2$ y $x_3 = x^3$, entonces se puede escribir la ecuación (3.3) como sigue:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \tag{3.4}$$

que es un modelo de regresión lineal múltiple con tres variables regresoras. Los modelos polinomiales se explicarán con más detalle en el capítulo 7.

También se pueden analizar modelos que incluyan **efectos de interacción** con métodos de regresión lineal múltiple. Por ejemplo, si el modelo es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \tag{3.5}$$

http://carlos2524.jimdo.com/

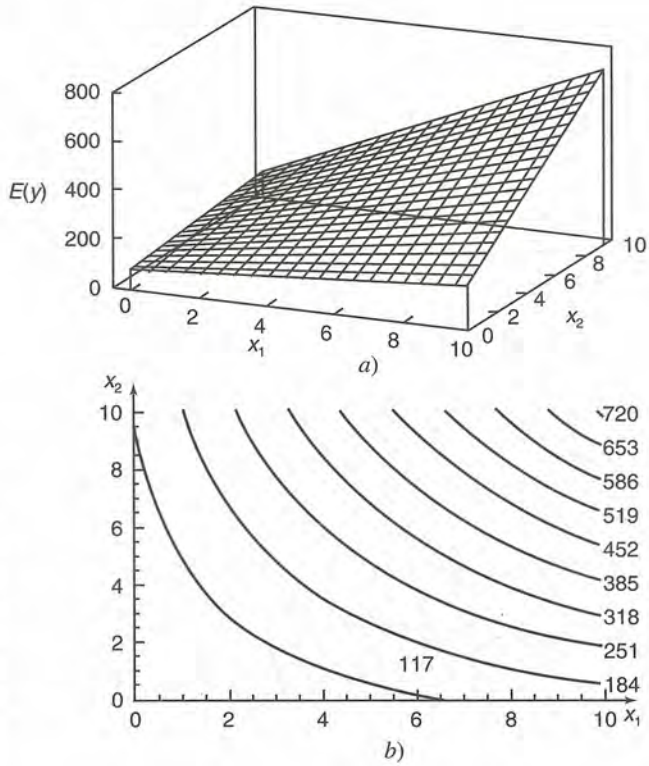


Figura 3.2 a) Gráfica tridimensional del modelo de regresión $E(y) = 50 + 10x_1 + 7x_2 + 5x_1x_2$. b) Gráfica de curvas de nivel.

Si se hace que $x_3 = x_1x_2$ y que $\beta_3 = \beta_{12}$, la ecuación (3.5) se podrá escribir en la forma

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \tag{3.6}$$

que es un modelo de regresión lineal.

La figura 3.2a muestra la gráfica tridimensional del modelo de regresión

$$y = 50 + 10x_1 + 7x_2 + 5x_1x_2$$

y la figura 3.2b la gráfica correspondiente bidimensional de curvas de nivel. Obsérvese que, aunque este es un modelo de regresión lineal, la forma de la superficie generada con el modelo no es lineal. En general, **todo modelo de regresión que es lineal en los parámetros** (las β) **es un modelo de regresión lineal, independientemente de la forma de la superficie que genera**.

La figura 3.2 es una buena interpretación gráfica de una interacción. En general, interacción implica que el efecto producido al cambiar una variable (por ejemplo x_1) depende del valor de la otra variable (x_2). Por ejemplo, la figura 3.2 muestra que al cambiar x_1 de 2 a 8 se produce un cambio mucho más pequeño de $E(y)$ cuando $x_2 = 2$ que cuando $x_2 = 10$. Los efectos de interacción se producen con frecuencia al estudiar y analizar los sistemas en el mundo real, y los métodos de regresión son una de las técnicas con las que se pueden describir.

Como ejemplo final, considérese el **modelo de segundo orden con interacción**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \tag{3.7}$$

http://carlos2524.jimdo.com/

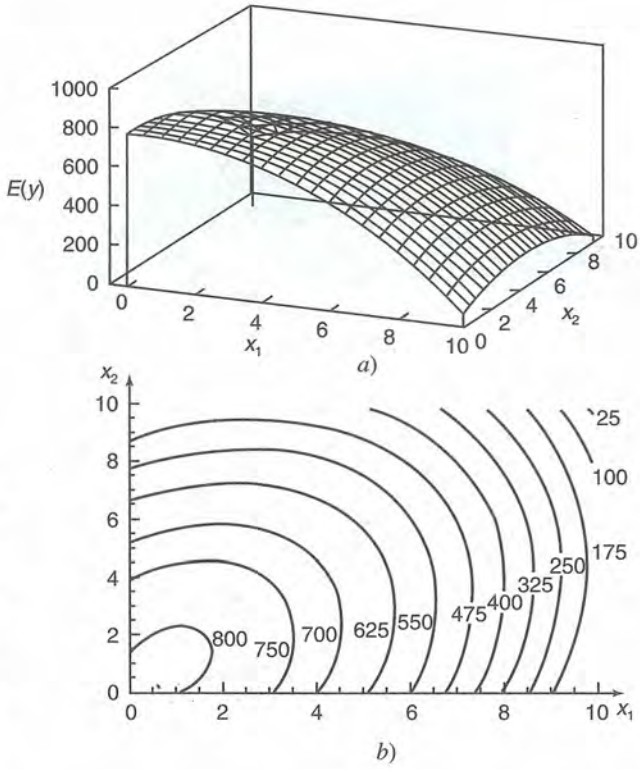


Figura 3.3 a) Gráfica tridimensional del modelo de regresión $E(y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$. b) Gráfica de curvas de nivel.

Si se igualan $x_3 = x_1^2$, $x_4 = x_2^2$, $x_5 = x_1x_2$, $\beta_3 = \beta_{11}$, $\beta_4 = \beta_{22}$ y $\beta_5 = \beta_{12}$, se podrá escribir la ecuación (3.7) como un modelo de regresión lineal múltiple, como sigue:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

La figura 3.3 muestra la gráfica tridimensional y la gráfica de curvas de nivel correspondiente para

$$E(y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$$

Estas gráficas indican que el cambio esperado de y cuando x_1 cambia (por ejemplo) una unidad, es una función de x_1 y x_2 **a la vez**. Los términos cuadrático y de interacción en este modelo producen una función en forma de colina. Dependiendo de los valores de los coeficientes de regresión, el modelo de segundo orden con interacción es capaz de tomar una gran variedad de formas, por lo que es un modelo de regresión muy flexible.

En la mayor parte de los problemas del mundo real no se conocen los valores de los parámetros (los coeficientes β_i) de regresión, ni la varianza del error σ^2, y se deben estimar a partir de datos muestrales. La ecuación o modelo de regresión ajustada se suele usar para pronosticar observaciones futuras de la variable de respuesta y, o para estimar la respuesta media a determinados valores de las y's.

http://carlos2524.jimdo.com/

3.2 ESTIMACIÓN DE LOS MODELOS DEL PARÁMETRO

3.2.1 Estimación de los coeficientes de regresión por mínimos cuadrados

Se puede aplicar el **método de mínimos cuadrados** para estimar los coeficientes de regresión de la ecuación (3.2). Supongamos que se dispone de $n > k$ observaciones, y sea y_i la i-ésima respuesta observada, y x_{ij} la i-ésima observación o nivel del regresor x_j. Los datos aparecerán como en la tabla 3.1. Se supone que el término de error ε del modelo tiene $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2$ y que los errores no están correlacionados.

TABLA 3.1 Datos para la regresión lineal múltiple

Observación	Respuesta	Regresores			
i	y	x_1	x_2	$\cdots$	x_k
1	y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}

En este capítulo se supondrá que las variables regresoras $x_1, x_2, \ldots, x_k$, son fijas, es decir, que son matemáticas o no aleatorias, y que se miden sin error. Sin embargo, tal como se describió en la sección 2.10 para el modelo de regresión lineal simple, todos los resultados obtenidos siguen siendo válidos para el caso en el que los regresores son variables aleatorias. Esto es realmente importante, porque cuando se toman datos de regresión en un **estudio observacional**, algunos o la mayor parte de los regresores son variables aleatorias. Cuando los datos son el resultado de un **experimento diseñado** es más probable que las x sean variables fijas. Cuando las x son variables aleatorias sólo es necesario que las observaciones con cada regresor sean independientes, y que la distribución no dependa de los coeficientes de regresión (las β) o de σ^2. Cuando se prueban hipótesis o se establecen intervalos de confianza, se debe suponer que la distribución condicional de y dadas $x_1, x_2, \ldots, x_k$ es normal, con promedio $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ y varianza σ^2.

Se puede escribir en la siguiente forma el modelo muestral de regresión que corresponde a la ecuación (3.2):

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n \end{aligned} \tag{3.8}$$

La función de mínimos cuadrados es

$$\begin{aligned} S(\beta_0, \beta_1, \ldots, \beta_k) &= \sum_{i=1}^{n} \varepsilon_i^2 \\ &= \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \end{aligned} \tag{3.9}$$

http://carlos2524.jimdo.com/

Se debe minimizar la función S respecto a $\beta_0, \beta_1, \ldots, \beta_k$. Los estimadores de $\beta_0, \beta_1, \ldots, \beta_k$ por mínimos cuadrados deben satisfacer

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k}\hat{\beta}_j x_{ij}\right) = 0 \tag{3.10a}$$

y

$$\left.\frac{\partial S}{\partial \beta_j}\right|_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k}\hat{\beta}_j x_{ij}\right)x_{ij} = 0, \quad j = 1, 2, \ldots, k \tag{3.10b}$$

Al simplificar la ecuación (3.10) se obtienen las **ecuaciones normales de mínimos cuadrados**

$$\begin{aligned}
n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n}x_{i1} + \hat{\beta}_2\sum_{i=1}^{n}x_{i2} + \cdots + \hat{\beta}_k\sum_{i=1}^{n}x_{ik} &= \sum_{i=1}^{n}y_i \\
\hat{\beta}_0\sum_{i=1}^{n}x_{i1} + \hat{\beta}_1\sum_{i=1}^{n}x_{i1}^2 + \hat{\beta}_2\sum_{i=1}^{n}x_{i1}x_{i2} + \cdots + \hat{\beta}_k\sum_{i=1}^{n}x_{i1}x_{ik} &= \sum_{i=1}^{n}x_{i1}y_i \\
&\vdots \\
\hat{\beta}_0\sum_{i=1}^{n}x_{ik} + \hat{\beta}_1\sum_{i=1}^{n}x_{ik}x_{i1} + \hat{\beta}_2\sum_{i=1}^{n}x_{ik}x_{i2} + \cdots + \hat{\beta}_k\sum_{i=1}^{n}x_{ik}^2 &= \sum_{i=1}^{n}x_{ik}y_i
\end{aligned} \tag{3.11}$$

Nótese que hay $p = k + 1$ ecuaciones normales, una para cada uno de los coeficientes desconocidos de regresión. La solución de las ecuaciones normales serán los **estimadores por mínimos cuadrados** $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$.

Es más cómodo manejar modelos de regresión múltiple cuando se expresan en notación matricial. Eso permite presentar en forma muy compacta al modelo, los datos y los resultados. En notación matricial el modelo expresado por la ecuación (3.8) es

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

en donde

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

http://carlos2524.jimdo.com/

En general, $\mathbf{y}$ es un vector de $n \times 1$ de las observaciones, $\mathbf{X}$ es una matriz de $n \times p$ de los niveles de las variables regresoras, β es un vector de $p \times 1$ de los coeficientes de regresión y ε es un vector de $n \times 1$ de errores aleatorios.

Se desea determinar el vector $\hat{\beta}$ de estimadores de mínimos cuadrados que minimice

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Nótese que $S(\beta)$ se puede expresar como sigue:

$$\begin{aligned} S(\beta) &= \mathbf{y}'\mathbf{y} - \beta'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta \\ &= \mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'\mathbf{X}'\mathbf{X}\beta \end{aligned}$$

ya que $\beta'\mathbf{X}'\mathbf{y}$ es una matriz de 1×1, es decir, un escalar, y que su transpuesta $(\beta'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\beta$ es el mismo escalar. Los estimadores de mínimos cuadrados deben satisfacer

$$\left.\frac{\partial S}{\partial \boldsymbol{\beta}}\right|_{\hat{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$$

que se simplifica a

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y} \tag{3.12}$$

Las ecuaciones (3.12) son las **ecuaciones normales de mínimos cuadrados**. Son la forma matricial de la presentación escalar, ecuaciones (3.11).

Para resolver las ecuaciones normales se multiplican ambos lados de (3.12) por la inversa de $\mathbf{X}'\mathbf{X}$. Así, el **estimador de β por mínimos cuadrados** es

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3.13}$$

siempre y cuando exista la matriz inversa $(\mathbf{X}'\mathbf{X})^{-1}$. La matriz $(\mathbf{X}'\mathbf{X})^{-1}$ siempre existe si los regresores son **linealmente independientes**, esto es, si ninguna columna de la matriz $\mathbf{X}$ es una combinación lineal de las demás columnas.

Es fácil de ver que la forma matricial de las ecuaciones normales (3.12) es idéntica a la forma escalar (3.11). Al escribir (3.12) con detalle se obtiene

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{ik}x_{i1} & \sum_{i=1}^{n} x_{ik}x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik}y_i \end{bmatrix}$$

Si se hace la multiplicación matricial indicada, se obtiene la forma escalar de las ecuaciones normales (3.11). En esta presentación se ve que $\mathbf{X}'\mathbf{X}$ es una matriz simétrica de $p \times p$, y que $\mathbf{X}'\mathbf{y}$ es un vector columna de $p \times 1$. Nótese la estructura especial de la

http://carlos2524.jimdo.com/

matriz $\mathbf{X}'\mathbf{X}$. Los elementos diagonales de $\mathbf{X}'\mathbf{X}$ son las sumas de los cuadrados de los elementos en las columnas de $\mathbf{X}$, y los elementos fuera de la diagonal son las sumas de los productos cruzados de los elementos de las columnas de $\mathbf{X}$. Además, nótese que los elementos de $\mathbf{X}'\mathbf{y}$ son las sumas de los productos cruzados de las columnas de $\mathbf{X}$ por las observaciones y_i.

El modelo ajustado de regresión que corresponde a los niveles de las variables regresoras $\mathbf{x}' = [1, x_1, x_2, \ldots, x_k]$ es

$$\hat{y} = \mathbf{x}'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_j$$

El vector de valores ajustados $\hat{y}_i$ que corresponden a los valores observados y_i es

$$\hat{\mathbf{y}} = \mathbf{X}\,\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y} \tag{3.14}$$

La $n \times n$ es la matriz $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ se suele llamar **matriz de sombrero**. Aplica el vector de valores observados en un vector de valores ajustados. La matriz de sombrero y sus propiedades desempeñan un papel central en el análisis de regresión.

La diferencia entre el valor observado y_i y el valor ajustado $\hat{y}_i$ correspondiente es el **residual** $e_i = y_i - \hat{y}_i$. Los n residuales se pueden escribir cómodamente con notación matricial como sigue:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \tag{3.15a}$$

Hay otras maneras de expresar el vector de residuales $\mathbf{e}$, que pueden ser útiles, como

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\,\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y} \tag{3.15b}$$

Ejemplo 3.1 *Datos del tiempo de entrega*

Un embotellador de bebidas gaseosas analiza las rutas de servicio de las máquinas expendedoras en su sistema de distribución. Le interesa predecir el tiempo necesario para que el representante de ruta atienda las máquinas expendedoras en una tienda. Esta actividad de servicio consiste en abastecer la máquina con productos embotellados, y algo de mantenimiento o limpieza. El ingeniero industrial responsable del estudio ha sugerido que las dos variables más importantes que afectan el tiempo de entrega y son la cantidad de cajas de producto abastecido, x_1, y la distancia caminada por el representante, x_2. El ingeniero ha reunido 25 observaciones de tiempo de entrega que se ven en la tabla 3.2. (Nótese que es una ampliación del conjunto de datos que se usó en el ejemplo 2.9.) Se ajustará el modelo de regresión lineal múltiple siguiente:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

con los datos de tiempo de entrega en la tabla 3.2.

Las gráficas son muy útiles para ajustar modelos de regresión múltiple. La figura 3.4 es una **matriz de dispersión** de los datos de tiempo de entrega. Es un arreglo bidimensional de gráficas bidimensionales, en las que, a excepción de los de la diagonal, cada cuadro contiene un diagrama de dispersión. Así, cada gráfica trata de arrojar alguna luz sobre la relación entre un par de variables. Con frecuencia esto es un mejor resumen de las relaciones

http://carlos2524.jimdo.com/

TABLA 3.2 Datos de tiempo de entrega para el ejemplo 3.1

Observación número	Tiempo de entrega (minutos) y	Cantidad de cajas x_1	Distancia (pies) x_2
1	16.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8.00	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.50	5	605
11	40.33	16	688
12	21.00	10	215
13	13.50	4	255
14	19.75	6	462
15	24.00	9	448
16	29.00	10	776
17	15.35	6	200
18	19.00	7	132
19	9.50	3	36
20	35.10	17	770
21	17.90	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

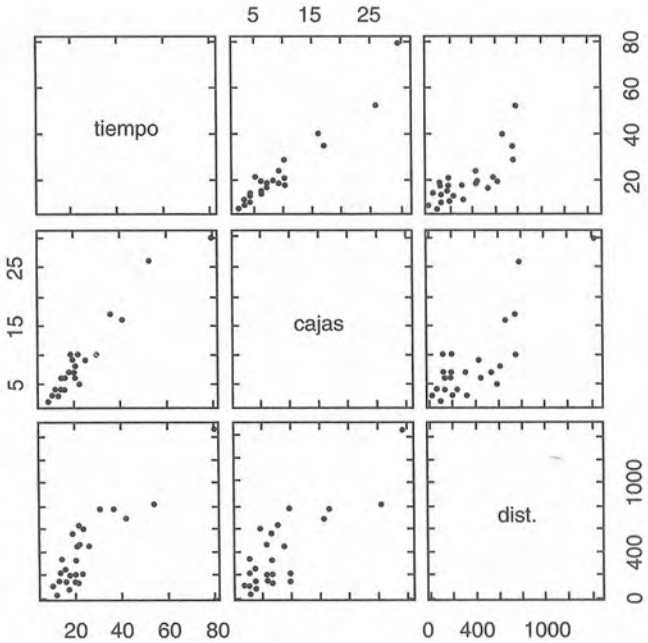


Figura 3.4 Matriz de diagramas de dispersión para los datos de tiempo de entrega del ejemplo 3.1.

http://carlos2524.jimdo.com/

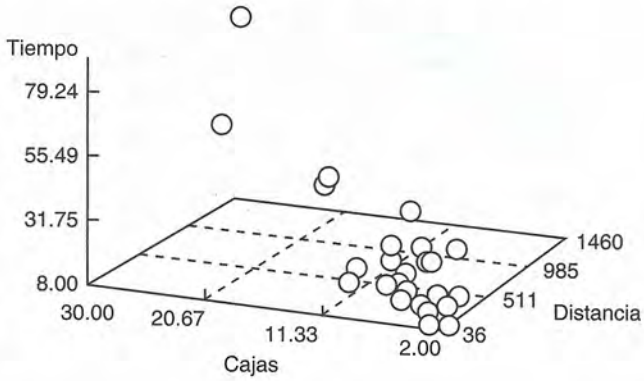


Figura 3.5 Diagrama tridimensional de dispersión para los datos de tiempo de entrega del ejemplo 3.1.

que una presentación numérica, como por ejemplo mostrar los coeficientes de correlación entre cada par de variables, porque muestra un sentido de linealidad o de no linealidad en la relación, y cierta percepción de cómo se arreglan los datos individuales en la región.

Cuando sólo hay dos regresores, a veces un diagrama tridimensional de dispersión es útil para visualizar la relación entre la respuesta y los regresores. La figura 3.5 muestra esta gráfica para los datos de tiempo de entrega. Al girar estas gráficas, algunos problemas de cómputo permiten captar diferentes vistas de la nube de puntos. Esta perspectiva proporciona una indicación de que un modelo de regresión lineal múltiple puede dar un ajuste razonable de los datos.

Para ajustar el modelo de regresión múltiple, primero se forman la matriz **X** y el vector **y**:

$$\mathbf{X} = \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ 1 & 3 & 340 \\ 1 & 4 & 80 \\ 1 & 6 & 150 \\ 1 & 7 & 330 \\ 1 & 2 & 110 \\ 1 & 7 & 210 \\ 1 & 30 & 1460 \\ 1 & 5 & 605 \\ 1 & 16 & 688 \\ 1 & 10 & 215 \\ 1 & 4 & 255 \\ 1 & 6 & 462 \\ 1 & 9 & 448 \\ 1 & 10 & 776 \\ 1 & 6 & 200 \\ 1 & 7 & 132 \\ 1 & 3 & 36 \\ 1 & 17 & 770 \\ 1 & 10 & 140 \\ 1 & 26 & 810 \\ 1 & 9 & 450 \\ 1 & 8 & 635 \\ 1 & 4 & 150 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 16.68 \\ 11.50 \\ 12.03 \\ 14.88 \\ 13.75 \\ 18.11 \\ 8.00 \\ 17.83 \\ 79.24 \\ 21.50 \\ 40.33 \\ 21.00 \\ 13.50 \\ 19.75 \\ 24.00 \\ 29.00 \\ 15.35 \\ 19.00 \\ 9.50 \\ 35.10 \\ 17.90 \\ 52.32 \\ 18.75 \\ 19.83 \\ 10.75 \end{bmatrix}$$

http://carlos2524.jimdo.com/

La matriz $\mathbf{X}'\mathbf{X}$ es

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 4 \\ 560 & 220 & \cdots & 150 \end{bmatrix} \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 4 & 150 \end{bmatrix}$$

$$= \begin{bmatrix} 25 & 219 & 10\,232 \\ 219 & 3\,055 & 133\,899 \\ 10\,232 & 133\,899 & 6\,725\,688 \end{bmatrix}$$

y el vector $\mathbf{X}'\mathbf{y}$ es

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 4 \\ 560 & 220 & \cdots & 150 \end{bmatrix} \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 10.75 \end{bmatrix} = \begin{bmatrix} 559.60 \\ 7\,375.44 \\ 337\,072.00 \end{bmatrix}$$

El estimador de β por mínimos cuadrados es

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

o sea

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 25 & 219 & 10\,232 \\ 219 & 3{,}055 & 133\,899 \\ 10\,232 & 133\,899 & 6\,725\,688 \end{bmatrix}^{-1} \begin{bmatrix} 559.60 \\ 7\,375.44 \\ 337\,072.00 \end{bmatrix}$$

$$= \begin{bmatrix} .11321518 & -.00444859 & -.00008367 \\ -.00454859 & .00274378 & -.00004786 \\ -.00008367 & -.00004786 & .00000123 \end{bmatrix} \begin{bmatrix} 559.60 \\ 7\,375.44 \\ 337\,072.00 \end{bmatrix}$$

$$= \begin{bmatrix} 2.34123115 \\ 1.61590712 \\ 0.01438483 \end{bmatrix}$$

El ajuste por mínimos cuadrados, con los coeficientes de regresión expresados con cinco decimales, es

$$\hat{y} = 2.34123 + 1.61591x_1 + 0.01438x_2$$

La tabla 3.3 muestra las y_i observadas junto con los valores ajustados $\hat{y}_i$ correspondientes, y los residuales e_i de este modelo.

http://carlos2524.jimdo.com/

TABLA 3.3 Observaciones, valores ajustados y residuales para el ejemplo 3.1

Observación número	y_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$
1	16.68	21.7081	–5.0281
2	11.50	10.3536	1.1464
3	12.03	12.0798	–0.0498
4	14.88	9.9556	4.9244
5	13.75	14.1944	–0.4444
6	18.11	18.3996	–0.2896
7	8.00	7.1554	0.8446
8	17.83	16.6734	1.1566
9	79.24	71.8203	7.4197
10	21.50	19.1236	2.3764
11	40.33	38.0925	2.2375
12	21.00	21.5930	–0.5930
13	13.50	12.4730	1.0270
14	19.75	18.6825	1.0675
15	24.00	23.3288	0.6712
16	29.00	29.6629	–0.6629
17	15.35	14.9136	0.4364
18	19.00	15.5514	3.4486
19	9.50	7.7068	1.7932
20	35.10	40.8880	–5.7880
21	17.90	20.5142	–2.6142
22	52.32	56.0065	–3.6865
23	18.75	23.3576	–4.6076
24	19.83	24.4028	–4.5728
25	10.75	10.9626	–0.2126

Resultados con SAS

La figura 3.6 muestra una parte de los resultados obtenidos en computadora con el paquete SAS (rutina SAS PROC REG) para los datos de tiempo de entrega de bebidas gaseosas del ejemplo 3.1. Si bien el formato de resultados es distinto entre los programas de cómputo, en este caso se muestra la información que se suele generar. Se ha llamado TIME a la variable de respuestas, y a las variables regresoras se le ha llamado CASES (cajas) y DISTANCE. La mayor parte de los resultados de la figura 3.6 son una ampliación directa, al caso de regresión múltiple de los resultados obtenidos en la regresión lineal simple. En las siguientes secciones se explicará esta información sobre los resultados. También se presentarán y describirán otras partes de los resultados obtenidos con SAS PROC REG.

3.2.2 Interpretación geométrica de mínimos cuadrados

A veces ayuda tener una interpretación geométrica intuitiva de mínimos cuadrados. Se puede imaginar que el vector de las observaciones $\mathbf{y}' = [y_1, y_2, \ldots, y_n]$ define un vector del origen al punto A de la figura 3.7. Nótese que $y_1, y_2, \ldots, y_n$ forman las coordenadas de un espacio muestral n dimensional. En la figura 3.7, el espacio muestral es tridimensional.

http://carlos2524.jimdo.com/

ANALISYS DE VARIANTE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB > F
MODEL	2	5550.81092	2775.40546	261.235	0.0001
ERROR	22	233.73168	10.62416716		
C TOTAL	24	5784.54260			
ROOT MSE		3.259473	R-SQUARE	0.9596	
DEP MEAN		22.384	ADJ R-SQ	0.9559	
C.V.		14.56162			

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HO: PARAMETER = 0	PROB > |T|	VARIANCE INFLATION
INTERCEP	1	2.34123115	1.09673017	2.135	0.0442	0
CASES	1	1.61590721	0.17073492	9.464	0.0001	3.11841
DISTANCE	1	0.01438483	0.003613086	3.981	0.0006	3.11841

Figura 3.6 Resultados para los datos de tiempo de entrega en el ejemplo 3.1, obtenidos con SAS PROC REG.

La matriz **X** consiste en p vectores columna de $n \times 1$, por ejemplo; **1** un vector columna de unos; y $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$. Cada una de esas columnas define a un vector desde el origen en el espacio muestral. Estos p vectores forman un subespacio p dimensional llamado **espacio de estimación**. El espacio de estimación para $p = 2$ se muestra en la figura 3.7. Se puede representar cualquier punto en este subespacio mediante una combinación lineal de los vectores $\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_k$. Así, cualquier punto del espacio de estimación tiene la forma $\mathbf{X}\beta$. Sea el punto B de la figura 3.7 el determinado por el vector $\mathbf{X}\beta$. La distancia de B a A elevada al cuadrado es

$$S(\beta) = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

Así, para minimizar la distancia al cuadrado del punto A definido por el vector de observación **y** al espacio de estimación, se requiere determinar el punto del espacio de estimación que está más cercano a A. La distancia al cuadrado será mínima cuando el punto del espacio de estimación esté al pie de la recta que llega de A y es normal (o perpendicular)

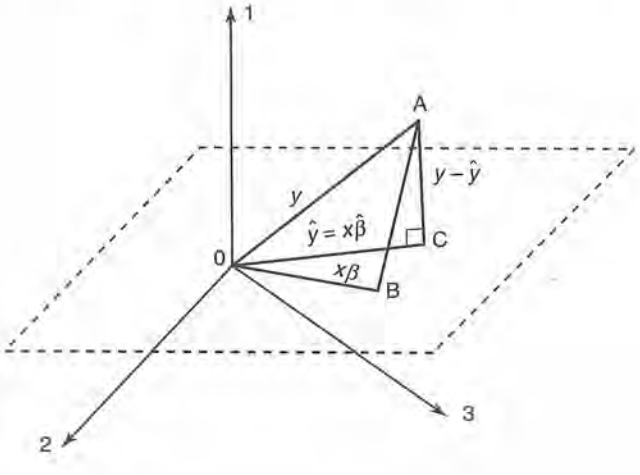


Figura 3.7 Interpretación geométrica de los mínimos cuadrados.

http://carlos2524.jimdo.com/

al espacio de estimación. Es el punto C de la figura 3.7. Este punto se define con el vector $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$. Por consiguiente, como $\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}$ es perpendicular al espacio de estimación, se podrá escribir

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{0}, \qquad \text{o bien} \qquad \mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$

que se reconocen como las ecuaciones normales de mínimos cuadrados.

3.2.3 Propiedades de los estimadores de mínimos cuadrados

Las propiedades estadísticas del estimador de mínimos cuadrados $\hat{\beta}$ se demuestran con facilidad. Examinaremos primero el sesgo:

$$\begin{aligned} E(\hat{\beta}) &= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right] = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon)\right] \\ &= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\right] = \beta \end{aligned}$$

porque $E(\varepsilon) = \mathbf{0}$ y $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{1}$. Entonces $\hat{\beta}$ es un **estimador insesgado** de β.

La propiedad de varianza de $\hat{\beta}$ se expresa con la **matriz de covarianza**

$$\mathrm{Cov}(\hat{\beta}) = E\left\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\right\}$$

que es una matriz simétrica de $p \times p$, cuyo j-ésimo elemento diagonal es la varianza de $\hat{\beta}_j$ y cuyo (ij)-ésimo elemento fuera de la diagonal es la covarianza entre $\hat{\beta}_i$ y $\hat{\beta}_j$. La covarianza de la matriz de $\hat{\beta}$ es

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Por consiguiente, si hacemos $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$, la varianza de $\hat{\beta}_j$ es $\sigma^2 C_{jj}$, y la covarianza entre $\hat{\beta}_i$ y $\hat{\beta}_j$ es $\sigma^2 C_{ij}$.

En el apéndice C.4 se demuestra que el estimador de mínimos cuadrados $\hat{\beta}$ es el mejor estimador lineal insesgado de β (el teorema de Gauss-Markov). Si además se supone que los errores ε_i tienen distribución normal, entonces, como se verá en la sección 3.2.6, $\hat{\beta}$ también es el estimador de máxima probabilidad de β. El estimador de máxima verosimilitud es el estimador insesgado de β de mínima varianza.

3.2.4 Estimación de σ^2

Como en la regresión lineal simple, se puede desarrollar un estimador de σ^2 a partir de la suma de cuadrados de residuales:

$$\begin{aligned} SS_{\text{Res}} &\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} e_i^2 \\ &= \mathbf{e}'\mathbf{e} \end{aligned}$$

http://carlos2524.jimdo.com/

Se sustituye $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$ y se obtiene

$$\begin{aligned} SS_{\text{Res}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned}$$

Como $\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$, la última ecuación se transforma en

$$SS_{\text{Res}} = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} \tag{3.16}$$

En el apéndice C.3 se demuestra que la suma de cuadrados de residuales tiene $n - p$ grados de libertad asociados con ella, porque se estiman p parámetros en el modelo de regresión. El **cuadrado medio residual**, o **cuadrado medio de residuales** es

$$MS_{\text{Res}} = \frac{SS_{\text{Res}}}{n - p} \tag{3.17}$$

También se demuestra en el apéndice C.3 que el valor esperado de MS_{Res} es σ^2, por lo que un **estimador insesgado** de σ^2 es

$$\hat{\sigma}^2 = MS_{\text{Res}} \tag{3.18}$$

Como se dijo en el caso de regresión lineal simple, el estimador de σ^2 **depende del modelo**.

Ejemplo 3.2 *Datos del tiempo de entrega*

Se estimará la varianza del error, σ^2, para el ajuste del modelo de regresión múltiple a los datos de tiempo de entrega de bebidas gaseosas en el ejemplo 3.1. Ya que

$$\mathbf{y}'\mathbf{y} = \sum_{i=1}^{25} y_i^2 = 18\,310.6290$$

y

$$\begin{aligned} \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} &= \begin{bmatrix} 2.34123115 & 1.61590721 & 0.01438483 \end{bmatrix} \begin{bmatrix} 559.60 \\ 7\,375.44 \\ 337\,072.00 \end{bmatrix} \\ &= 18\,076.90304 \end{aligned}$$

la suma de cuadrados de residuales es

$$\begin{aligned} SS_{\text{Res}} &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \\ &= 18\,310.6290 - 18\,076.9030 = 233.7260 \end{aligned}$$

http://carlos2524.jimdo.com/

Por consiguiente, el estimado de σ^2 es el cuadrado medio de residuales

$$\hat{\sigma}^2 = \frac{SS_{\text{Res}}}{n-p} = \frac{233.7260}{25-3} = 10.6239$$

Los resultados de SAS en la figura 3.6 reportan el cuadrado medio de los errores (sinónimo de cuadrado medio de residuales) como igual a 10.62416716. Esta diferencia relativamente pequeña se debe al redondeo en los cálculos.

La naturaleza dependiente del modelo de este estimado de σ^2 se puede demostrar con facilidad. La figura 2.13 muestra los resultados en computadora de un ajuste por mínimos cuadrados para los datos del tiempo de entrega usando sólo un regresor, las cajas x_1. El cuadrado medio de residuales para este modelo es 17.4841, que es bastante mayor que el resultado obtenido arriba para el modelo con dos regresores. ¿Cuál estimado es el "correcto"? En cierto sentido, ambos estimados son correctos, pero dependen mucho de la elección del modelo. Quizá una pregunta mejor sea ¿cuál **modelo** es el correcto? Como σ^2 es la varianza de los errores (el ruido no explicado en torno a la recta de regresión), normalmente se preferiría un modelo con un cuadrado medio de residuales menor, a uno con cuadrado medio más grande.

3.2.5 Inadecuación de los diagramas de dispersión en regresión múltiple

Se vio en el capítulo 2 que el diagrama de dispersión es una herramienta importante para analizar la relación entre y y x en la regresión lineal simple. También se vio en el ejemplo 3.1 que la **matriz de diagramas de dispersión** es útil para visualizar la relación entre y y dos regresores. Es tentador llegar a la conclusión que éste es un concepto general; ya, que siempre es útil examinar los diagramas de dispersión de y en función de x_1, y en función de $x_2, \ldots, y$ en función de x_k, para evaluar las relaciones entre y y cada uno de los regresores $x_1, x_2, \ldots, x_k$. Desafortunadamente, y en general, eso no es cierto.

De acuerdo con Daniel y Wood [1980] se ilustrará lo inadecuado de los diagramas de dispersión en un problema con dos regresores. Se tienen los siguientes datos:

y	x_1	x_2
10	2	1
17	3	2
48	4	5
27	1	2
55	5	6
26	6	4
9	7	3
16	8	4

Estos datos se generaron con la ecuación

$$y = 8 - 5x_1 + 12x_2$$

La matriz de diagramas de dispersión se muestra en la figura 3.8. La gráfica de y en función de x_1 no muestra relación alguna aparente entre las dos variables. La gráfica de y en

http://carlos2524.jimdo.com/

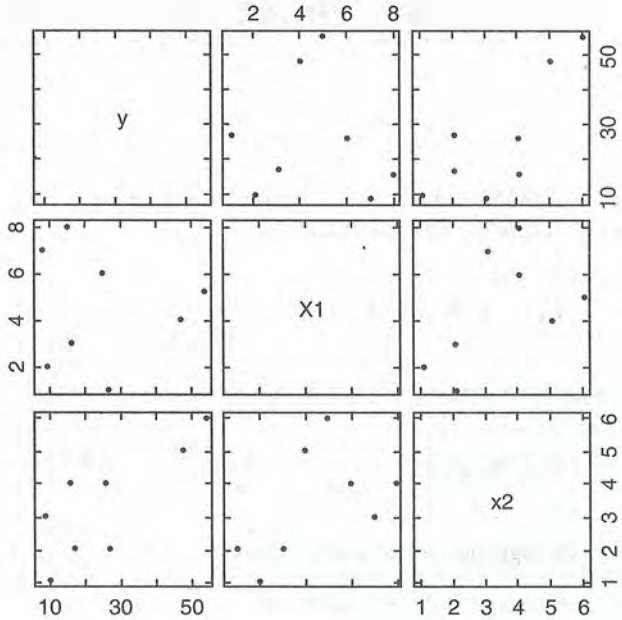


Figura 3.8 Una matriz de diagramas de dispersión.

función de x_2 indica que existe una relación lineal, con una pendiente aproximada de 8. Nótese que ambos diagramas de dispersión contienen información errónea. Como en este conjunto de datos hay dos pares de puntos que tienen los mismos valores de x_2 ($x_2 = 2$ y $x_2 = 4$), se podría medir el efecto de x_1 con x_2 fija con ambos pares. Así se obtiene $x_2 = 2$, $\hat{\beta}_1 = (17 - 27)/(3 - 1) = -5$ para $x_2 = 4$, y $\hat{\beta}_1 = (26 - 16)/(6 - 8) = -5$, que son los resultados correctos. Conocido $\hat{\beta}_1$, se puede estimar ya el efecto de x_2. Sin embargo, y en general, este procedimiento no es útil, porque muchos conjuntos de datos no tienen puntos duplicados.

Este ejemplo ilustra que puede ser engañoso trazar diagramas de dispersión de y en función de x_j ($j = 1, 2, \ldots, k$), aun en el caso de sólo dos regresores operando en una forma perfectamente aditiva y sin ruido. Un caso más realista de regresión con varios regresores y error en las y confundiría todavía más la situación. Si sólo hay un (o unos pocos) regresor dominante, o si los regresores funcionan casi en forma independiente, la matriz de diagramas de dispersión es más útil. Sin embargo, cuando varios regresores importantes están interrelacionados, estos diagramas de dispersión pueden ser muy engañosos. En el capítulo 9 se describirán métodos analíticos para eliminar las relaciones entre varios regresores y una respuesta.

3.2.6 Estimación de máxima posibilidad

Así como en el caso de la regresión lineal simple, se puede demostrar que los estimadores de máxima posibilidad para los parámetros del modelo también son estimadores de mínimos cuadrados, cuando los errores del modelo están distribuidos en forma normal e independiente. El modelo es

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

http://carlos2524.jimdo.com/

y los errores se distribuyen en forma normal e independiente con varianza constante igual a σ^2, o bien ε se distribuye como $N(\mathbf{0}, \sigma^2\mathbf{I})$. La función de densidad normal de los errores es

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2}} e^{-\frac{1}{2\sigma^2}\varepsilon_i^2}$$

La función de posibilidad es la densidad conjunta de $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, es decir, $\Pi_{i=1}^n f(\varepsilon_i)$. Por consiguiente, la función de posibilidad es

$$L(\varepsilon, \beta, \sigma^2) = \prod_{i=1}^{n} f(\varepsilon_i) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\varepsilon_i'\varepsilon}$$

Ahora, como se puede escribir $\varepsilon = \mathbf{y} - \mathbf{X}\beta$, la función de posibilidad se transforma en

$$L(\mathbf{y}, \mathbf{X}, \beta, \sigma^2) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}$$

Como en el caso de regresión lineal simple, conviene trabajar con el logaritmo de la posibilidad:

$$\ln L(\mathbf{y}, \mathbf{X}, \beta, \sigma^2) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

Es claro que para un valor fijo de σ, el logaritmo de posibilidad se maximiza cuando se minimiza el término

$$(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

Por consiguiente, el estimador de máxima posibilidad de β bajo errores normales equivale al estimador por mínimos cuadrados $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. El estimador de máxima posibilidad de σ^2 es

$$\tilde{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})}{n}$$

Estas generalizaciones de los resultados obtenidos para la regresión lineal simple, en el capítulo 2, sección 2.7, para el caso de la regresión múltiple. Las propiedades estadísticas de los estimadores de máxima posibilidad se resumen en la sección 2.7.

3.3 PRUEBA DE HIPÓTESIS EN LA REGRESIÓN LINEAL MÚLTIPLE

Una vez estimados los parámetros del modelo, surgen de inmediato dos preguntas:

1. ¿Cuál es la adecuación general del modelo?
2. ¿Cuáles regresores específicos parecen importantes?

Hay varios procedimientos de prueba de hipótesis que demuestran su utilidad para contestar estas preguntas. Las pruebas formales requieren que los errores aleatorios sean independientes y tengan una distribución normal con promedio $E(\varepsilon_i) = 0$ y una varianza $\text{Var}(\varepsilon_i) = \sigma^2$.

http://carlos2524.jimdo.com/

3.3.1 Prueba de la significancia de la regresión

La prueba de la **significancia de la regresión** es para determinar si hay una **relación lineal** entre la respuesta y y cualquiera de las variables regresoras $x_1, x_2, \ldots, x_k$. Este procedimiento suele considerarse como una prueba general o global de la adecuación del modelo. Las hipótesis pertinentes son:

$$H_0\colon \beta_1 = \beta_1 = \cdots = \beta_k = 0$$

$$H_1\colon \beta_j \neq 0 \text{ al menos para una } j.$$

El rechazo de la hipótesis nula implica que al menos uno de los regresores $x_1, x_2, \ldots, x_k$ contribuye al modelo en forma significativa.

El procedimiento de prueba es una generalización del **análisis de varianza** que se usó en la regresión lineal simple. La **suma total de cuadrados** SS_T se divide en una **suma de cuadrados debidos a la regresión**, SS_R, y a una **suma de cuadrados de residuales**, SS_{Res}. Así,

$$SS_T = SS_R + SS_{Res}$$

En el apéndice C.3 se demuestra que si es cierta la hipótesis nula, entonces SS_R/σ^2 tiene una distribución χ^2_k, con la misma cantidad de grados de libertad que la cantidad de variables regresoras en el modelo. También se demuestra en el apéndice C.3 que $SS_{Res}/\sigma^2 \sim \chi^2_{n-k-1}$ y que SS_{Res} y SS_R son independientes. De acuerdo con la definición de un estadístico F que aparece en el apéndice C.1,

$$F_0 = \frac{SS_R/k}{SS_{Res}/(n-k-1)} = \frac{MS_R}{MS_{Res}}$$

tiene la distribución $F_{k,\,n-k-1}$. En el apéndice C.3 se demuestra que

$$E(MS_{Res}) = \sigma^2$$

$$E(MS_R) = \sigma^2 + \frac{\boldsymbol{\beta}^{*\prime}\mathbf{X}_c'\mathbf{X}_c\boldsymbol{\beta}^*}{k\sigma^2}$$

Siendo $\boldsymbol{\beta}^* = (\beta_1, \beta_2, \ldots, \beta_k)'$ y $\mathbf{X}_c$ es la matriz "centrada" del modelo, definida por

$$\mathbf{X}_c = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1k} - \bar{x}_k \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2k} - \bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{i1} - \bar{x}_1 & x_{i2} - \bar{x}_2 & \cdots & x_{ik} - \bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{nk} - \bar{x}_k \end{bmatrix}$$

Estos cuadrados medios esperados indican que si el valor observado de F_0 es grande, es probable que al menos una $\beta_j \neq 0$. También se demuestra en el apéndice C.3 que si al menos una $\beta_j \neq 0$, entonces F_0 tiene una distribución F no central, con k y $n-k-1$ grados de libertad, y parámetro de no centralidad definido por

$$\lambda = \frac{\boldsymbol{\beta}^*\mathbf{X}_c'\mathbf{X}_c\boldsymbol{\beta}^*}{\sigma^2}$$

http://carlos2524.jimdo.com/

Este parámetro de no centralidad también indica que el valor observado de F_0 debe ser grande para que al menos una $\beta_j \neq 0$. Por consiguiente, para probar la hipótesis H_0: $\beta_1 = \beta_2 \cdots = \beta_k = 0$, se calcula el estadístico de prueba F_0 y se rechaza H_0 si

$$F_0 > F_{\alpha,k,n-k-1}$$

El procedimiento de prueba se resume normalmente en una **tabla de análisis de varianza**, como la tabla 3.4.

Una fórmula de cálculo para SS_R se deduce partiendo de

$$SS_{Res} = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} \tag{3.21}$$

y ya que

$$SS_T = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

se puede escribir la ecuación anterior en la forma

$$SS_{Res} = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} - \left[\hat{\beta}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] \tag{3.22}$$

o bien,

$$SS_{Res} = SS_T - SS_R \tag{3.23}$$

Por consiguiente, la **suma de cuadrados de la regresión** es

$$SS_R = \hat{\beta}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{3.24}$$

la **suma de cuadrados de residuales**, o **suma residual de cuadrados** es

$$SS_{Res} = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} \tag{3.25}$$

y la **suma total de cuadrados** es

$$SS_T = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{3.26}$$

TABLA 3.4 Análisis de varianza para determinar el significado en la regresión múltiple

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0
Regresión	SS_R	k	MS_R	MS_R/MS_{Res}
Residuales	SS_{Res}	$n-k-1$	MS_{Res}	
Total	SS_T	$n-1$		

http://carlos2524.jimdo.com/

Ejemplo 3.3 *Datos del tiempo de entrega*

Se probará el significado de una regresión con los datos de tiempo de entrega del ejemplo 3.1. Algunas cantidades numéricas que se requieren se calcularon en el ejemplo 3.2. Nótese que

$$SS_{\mathrm{T}} = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$= 18\,310.6290 - \frac{(559.60)^2}{25} = 5\,784.5426$$

$$SS_{\mathrm{R}} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$= 18\,076.9030 - \frac{(559.60)^2}{25} = 5\,550.8166$$

y

$$SS_{\mathrm{Res}} = SS_{\mathrm{T}} - SS_{\mathrm{R}}$$

$$= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 233.7260$$

El análisis de varianza se muestra en la tabla 3.5. Para probar H_0: $\beta_1 = \beta_2 = 0$, se calcula el estadístico

$$F_0 = \frac{MS_{\mathrm{R}}}{MS_{\mathrm{Res}}} = \frac{2\,775.4083}{10.6239} = 261.24$$

Como el valor P es muy pequeño, la conclusión es que el tiempo de entrega se relaciona con el volumen de entrega y/o con la distancia. Sin embargo, eso no implica necesariamente que la relación que se encontró sea adecuada para predecir el tiempo de entrega en función del volumen y de la distancia. Se requieren más pruebas de adecuación del modelo.

TABLA 3.5 Prueba para la significancia de la regresión del ejemplo 3.3

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Regresión	5 550.8166	2	2 775.4083	261.24	4.7×10^{-16}
Residuales	233.7260	22	10.6239		
Total	5 784.5426	24			

Resultados con SAS

La parte superior del reporte de SAS (rutina SAS PROC REG) de la figura 3.6 presenta también el análisis de varianza para probar el significado de la regresión. Salvo el redondeo, los resultados concuerdan con los de la tabla 3.5.

http://carlos2524.jimdo.com/

R^2 y R^2 ajustada

Otras dos maneras de evaluar la adecuación general del modelo son los estadísticos R^2 y R^2 ajustada; esta última se representa por R^2_{Adj}. Los resultados de SAS de la figura 3.6 reportan la R^2 para el modelo de regresión múltiple de los datos de tiempo de entrega como $R^2 = 0.9596$. En el ejemplo 2.9, donde sólo se usó el único regresor CASES (cajas), el valor de R^2 era menor, de $R^2 = 0.9305$ (véase la Fig. 2.13). En general, R^2 aumenta siempre, cuando se agrega un regresor al modelo, independientemente del valor de la contribución de esa variable. En consecuencia, es difícil juzgar si un aumento de R^2 dice en realidad algo importante.

Algunas personas que trabajan con modelos de regresión prefieren usar el estadístico **R^2 ajustada**, que se define como sigue:

$$R^2_{\text{Adj}} = 1 - \frac{SS_{\text{Res}}/(n-p)}{SS_{\text{T}}/(n-1)} \tag{3.27}$$

En vista de que $SS_{\text{Res}}/(n-p)$ es el cuadrado medio de residuales y $SS_{\text{T}}/(n-1)$ es constante, independientemente de cuántas variables hay en el modelo, R^2_{Adj} sólo aumentará al agregar una variable al modelo si esa adición reduce el cuadrado medio residual. Los resultados SAS PROC REG (Fig. 3.6) indican que $R^2_{\text{Adj}} = 0.9559$ para el modelo con dos variables, mientras que para el modelo de regresión lineal simple sólo con CASES, $R^2_{\text{Adj}} = 0.9275$ (Fig. 2.13). Por consiguiente, se llegaría a la conclusión que al agregar DISTANCE (distancia) al modelo se obtiene una reducción significativa de la variabilidad total.

En los capítulos posteriores, al describir **la construcción de modelos** y la **selección de variables** será útil contar con un procedimiento que pueda evitar un **sobreajuste del modelo**, esto es, agregar términos que son innecesarios. La R^2 ajustada penaliza la adición de términos que no son útiles, además que es ventajoso para evaluar y comparar los modelos posibles de regresión.

3.3.2 Pruebas sobre coeficientes individuales de regresión

Una vez determinado que al menos uno de los regresores es importante, la pregunta lógica es ¿cuál(es) sirve(n) de ellos? Si se agrega una variable a un modelo de regresión, la suma de cuadrados de la regresión aumenta, y la suma de cuadrados residuales disminuye. Se debe decidir si el aumento de la suma de cuadrados de la regresión es suficiente para garantizar el uso del regresor adicional en el modelo. La adición de un regresor también aumenta la varianza del valor ajustado $\hat{y}$, por lo que se debe tener cuidado de incluir sólo regresores que tengan valor para explicar la respuesta. Además, si se agrega un regresor no importante se puede aumentar el cuadrado medio de residuales, y con eso se disminuye la utilidad del modelo.

Las hipótesis para probar la significancia de cualquier coeficiente individual de regresión, como por ejemplo β_j, son

$$\begin{aligned} H_0&: \beta_j = 0 \\ H_1&: \beta_j \neq 0 \end{aligned} \tag{3.28}$$

http://carlos2524.jimdo.com/

Si no se rechaza H_0: $\beta_j = 0$, quiere decir que se puede eliminar el regresor x_j del modelo. El **estadístico de prueba** para esta hipótesis es

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{\text{se}\left(\hat{\beta}_j\right)} \tag{3.29}$$

donde C_{jj} es el elemento diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$ que corresponde a $\hat{\beta}_j$. Se rechaza la hipótesis nula H_0: $\beta_j = 0$ si $|t_0| > t_{\alpha/2,\, n-k-1}$. Nótese que ésta es en realidad una **prueba parcial** o **marginal**, porque el coeficiente de regresión $\hat{\beta}_j$ depende de todas las demás variables regresoras x_i $(i \neq j)$, que hay en el modelo. Así, se trata de una prueba de la **contribución** de x_j **dados los demás regresores del modelo**.

Ejemplo 3.4 *Datos del tiempo de entrega*

Para ilustrar el procedimiento se usarán los datos de tiempos de entrega del ejemplo 3.1. Supondremos que se desea evaluar la importancia de la variable regresora DISTANCE (distancia, x_{12}) dado que el regresor CASES (cajas, x_1) está en el modelo. Las hipótesis son:

$$H_0\colon \beta_2 = 0$$
$$H_1\colon \beta_2 \neq 0$$

El elemento de la diagonal principal de $(\mathbf{X}'\mathbf{X})^{-1}$ que corresponde a β_2 es $C_{22} = 0.00000123$, por lo que el estadístico t (3.29) es

$$t_0 = \frac{\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 C_{22}}} = \frac{0.01438}{\sqrt{(10.6239)(0.00000123)}} = 3.98$$

En vista de que $t_{0.025,\,22} = 2.074$, se rechaza H_0: $\beta_2 = 0$, y la conclusión es que el regresor DISTANCE, o x_2, contribuye en forma significativa al modelo, dado que CASES, o x_1, ya está también en el modelo.

También se puede determinar directamente la contribución de la suma de cuadrados de un regresor en la regresión, por ejemplo de x_j, dado que otros regresores x_i $(i \neq j)$, están ya en el modelo; para eso se usa el **método de suma extra de cuadrados**. Con este procedimiento también se puede investigar la contribución de un **subconjunto** de las variables regresoras para el modelo.

Considérese el modelo de regresión con k regresores

$$\mathbf{y} = \mathbf{X}\,\beta + \varepsilon$$

donde y es de $n \times 1$, $\mathbf{X}$ es de $n \times p$, β es de $p \times 1$, ε es de $n \times 1$, y $p = k + 1$. Se desea determinar si algún subconjunto de $r < k$ regresores contribuye en forma significativa al modelo de regresión. Sea seccionado como sigue el vector de los coeficientes de regresión:

$$\beta = \left[\frac{\beta_1}{\beta_2}\right]$$

http://carlos2524.jimdo.com/

donde β_1 es de $(p-r) \times 1$ y β_2 es de $r \times 1$. Se desean probar las hipótesis

$$H_0: \boldsymbol{\beta}_2 = \mathbf{0}$$
$$H_1: \boldsymbol{\beta}_2 \neq \mathbf{0} \tag{3.30}$$

Este modelo se puede escribir como sigue:

$$\mathbf{y} = \mathbf{X}\,\beta + \varepsilon = \mathbf{X}_1\,\beta_1 + \mathbf{X}_2\,\beta_2 + \varepsilon \tag{3.31}$$

en el que la matriz $\mathbf{X}_1$ de $n \times (p-r)$ representa a las columnas de $\mathbf{X}$ asociadas con β_1 y la matriz $\mathbf{X}_2$, de $n \times r$, representa a las columnas de $\mathbf{X}$ asociadas con β_2. A éste se le llama el **modelo completo**.

Para el modelo completo, se sabe que $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. La suma de cuadrados de regresión para este modelo es

$$SS_{\mathrm{R}}(\beta) = \hat{\beta}'\,\mathbf{X}'\mathbf{y} \quad (p \text{ grados de libertad})$$

y

$$MS_{\mathrm{Res}} = \frac{\mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y}}{n-p}$$

Para determinar la contribución de los términos de β_2 a la regresión se ajusta el modelo suponiendo que es cierta la hipótesis nula H_0: $\beta_2 = \mathbf{0}$. Este **modelo reducido** es

$$\mathbf{y} = \mathbf{X}_1\,\beta_1 + \varepsilon \tag{3.32}$$

El estimador de β_1 por mínimos cuadrados en el modelo reducido es $\hat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$. La suma de cuadrados de la regresión es

$$SS_{\mathrm{R}}(\beta_1) = \hat{\beta}_1\mathbf{X}_1'\mathbf{y} \ (p-r \text{ grados de libertad}) \tag{3.33}$$

La suma de cuadrados de la regresión debida a β_2 dado que β_1 ya está en el modelo es

$$SS_{\mathrm{R}}(\beta_2 \mid \beta_1) = SS_{\mathrm{R}}(\beta) - SS_R(\beta_1) \tag{3.34}$$

con $p-(p-r) = r$ grados de libertad. Esta suma de cuadrados se llama **suma extra de cuadrados debida a β_2**, porque mide el aumento de la suma de cuadrados de la regresión debida a agregar los regresores $x_{k-r+1}, x_{k-r+2}, \ldots, x_k$ a un modelo que ya contiene $x_1, x_2, \ldots, x_{k-r}$. Ahora, $SS_{\mathrm{R}}(\beta_2 \mid \beta_1)$ es independiente de MS_{Res}, y se puede probar la hipótesis nula $\beta_2 = \mathbf{0}$ mediante el estadístico

$$F_0 = \frac{SS_{\mathrm{R}}(\beta_2 \mid \beta_1)/r}{MS_{\mathrm{Res}}} \tag{3.35}$$

Si $\beta_2 \neq \mathbf{0}$, entonces F_0 sigue una distribución F no central, con parámetro de no centralidad igual a

$$\lambda = \frac{1}{\sigma^2}\beta_2'\mathbf{X}_2'\left[\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\right]\mathbf{X}_2\,\beta_2$$

http://carlos2524.jimdo.com/

Este resultado es muy importante. Si hay multicolinealidad en los datos, hay casos en los que β_2 es definidamente distinto de cero, pero esta prueba en realidad casi no tiene potencia (capacidad para indicar esta diferencia) porque hay una relación casi colineal entre $\mathbf{X}_1$ y $\mathbf{X}_2$. En este caso, λ es casi cero aun cuando β_2 sea realmente importante. Esta relación también hace destacar que la máxima potencia de la prueba se alcanza cuando $\mathbf{X}_1$ y $\mathbf{X}_2$ son ortogonales entre sí. Por ortogonales se entiende que $\mathbf{X}_2'\mathbf{X}_1 = \mathbf{0}$.

Si $F_0 > F_{\alpha,r,n-p}$, se rechaza H_0 y se concluye que al menos uno de los parámetros en β_2 es distinto de cero, y en consecuencia que al menos uno de los regresores $x_{k-r+1}, x_{k-r+2}, \ldots, x_k$ en $\mathbf{X}_2$ contribuyen en forma significativa al modelo de regresión. Algunos autores llaman la prueba (3.35) **prueba parcial F**, o **prueba F parcial**, porque mide la contribución de los regresores en $\mathbf{X}_2$, dado que los demás regresores en $\mathbf{X}_1$ ya están en el modelo. Para ilustrar la utilidad de este procedimiento, considérese el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Las sumas de cuadrados

$$SS_R(\beta_1 | \beta_0, \beta_2, \beta_3)$$
$$SS_R(\beta_2 | \beta_0, \beta_1, \beta_3)$$

y

$$SS_R(\beta_3 | \beta_0, \beta_1, \beta_2)$$

son sumas de cuadrados de un grado de libertad que miden la contribución de cada regresor x_j, $j = 1, 2, 3$, al modelo, dado que todos los demás regresores ya estaban en él. Esto es, se evalúa la ventaja de agregar x_j a un modelo que no incluía a este regreso. En general, se puede determinar

$$SS_R(\beta_j | \beta_0, \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k), \quad 1 \le j \le k$$

que es el aumento en la suma de cuadrados de regresión, debido a agregar x_j a un modelo que ya contiene $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$. Hay quienes creen de utilidad imaginar que esto mide la **contribución de x_j como si fuera la última variable agregada al modelo**.

Se puede demostrar que la prueba F parcial sobre una variable única x_j equivale a la prueba t en (3.29). Sin embargo, la prueba F parcial es un procedimiento más general, porque se puede medir el efecto de **conjuntos** de variables. En el capítulo 9 se mostrará cómo la prueba F parcial desempeña un papel importante en la **formación de modelos**, esto es, en la búsqueda del mejor conjunto de regresores que se deben usar en el modelo.

El método de la suma extra de cuadrados se puede usar para probar hipótesis acerca de cualquier **subconjunto** de variables regresoras que parezcan razonables para el problema que se esté analizando. A veces se ve que hay una jerarquía u ordenamiento natural de los regresores, y eso forma la base de una prueba. Por ejemplo, en el polinomio cuadrático

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

Se podría desear calcular

$$SS_R(\beta_1, \beta_2 | \beta_0)$$

http://carlos2524.jimdo.com/

que mediría la contribución de los términos de primer orden al modelo, y

$$SS_R(\beta_{12}, \beta_{11}, \beta_{22} | \beta_0, \beta_1, \beta_2)$$

que mediría la contribución de términos de segundo orden a un modelo que ya incluyera los términos de primer orden.

Cuando se imagina uno la adición de regresores uno por uno a un modelo, y se examina la contribución del regresor agregado en cada paso dados todos los regresores que se agregaron antes, se puede partir la suma de cuadrados de la regresión en componentes marginales de un solo grado de libertad. Por ejemplo, considérese el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

con la identidad del correspondiente análisis de varianza

$$SS_T = SS_R(\beta_1, \beta_2, \beta_3 | \beta_0) + SS_{Res}$$

Se puede descomponer la suma de cuadrados de la regresión con tres grados de libertad como sigue:

$$SS_R(\beta_1, \beta_2, \beta_3 | \beta_0) = SS_R(\beta_1 | \beta_0) + SS_R(\beta_2 | \beta_1, \beta_0) + SS_R(\beta_3 | \beta_1, \beta_2, \beta_0)$$

en donde cada suma de cuadrados del lado derecho sólo tiene un grado de libertad. Nótese que el orden de los regresores en esos componentes marginales es arbitrario. Una partición alternativa de $SS_R(\beta_1, \beta_2, \beta_3 | \beta_0)$ es

$$SS_R(\beta_1, \beta_2, \beta_3 | \beta_0) = SS_R(\beta_2 | \beta_0) + SS_R(\beta_1 | \beta_2, \beta_0) + SS_R(\beta_3 | \beta_1, \beta_2, \beta_0)$$

Sin embargo, el método de la suma extra de cuadrados no siempre produce una partición de la suma de cuadrados de la regresión, porque en general

$$\begin{aligned} SS_R(\beta_1, \beta_2, \beta_3 | \beta_0) \neq\ & SS_R(\beta_1 | \beta_2, \beta_3, \beta_0) + SS_R(\beta_2 | \beta_1, \beta_3, \beta_0) \\ & + SS_R(\beta_3 | \beta_1, \beta_2, \beta_0) \end{aligned}$$

Ejemplo 3.5 *Datos del tiempo de entrega*

En los datos de tiempo de entrega de gaseosas del ejemplo 3.1, supóngase que se trata de investigar la contribución de la variable distancia (x_2) al modelo. Las hipótesis correspondientes son

$$\begin{aligned} H_0 &: \beta_2 = 0 \\ H_1 &: \beta_2 \neq 0 \end{aligned}$$

Para probar esas hipótesis se necesita la suma extra de cuadrados debida a β_2, que es

$$\begin{aligned} SS_R(\beta_2 | \beta_1, \beta_0) &= SS_R(\beta_1, \beta_2, \beta_0) - SS_R(\beta_1, \beta_0) \\ &= SS_R(\beta_1, \beta_2 | \beta_0) - SS_R(\beta_1, \beta_2 | \beta_0) \end{aligned}$$

http://carlos2524.jimdo.com/

De acuerdo con el ejemplo 3.3,

$$SS_R(\beta_1, \beta_2|\beta_0) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = 5\,550.8166 \text{ (2 grados de libertad)}$$

El modelo reducido $y = \beta_0 + \beta_1 x_1 + \varepsilon$ se ajustó en el ejemplo 2.9, y se obtuvo $\hat{y} = 3.3208 + 2.1762x_1$. La suma de cuadrados de regresión para este modelo es

$$\begin{aligned} SS_R(\beta_1|\beta_0) &= \hat{\beta}_1 S_{xy} = (2.1762)(2\,473.3440) \\ &= 5\,382.4077 \quad (1 \text{ grado de libertad}) \end{aligned}$$

Por consiguiente,

$$\begin{aligned} SS_R(\beta_2|\beta_1, \beta_0) &= 5\,550.8166 - 5\,382.4088 \\ &= 168.4078 \ (1 \text{ grado de libertad}) \end{aligned}$$

Es un aumento de la suma de cuadrados de la regresión, que se debe a agregar x_2 al modelo que ya contenía a x_1. Para probar H_0: $\beta_2 = 0$, se forma el estadístico de prueba

$$F_0 = \frac{SS_R(\beta_2|\beta_1, \beta_0)/1}{MS_{Res}} = \frac{168.4078/1}{10.6239} = 15.85$$

Obsérvese que la MS_{Res} del modelo **completo**, que contiene a x_1 y a x_2, se usa en el denominador del estadístico. Como $F_{0.05,\,1,\,22} = 4.30$, se rechaza H_0: $\beta_2 = 0$ y se concluye que la distancia (x_2) contribuye al modelo en forma significativa.

Como esta prueba F parcial implica a una sola variable, equivale a la prueba t. Para visualizarlo, recuérdese que la prueba t de H_0: $\beta_2 = 0$ dio como resultado el estadístico $t_0 = 3.98$. De acuerdo con el apéndice C.1, el cuadrado de una variable aleatoria t con ν grados de libertad es una variable aleatoria F con un grado de libertad en el numerador y ν en el denominador; entonces $t_0^2 = (3.98)^2 = 15.84 \simeq F_0$.

3.3.3 Caso especial de columnas ortogonales en X

Se tiene el modelo (3.31)

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \end{aligned}$$

El método de la suma extra de cuadrados permite medir el efecto de los regresores en $\mathbf{X}_2$, condicionado a los de $\mathbf{X}_1$, calculando $SS_R(\beta_2|\beta_1)$. En general, no se puede hablar de determinar la suma de cuadrados debida a β_2, que es $SS_R(\beta_2)$ sin tener en cuenta la dependencia de esta cantidad respecto a los regresores en $\mathbf{X}_1$. Sin embargo, si las columnas de $\mathbf{X}_1$ son **ortogonales** a las de $\mathbf{X}_2$, se puede determinar la suma de cuadrados debida a β_2, que está libre de cualquier dependencia respecto a los regresores de $\mathbf{X}_2$.

Para demostrar lo anterior, se forman las ecuaciones normales $(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$, para el modelo (3.31). Las ecuaciones normales son

$$\left[\begin{array}{c|c} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \hline \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{array}\right] \left[\begin{array}{c} \hat{\boldsymbol{\beta}}_1 \\ \hline \hat{\boldsymbol{\beta}}_2 \end{array}\right] = \left[\begin{array}{c} \mathbf{X}_1'\mathbf{y} \\ \hline \mathbf{X}_2'\mathbf{y} \end{array}\right]$$

http://carlos2524.jimdo.com/

Si las columnas de $\mathbf{X}_1$ son ortogonales a las de $\mathbf{X}_2$, entonces $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, y $\mathbf{X}_2'\mathbf{X}_1 = \mathbf{0}$, además que las ecuaciones normales se transforman en

$$\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 = \mathbf{X}_1'\mathbf{y}$$

$$\mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_2'\mathbf{y}$$

cuya solución es

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$$

$$\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}$$

Nótese que el estimador de $\boldsymbol{\beta}_1$ por mínimos cuadrados es $\hat{\boldsymbol{\beta}}_1$, independientemente de si $\mathbf{X}_2$ está en el modelo o no, y que el estimador de $\boldsymbol{\beta}_2$ por mínimos cuadrados es $\hat{\boldsymbol{\beta}}_2$, independientemente de si $\mathbf{X}_1$ está en el modelo o no.

La suma de cuadrados de la regresión, para el modelo completo, es

$$\begin{aligned} SS_R(\boldsymbol{\beta}) &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \\ &= \left[\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2'\right]\begin{bmatrix}\mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y}\end{bmatrix} \\ &= \hat{\boldsymbol{\beta}}'\mathbf{X}_1'\mathbf{y} + \boldsymbol{\beta}_2'\mathbf{X}_2'\mathbf{y} \\ &= \mathbf{y}'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} + \mathbf{y}'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y} \end{aligned} \tag{3.36}$$

Sin embargo, las ecuaciones normales forman dos conjuntos, y en cada uno de ellos se ve que

$$\begin{aligned} SS_R(\boldsymbol{\beta}_1) &= \hat{\boldsymbol{\beta}}_1'\mathbf{X}_1'\mathbf{y} = \mathbf{y}'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} \\ SS_R(\boldsymbol{\beta}_2) &= \hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y} = \mathbf{y}'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y} \end{aligned} \tag{3.37}$$

Al comparar la ecuación (3.37) con la (3.36) se ve que

$$SS_R(\boldsymbol{\beta}) = SS_R(\boldsymbol{\beta}_1) + SS_R(\boldsymbol{\beta}_2) \tag{3.38}$$

Por consiguiente,

$$SS_R(\boldsymbol{\beta}_1 \mid \boldsymbol{\beta}_2) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_2) \equiv SS_R(\boldsymbol{\beta}_1)$$

y

$$SS_R(\boldsymbol{\beta}_2 \mid \boldsymbol{\beta}_1) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_1) \equiv SS_R(\boldsymbol{\beta}_2)$$

Por lo anterior, $SS_R(\boldsymbol{\beta}_1)$ mide la contribución de los regresores en $\mathbf{X}_1$ al modelo en **forma incondicional**, y $SS_R(\boldsymbol{\beta}_2)$ mide la contribución **incondicional** de los regresores en $\mathbf{X}_2$ al modelo. Como se puede determinar en forma no ambigua el efecto de cada regresor, cuando los regresores son ortogonales, los experimentos de recolección de datos se suelen diseñar en tal forma que se obtengan variables ortogonales.

http://carlos2524.jimdo.com/

Como ejemplo de un modelo de regresión con regresores ortogonales, se examinará el modelo $y = \beta_0 + \beta_1 x_2 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$, en el que la matriz $\mathbf{X}$ es

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccc} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{array} \\ \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

Los niveles de los regresores corresponden al diseño factorial 2^3. Es fácil de ver que las columnas de $\mathbf{X}$ son ortogonales. Así, $SS_R(\beta_j)$, $j = 1, 2, 3$, mide la contribución del regresor x_j al modelo, independientemente de si alguno de los demás regresores se incluye en el ajuste.

3.3.4 Prueba de la hipótesis general lineal

Se pueden probar muchas hipótesis acerca de los coeficientes de regresión, si se usa un método unificado. El método de suma extra de cuadrados es un caso especial de este procedimiento. En el procedimiento más general, la suma de cuadrados con la que se calcula la hipótesis es como la diferencia de dos sumas de cuadrados de residuales. A continuación se describirá el procedimiento. Para conocer demostraciones y descripciones más detalladas, consúltese Graybill [1976], Searle [1971] o Seber [1977].

Supóngase que la hipótesis nula de interés se expresa en la forma H_0: $\mathbf{T}\beta = \mathbf{0}$, donde $\mathbf{T}$ es una matriz de constantes de $m \times p$, tal que sólo r de las m ecuaciones de $\mathbf{T}\beta = \mathbf{0}$ son independientes. El **modelo completo** (*FM*, de *full model*) es $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, siendo $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, y la suma de cuadrados de residuales, para el modelo completo es

$$SS_{\text{Res}}(FM) = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} \ (n - p \text{ grados de libertad})$$

Para obtener el **modelo reducido**, se usan las r ecuaciones independientes en $\mathbf{T}\beta = \mathbf{0}$ para calcular los r coeficientes de regresión en el modelo completo, en función de los $p - r$ coeficientes restantes de regresión. Esto conduce al modelo reducido $\mathbf{y} = \mathbf{Z}\gamma + \varepsilon$, por ejemplo, donde $\mathbf{Z}$ es una matriz de $n \times (p - r)$ y γ es un vector de $(p - r) \times 1$, de coeficientes desconocidos de regresión. El estimado de γ es

$$\hat{\gamma} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

y la suma de cuadrados de residuales, para el modelo reducido (*RM*, de *reduced model*), es

$$SS_{\text{Res}}(RM) = \mathbf{y}'\mathbf{y} - \hat{\gamma}'\mathbf{Z}'\mathbf{y} \ (n - p + r \text{ grados de libertad})$$

El modelo reducido contiene menos parámetros que el modelo completo, así que $SS_{\text{Res}}(RM) \geq SS_{\text{Res}}(FM)$. Para probar la hipótesis H_0: $\mathbf{T}\beta = \mathbf{0}$ se emplea la diferencia de sumas de cuadrados residuales

$$SS_{\text{H}} = SS_{\text{Res}}(RM) - SS_{\text{Res}}(FM) \tag{3.39}$$

http://carlos2524.jimdo.com/

con $n - p + r - (n - p) = r$ grados de libertad. En ella, SS_H se llama suma de cuadrados debida a la **hipótesis** H_0: $\mathbf{T}\beta = \mathbf{0}$. El estadístico de prueba para esta hipótesis es

$$F_0 = \frac{SS_H / r}{SS_{Res}(FM)/(n - p)} \tag{3.40}$$

Se rechaza H_0: $\mathbf{T}\beta = \mathbf{0}$ si $F_0 > F_{\alpha, r, n-p}$.

Ejemplo 3.6 *Prueba de igualdad de coeficientes de regresión*

Para probar la igualdad de los coeficientes de regresión se puede usar el método de la hipótesis lineal general. Se tiene el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Para el modelo completo (FM), $SS_{Res}(FM)$ tiene $n - p = n - 4$ grados de libertad. Se desea probar H_0: $\beta_1 = \beta_3$. Esta hipótesis se puede enunciar como H_0: $\mathbf{T}\beta = \mathbf{0}$, siendo

$$\mathbf{T} = [0, \quad 1, \quad 0, \quad -1]$$

un vector renglón de 1×4. Hay sólo una ecuación en $\mathbf{T}\beta = \mathbf{0}$, que es $\beta_1 - \beta_3 = 0$. Si se sustituye esta ecuación en el modelo completo, se obtiene el modelo reducido (RM)

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_1 x_3 + \varepsilon \\ &= \beta_0 + \beta_1 (x_1 + x_3) + \beta_2 x_2 + \varepsilon \\ &= \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \varepsilon \end{aligned}$$

donde $\gamma_0 = \beta_0$, $\gamma_1 = \beta_1 (= \beta_3)$, $z_1 = x_1 + x_3$, $\gamma_2 = \beta_2$ y $z_2 = x_2$. Al ajustar el modelo reducido, se calcularía la $SS_{Res}(RM)$ con $n - 4 + 1 = n - 3$ grados de libertad. La suma de cuadrados debidos a la hipótesis $SS_H = SS_{Res}(RM) - SS_{Res}(FM)$ tiene $n - 3 - (n - 4) = 1$ grado de libertad. El cociente F (3.40) es $F_0 = (SS_H/1) / [SS_{Res}(FM)/ (n - 4)]$. Nótese que esta hipótesis también se podría probar con el estadístico t:

$$t_0 = \frac{\hat{\beta}_1 - \hat{\beta}_3}{\text{se}\left(\hat{\beta}_1 - \hat{\beta}_3\right)} = \frac{\hat{\beta}_1 - \hat{\beta}_3}{\sqrt{\hat{\sigma}^2 (C_{11} + C_{33} - 2C_{13})}}$$

con $n - 4$ grados de libertad. Esto equivale a la prueba F.

Ejemplo 3.7

Supongamos que el modelo es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

http://carlos2524.jimdo.com/

y que se desea probar H_0: $\beta_1 = \beta_3$, $\beta_2 = 0$. Para enunciar lo anterior en forma de una hipótesis lineal general, sea

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Ahora hay dos ecuaciones en $\mathbf{T}\boldsymbol{\beta} = \mathbf{0}$, que son $\beta_1 - \beta_3 = 0$ y $\beta_2 = 0$. Estas ecuaciones producen el modelo reducido

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_1 x_3 + \varepsilon \\ &= \beta_0 + \beta_1(x_1 + x_3) + \varepsilon \\ &= \gamma_0 + \gamma_1 z_1 + \varepsilon \end{aligned}$$

En este ejemplo, $SS_{\text{Res}}(RM)$ tiene $n - 2$ grados de libertad, por lo que SS_{R} tiene $n - 2 - (n - 4) = 2$ grados de libertad. La relación F (3.40) es $F_0 = (SS_{\text{H}}/2) / [SS_{\text{Res}}(FM) / (n - 4)]$.

El estadístico de prueba (3.40) para la hipótesis lineal general se puede escribir en otra forma, como sigue:

$$F_0 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{T}'\left[\mathbf{T}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}'\right]^{-1}\mathbf{T}\hat{\boldsymbol{\beta}}/r}{SS_{\text{Res}}(FM)/(n-p)} \tag{3.41}$$

También se podría haber usado esta forma para desarrollar los procedimientos de prueba que se ilustraron en los ejemplos 3.6 y 3.7.

Hay una ligera extensión de la hipótesis lineal general, que a veces es útil; la presentamos a continuación:

$$\begin{aligned} H_0 &: \mathbf{T}\boldsymbol{\beta} = \mathbf{c} \\ H_1 &: \mathbf{T}\boldsymbol{\beta} \neq \mathbf{c} \end{aligned} \tag{3.42}$$

para la cual, el estadístico de prueba es

$$F_0 = \frac{(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{c})'\left[\mathbf{T}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}'\right]^{-1}\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{c}\ /r}{SS_{\text{Res}}(FM)/(n-p)} \tag{3.43}$$

Como de acuerdo con la hipótesis nula $\mathbf{T}\boldsymbol{\beta} = \mathbf{c}$, la distribución de F_0 en la ecuación (3.43) es $F_{r,n-p}$, se rechazaría H_0: $\mathbf{T}\boldsymbol{\beta} = \mathbf{c}$ si $F_0 > F_{\alpha,r,n-p}$. Esto es, el procedimiento de prueba es una prueba F unilateral superior. Nótese que el numerador de la ecuación (3.43) expresa una medida de la distancia al cuadrado entre $\mathbf{T}\boldsymbol{\beta}$ y $\mathbf{c}$, normalizada por la matriz de covarianza de $\mathbf{T}\hat{\boldsymbol{\beta}}$.

Para ilustrar cómo se usa este procedimiento ampliado, considérese el caso descrito en el ejemplo 3.6, y supóngase que se desea probar

$$H_0: \beta_1 - \beta_3 = 2$$

Es claro que $\mathbf{T} = [0, 1, 0, -1]$ y $\mathbf{c} = [2]$. Para conocer otros usos de este procedimiento, véase el problema 3.20.

http://carlos2524.jimdo.com/

Por último, si no se puede rechazar la hipótesis H_0: $\mathbf{T}\beta = \mathbf{0}$ (H_0: $\mathbf{T}\beta = \mathbf{c}$), será razonable estimar a β sujeta a la restricción impuesta por la hipótesis nula. No es probable que el usual estimador de mínimos cuadrados satisfaga en forma automática la restricción. En esos casos se podrá usar un **estimador restringido de mínimos cuadrados**. Véase el problema 3.28.

3.4 INTERVALOS DE CONFIANZA EN REGRESIÓN MÚLTIPLE

Los intervalos de confianza de los coeficientes de regresión individuales, y los intervalos de confianza de la respuesta media, para niveles específicos de los regresores, juegan el mismo papel importante que en la regresión lineal simple. En esta sección se desarrollan los intervalos de confianza, uno por uno, para estos casos. También se presentarán en forma breve los intervalos simultáneos de confianza para los coeficientes de regresión.

3.4.1 Intervalos de confianza de los coeficientes de regresión

Para construir **estimados de intervalo de confianza** de los coeficientes de regresión β_j, se continuará suponiendo que los errores ε_i están distribuidos normal e independientemente, con promedio cero y varianza σ^2. En consecuencia, las observaciones y_i están distribuidas en forma normal e independiente, con media $\beta_0 + \Sigma_{j=1}^{k} \beta_j x_{ij}$, y varianza σ^2. Como el estimador $\hat{\beta}$ por mínimos cuadrados es una combinación lineal de las observaciones, también está distribuido normalmente, con vector medio β y matriz de covarianza $\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$. Esto implica que la distribución marginal de cualquier coeficiente de regresión $\hat{\beta}_j$ es normal, con media β_j y varianza $\sigma^2 C_{jj}$, donde C_{jj} es el j-ésimo elemento diagonal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. En consecuencia, cada una de los estadísticos

$$\frac{\hat{\beta}_j - \hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}, \quad j = 0, 1, \ldots, k \tag{3.44}$$

se distribuye como t, con $n - p$ grados de libertad, donde $\hat{\sigma}^2$ es el estimado de la varianza de error obtenida con la ecuación (3.18).

De acuerdo con el resultado de la ecuación (3.44), se puede definir un **intervalo de confianza de 100(1 – α) por ciento para el coeficiente de regresión** β_j, $j = 0, 1, \ldots, k$, como sigue:

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \le \beta_j \le \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \tag{3.45}$$

Recuérdese que la cantidad

$$\text{se}\left(\hat{\beta}_j\right) = \sqrt{\hat{\sigma}^2 C_{jj}} \tag{3.46}$$

el error estándar del coeficiente de regresión $\hat{\beta}_j$.

http://carlos2524.jimdo.com/

Ejemplo 3.8 *Datos del tiempo de entrega*

Se calculará un intervalo de confianza de 95% para el parámetro β_1 en el ejemplo 3.1. El estimado puntual de β_1 es $\hat{\beta}_1 = 1.61591$; el elemento diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$ que corresponde a β_1 es $C_{11} = 0.00274378$, y $\hat{\sigma}^2 = 10.6239$ (de acuerdo con el Ej. 3.2). Se aplica la ecuación (3.45) y se ve que

$$\hat{\beta}_1 - t_{0.025,22}\sqrt{\hat{\sigma}^2 C_{11}} \le \beta_1 \le \hat{\beta}_1 + t_{0.025,22}\sqrt{\hat{\sigma}^2 C_{11}}$$

$$1.61591 - (2.074)\sqrt{(10.6239)(0.00274378)}$$

$$\le \beta_1 \le 1.61591 + (2.074)\sqrt{(10.6239)(0.00274378)}$$

$$1.61591 - (2.074)(0.17073) \le \beta_1 \le 1.61591 + (2.074)(0.17073)$$

y el intervalo de confianza de 95% para β_1 es

$$1.26181 \le \beta_1 \le 1.97001$$

Obsérvese que los resultados de SAS, en la figura 3.6, mencionan el error estándar de cada coeficiente de regresión. Esto hace que en la práctica sea muy fácil establecer esos intervalos.

3.4.2 Estimación del intervalo de confianza de la respuesta media

Se puede establecer un intervalo de confianza para la respuesta media en determinado punto, como $x_{01}, x_{02}, \ldots, x_{0k}$. Defínase el vector $\mathbf{x}_0$ como sigue:

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0k} \end{bmatrix}$$

El valor ajustado en este punto es

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} \tag{3.47}$$

Es un estimador insesgado de $E(y|\mathbf{x}_0)$, porque $E(\hat{y}_0) = \mathbf{x}_0'\boldsymbol{\beta} = E(y|\mathbf{x}_0)$, y la varianza de $\hat{y}_0$ es

$$\text{Var}(\hat{y}_0) = \sigma^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \tag{3.48}$$

Por consiguiente, un **intervalo de confianza de 100(1 – α) por ciento de la respuesta media** en el punto $x_{01}, x_{02}, \ldots, x_{0k}$ es

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$$

$$\le E(y|\mathbf{x}_0) \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \tag{3.49}$$

Es la generalización de la ecuación (2.43) al caso de la regresión múltiple.

http://carlos2524.jimdo.com/

Ejemplo 3.9 *Datos del tiempo de entrega*

El embotellador de gaseosas del ejemplo 3.1 quiere establecer un intervalo de confianza de 95% para el tiempo medio de entrega, para una tienda donde se requieran $x_1 = 8$ cajas, y donde la distancia es $x_2 = 275$ pies. Entonces,

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix}$$

El valor ajustado en este punto se calcula con la ecuación (3.47) como sigue:

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = [1 \;\; 8 \;\; 275]\begin{bmatrix} 2.34123 \\ 1.61591 \\ 0.01438 \end{bmatrix} = 19.22 \text{ minutos}$$

La varianza de $\hat{y}_0$ se estima con

$$\begin{aligned}\hat{\sigma}^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 &= 10.6239[1 \;\; 8 \;\; 275] \\ &\quad \times \begin{bmatrix} 0.11321518 & -0.00444859 & -0.00008367 \\ -0.00444859 & 0.00274378 & -0.00004786 \\ -0.00008367 & -0.00004786 & 0.00000123 \end{bmatrix}\begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix} \\ &= 10.6239(0.05346) = 0.56794\end{aligned}$$

Por consiguiente, un intervalo de confianza de 95% para el tiempo promedio de entrega en este punto se calcula con la ecuación (3.49):

$$19.22 - 2.074\sqrt{0.56794} \le E(y|x_0) \le 19.22 + 2.074\sqrt{0.56794}$$

que se reduce a

$$17.66 \le E(y|x_0) \le 20.78$$

El 95% de esos intervalos contienen el tiempo real de entrega.

3.4.3 Intervalos simultáneos de confianza para coeficientes de regresión

Se han descrito los procedimientos para establecer diversos tipos de intervalos de confianza y de predicción para el modelo de regresión lineal. Se ha hecho notar que éstos son intervalos de uno por uno, esto es, son los tipos usuales de intervalo de confianza o de predicción, en donde el coeficiente de confianza $1 - \alpha$ indica la proporción de estimaciones correctas que resulta cuando se seleccionan muestras aleatorias repetidas, y para cada muestra se establece el estimado adecuado de intervalo. En algunos problemas se necesita construir varios intervalos de confianza o de predicción con los mismos datos de muestra. En esos casos, el analista suele interesarse en la especificación de un coeficiente de confianza que se aplique **en forma simultánea**, o al mismo tiempo, a todo el **conjunto** de estimados

http://carlos2524.jimdo.com/

de intervalo. Un conjunto de intervalos de confianza o de predicción que son todos ciertos en forma simultánea, con $1 - \alpha$ de probabilidad, se llama conjunto de intervalos **simultáneos** o **conjuntos de confianza** o **de predicción**.

Por ejemplo, se tiene un modelo de regresión lineal simple. Supongamos que el analista desea sacar inferencias acerca de la ordenada al origen β_0 y la pendiente β_1, una posibilidad sería establecer intervalos de confianza, por ejemplo de 95%, para ambos parámetros, sin embargo, si esos estimados son independientes, la probabilidad de que ambas afirmaciones sean correctas es $(0.95)^2 = 0.9025$. Así, no se tiene un nivel de confianza de 95% asociado con ambas afirmaciones. Además, como los intervalos se establecen usando el mismo conjunto de datos muestrales, no son independientes. Esto introduce mayor complicación en la determinación del nivel de confianza para el conjunto de afirmaciones.

Es relativamente fácil definir una **región de confianza conjunta** para los parámetros β del modelo de regresión múltiple. Se puede demostrar que

$$\frac{(\hat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)}{pMS_{\text{Res}}} \sim F_{p,n-p}$$

y eso implica que

$$P\left\{\frac{(\hat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)}{pMS_{\text{Res}}} \leq F_{\alpha,n-p}\right\} = 1 - \alpha$$

En consecuencia, una **región de confianza conjunta** de $100(1 - \alpha)$ por ciento, para todos los parámetros en β es

$$\frac{(\hat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)}{pMS_{\text{Res}}} \leq F_{\alpha,p,n-p} \tag{3.50}$$

Esta desigualdad describe una región de forma elíptica. La determinación de esta región de confianza conjunta es relativamente directa para la regresión lineal simple, con $p = 2$. Es más difícil para $p = 3$, y se necesitaría un programa especial de gráficas en 3D.

Ejemplo 3.10 *Datos del propelente de reacción*

Para el caso de la regresión lineal simple, se puede demostrar que la ecuación (3.50) se reduce a

$$\frac{n(\hat{\beta}_0 - \beta_0)^2 + 2\sum_{i=1}^{n} x_i(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) + \sum_{i=1}^{n} x_i^2(\hat{\beta}_1 - \beta_1)^2}{2\,MS_{\text{Res}}} \leq F_{\alpha,2,n-2}$$

Para ilustrar la determinación de esta región de confianza, se usarán los datos del combustible para cohetes del ejemplo 2.1. Se determinará una región de confianza de 95% para β_0 y β_1. Como $\hat{\beta}_0 = 2\,627.82$, $\hat{\beta}_1 = -37.15$, $\Sigma_{i=1}^{20} x_i^2 = 4\,677.69$, $MS_{\text{Res}} = 9\,244.59$ y $F_{0.05,2,18} = 3.55$, se sustituyen en la ecuación anterior y se obtiene, como contorno de la elipse,

http://carlos2524.jimdo.com/

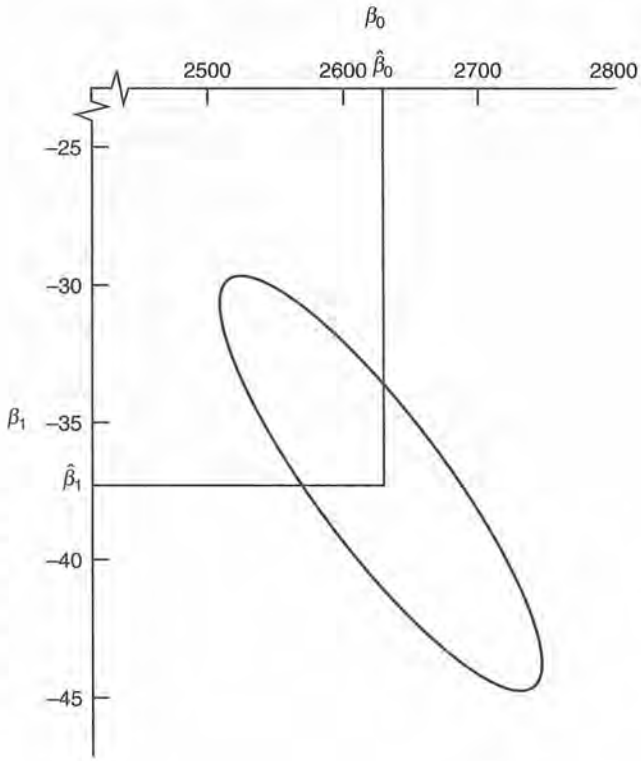


Figura 3.9 Región de confianza conjunta de 95% de β_0 y β_1, para los datos del propelente de reacción.

$$\left[20(2\,627.82-\beta_0)^2+2(267.25)(2\,627.82-\beta_0)(-37.15-\beta_1)\right.$$
$$\left.+(4\,677.69)(-37.15-\beta_1)^2\right]/[2(9\,244.59)]=3.55$$

En la figura 3.9 se muestra la región de confianza conjunta. Nótese que esta elipse no es paralela al eje β_1. Su inclinación es una función de la covarianza entre $\hat{\beta}_0$ y $\hat{\beta}_1$, que es $-\bar{x}\sigma^2/S_{xx}$. Una covarianza positiva implica que los errores en los estimados puntuales de β_0 y β_1 probablemente tengan la misma dirección, mientras que una covarianza negativa indica que esos errores probablemente tengan direcciones opuestas. En nuestro ejemplo, $\bar{x}$ es positiva por lo que $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$ es negativa. Así, si el estimado de la pendiente es muy inclinado (β_1 está sobreestimada), es probable que el estimado de la ordenada al origen sea demasiado pequeño (β_0 está subestimada). El alargamiento de la región depende de los tamaños relativos de las varianzas de β_0 y β_1. En general, si la elipse está alargada en dirección (por ejemplo) de β_0, ello implica que β_0 no está estimada con tanta precisión como la de β_1. Este caso es el de nuestro ejemplo.

Hay otro método general para obtener estimados simultáneos de intervalo para los parámetros de un modelo de regresión lineal. Esos intervalos de confianza se pueden establecer mediante

$$\hat{\beta}_j \pm \Delta\,\text{se}\left(\hat{\beta}_j\right), \quad j=0,1,\ldots,k \tag{3.51}$$

donde la constante Δ se escoge de tal manera que se obtenga una probabilidad especificada de que todos los intervalos sean correctos.

http://carlos2524.jimdo.com/

Hay varios métodos que se pueden emplear para determinar Δ en (3.51). Uno es el **método de Bonferroni**, en el que se define $\Delta = t_{\alpha/2p,\, n-p}$ de tal manera que la (3.51) se transforme en

$$\hat{\beta}_j \pm t_{\alpha/2p,\, n-p}\, \mathrm{se}\left(\hat{\beta}_j\right), \quad j = 0, 1, \ldots, k \tag{3.52}$$

La probabilidad mínima es $1 - \alpha$ de que todos los intervalos sean correctos. Obsérvese que los **intervalos de confianza de Bonferroni** se parecen algo a los intervalos ordinarios de uno por uno, basados en la distribución t, pero cada intervalo de Bonferroni tiene un coeficiente de confianza de $1 - \alpha/p$, en lugar de $1 - \alpha$.

Ejemplo 3.11 *Datos del propelente de reacción*

Se calcularán los intervalos de confianza conjuntos para β_0 y β_1 con los datos del combustible de cohetes en el ejemplo 2.1, estableciendo para cada uno un intervalo de confianza de 95%. Ya que

$$\hat{\beta}_0 = 2\,627.822, \quad \mathrm{se}(\hat{\beta}_0) = 44.184$$
$$\hat{\beta}_1 = -37.154, \quad \mathrm{se}(\hat{\beta}_1) = 2.889$$

y $t_{0.05/2,\,18} = t_{0.025,\,18} = 2.101$, los intervalos de confianza conjuntos son

$$\begin{aligned} \hat{\beta}_0 - t_{0.0125,18}\mathrm{se}\left(\hat{\beta}_0\right) \le \beta_0 &\le \hat{\beta}_0 + t_{0.0125,18}\mathrm{se}\left(\hat{\beta}_0\right) \\ 2\,627.822 - (2.445)(44.184) \le \beta_0 &\le 2\,627.822 + (2.445)(44.184) \\ 2\,519.792 \le \beta_0 &\le 2\,735.852 \end{aligned}$$

y

$$\begin{aligned} \hat{\beta}_1 - t_{0.0125,18}\mathrm{se}\left(\hat{\beta}_1\right) \le \beta_1 &\le \hat{\beta}_1 + t_{0.0125,18}\mathrm{se}\left(\hat{\beta}_1\right) \\ -37.154 - (2.445)(2.889) \le \beta_1 &\le -37.154 + (2.445)(2.889) \\ -44.218 \le \beta_1 &\le -30.090 \end{aligned}$$

Se llega a la conclusión, con 90% de confianza, que este procedimiento conduce a estimados correctos de intervalo para los dos parámetros.

La elipse de confianza siempre es un procedimiento más eficiente que el método de Bonferroni, porque el volumen de la elipse siempre es menor que el del espacio que abarcan los intervalos de Bonferroni. Sin embargo, son más fáciles de establecer estos intervalos.

Con frecuencia se requiere, para establecer los intervalos de confianza de Bonferroni, niveles de significancia que no aparecen en las tablas t usuales. Muchas de las calculadoras modernas tienen el cálculo de valores de $t_{\alpha,\,\nu}$ como función de biblioteca.

El método de Bonferroni no es el único para establecer Δ en (3.51). Entre otros métodos están el **método S de Scheffé** (véase Scheffé [1953, 1959]), en el cual

$$\Delta = (2F_{\alpha,\,p,\,n-p})^{1/2}$$

http://carlos2524.jimdo.com/

y el **procedimiento del módulo t máximo** (véase Hahn [1972] y Hahn y Hendrickson [1971]), en el cual

$$\Delta = u_{\alpha, p, n-p}$$

siendo $u_{\alpha, p, n-p}$ el punto de la cola superior de la distribución de valor absoluto máximo de dos variables aleatorias independientes t de Student, cada una basada en $n - 2$ grados de libertad. Una forma obvia de comparar estas tres técnicas es en términos de las longitudes de los intervalos de confianza que generan. En general, los intervalos de Bonferroni son más cortos que los de Scheffé, y los intervalos del módulo t máximo son más cortos que los de Bonferroni.

3.5 PREDICCIÓN DE NUEVAS OBSERVACIONES

Con el modelo de regresión se pueden predecir observaciones futuras de y que correspondan a determinados valores de las variables regresoras, por ejemplo $x_{01}, x_{02}, \ldots, x_{0k}$. Si $\mathbf{x}_0' = [1, x_{01}, x_{02}, \ldots, x_{0k}]$, entonces un **estimado puntual de la observación futura** y_0 en el punto $x_{01}, x_{02}, \ldots, x_{0k}$ es

$$\hat{y}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}} \tag{3.53}$$

Un **intervalo de predicción de 100(1 – α) por ciento** para esta futura observación es

$$\hat{y}_0 - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 \left(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right)}$$
$$\leq y_0 \leq \hat{y}_0 + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 \left(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right)} \tag{3.54}$$

Es una generalización de intervalo de predicción para una futura observación en la regresión lineal simple, ecuación (2.45).

Ejemplo 3.12 *Datos del tiempo de entrega*

Supongamos que el embotellador de gaseosas del ejemplo 3.1 desea establecer un intervalo de predicción de 95% para el tiempo de entrega en una tienda, donde se entregan $x_1 = 8$ cajas, y la distancia que camina el repartidor es $x_2 = 275$ pies. Nótese que $\mathbf{x}_0' = [1, 8, 275]$, y que el estimado puntual para el tiempo de entrega es $\hat{y}_0 = \mathbf{x}_0' = 19.22$ minutos. También, en el ejemplo 3.9 se calculó $\mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = 0.05346$. Entonces, de acuerdo con la ecuación (3.54),

$$19.22 - 2.074\sqrt{10.6239(1 + 0.05346)}$$
$$\leq y_0 \leq 19.22 + 2.074\sqrt{10.6239(1 + 0.05346)}$$

y el intervalo de predicción de 95% es

$$12.28 \leq y_0 \leq 26.16$$

http://carlos2524.jimdo.com/

3.6 EXTRAPOLACIÓN OCULTA EN LA REGRESIÓN MÚLTIPLE

Al pronosticar nuevas respuestas y para estimar la respuesta media en un punto dado x_{01}, $x_{02}, \ldots, x_{0k}$, se debe tener cuidado para no **extrapolar** fuera de la región que contiene las observaciones originales. Es muy posible que un modelo que ajusta bien en la región de los datos originales funcione mal fuera de ella. En la regresión múltiple es fácil extrapolar en forma inadvertida, porque los niveles de los regresores $(x_{i1}, x_{i2}, \ldots, x_{ik})$, $i = 1, 2, \ldots, n$ definen, en conjunto, la región que contiene los datos. Por ejemplo, se tiene la figura 3.10, que ilustra la región que contiene los datos originales para un modelo de dos regresores. Nótese que el punto (x_{01}, x_{02}) está dentro de los intervalos de ambos regresores x_1 y x_2, pero fuera de la región de los datos originales. Así, tanto la predicción del valor de una nueva observación, o la estimación de la respuesta media en este punto es una extrapolación del modelo original de regresión.

En vista de que si sólo se comparan los valores de las x para un nuevo punto de datos, con los intervalos de las x originales, no siempre se detectará una extrapolación oculta, sería útil contar con un procedimiento formal para detectarlas. Se definirá el conjunto convexo mínimo que contiene todos los n datos originales $(x_{i1}, x_{i2}, \ldots, x_{ik})$, $i = 1, 2, \ldots, n$ como la envolvente, o *cáscara* de variables regresoras (RVH, de *regressor variable hull*). Si un punto $x_{01}, x_{02}, \ldots, x_{0k}$ está dentro o en la frontera de la RVH, una predicción o una estimación implica interpolaciones, mientras que si este punto está fuera de la RVH, se requiere extrapolar.

Los elementos diagonales h_{ii} de la matriz sombrero $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ son útiles para detectar la **extrapolación oculta**. Los valores de h_{ii} dependen tanto de la distancia euclidiana del punto $\mathbf{x}_i$ al centroide, como de la densidad de los puntos en la RVH. En general, el punto que tiene mayor valor de h_{ii}, que sería $h_{\text{máx}}$, estará en la frontera de la región RVH del espacio x donde la densidad de las observaciones es relativamente baja. El conjunto de los puntos $\mathbf{x}$ (no necesariamente los puntos de datos con los que se ajustó el modelo) que satisfacen

$$\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} \leq h_{\text{máx}}$$

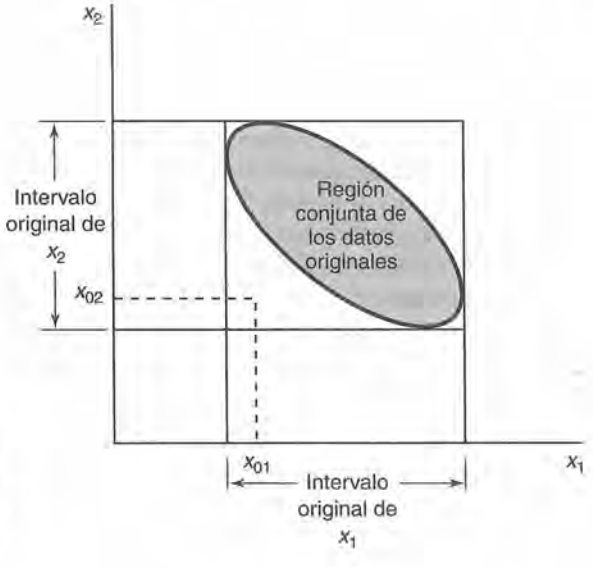


Figura 3.10 Ejemplo de extrapolación en regresión múltiple.

http://carlos2524.jimdo.com/

TABLA 3.6 Valores de h_{ii} para los datos de tiempo de entrega

Observación i	Cajas x_{i1}	Distancia x_{i2}	h_{ii}
1	7	560	0.10180
2	3	220	0.07070
3	3	340	0.09874
4	4	80	0.08538
5	6	150	0.07501
6	7	330	0.04287
7	2	110	0.08180
8	7	210	0.06373
9	30	1460	0.49829 = $h_{\text{máx}}$
10	5	605	0.19630
11	16	688	0.08613
12	10	215	0.11366
13	4	255	0.06113
14	6	462	0.07824
15	9	448	0.04111
16	10	776	0.16594
17	6	200	0.05943
18	7	132	0.09626
19	3	36	0.09645
20	17	770	0.10169
21	10	140	0.16528
22	26	810	0.39158
23	9	450	0.04126
24	8	635	0.12061
25	4	150	0.06664

es un elipsoide que encierra todos los puntos dentro de la RVH (véase Cook [1979] y Weisberg [1985]). Así, si lo que interesa es predecir o estimar en el punto $\mathbf{x}_0' = [1, x_{01}, x_{02}, \ldots, x_{0k}]$, el lugar de ese punto en relación con la RVH se refleja mediante

$$h_{00} = \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$$

Los puntos para los que $h_{00} > h_{\text{máx}}$ están fuera del elipsoide que encierra a la RVH son de extrapolación. Sin embargo, si $h_{00} < h_{\text{máx}}$, el punto está dentro del elipsoide, y posiblemente dentro de la RVH, y se consideraría punto de interpolación, porque está cerca de la nube de puntos que se usaron para ajustar el modelo. En general, mientras menor es el valor de h_{00}, el punto $\mathbf{x}_0$ está más cerca del centroide del espacio x.[1]

Weisberg [1985] hace notar que este procedimiento no produce el elipsoide de volumen mínimo que contiene la RVH. A esto se le llama el elipsoide de cobertura mínimo (MCE, de *minimum covering ellipsoid*). Él presenta un algoritmo alternativo para generar el elipsoide de cobertura mínimo. Sin embargo, la prueba de extrapolación basada en el

[1] Si $h_{\text{máx}}$ es mucho mayor que el siguiente valor mayor, el punto es un atípico severo en el espacio x. La presencia de ese valor atípico puede hacer que la elipse sea mucho mayor que lo deseable. En estos casos se podría usar el segundo valor máximo de h_{ii} como $h_{\text{máx}}$. Este método puede ser de utilidad cuando se ha subapreciado el punto más remoto, por ejemplo, mediante las técnicas robustas de ajuste que se describirán en el capítulo 11.

http://carlos2524.jimdo.com/

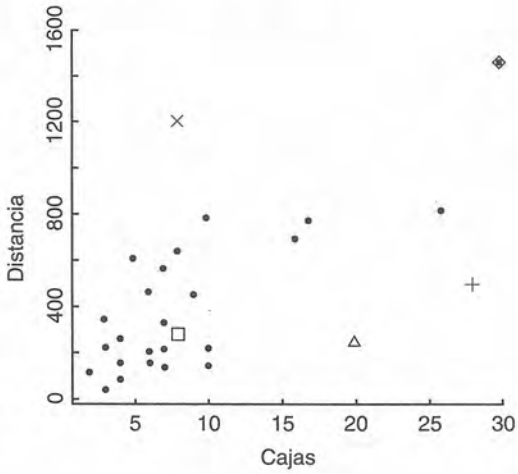


Figura 3.11 Diagrama de dispersión de cajas y distancia para los datos de tiempo de entrega.

MCE sigue siendo una aproximación, porque puede haber regiones dentro de ese elipsoide donde no hayan puntos de muestra.

Ejemplo 3.13 *Extrapolación oculta - los datos del tiempo de entrega*

Se ilustrará la detección de la extrapolación oculta con los datos de tiempo de entrega de bebidas gaseosas en el ejemplo 3.1. En la tabla 3.6 se muestran los valores de h_{ii} para los 25 datos. Nótese que la observación 9, representada por ♦ en la figura 3.11, tiene el valor máximo de h_{ii}. La figura 3.11 confirma que la observación 9 está en la frontera de la RVH. Ahora se supondrá que se desea hacer la predicción o estimación en los cuatro puntos siguientes:

Punto	Símbolo en la figura 3.11	x_{10}	x_{20}	h_{00}
a	□	8	275	0.05346
b	△	20	250	0.58917
c	+	28	500	0.89874
d	×	8	1200	0.86736

Todos estos puntos están dentro de los intervalos de los regresores x_1 y x_2. En la figura 3.11, el punto a (que se usó en los ejemplos 3.9 y 3.12 para estimar y predecir), para el que $h_{00} = 0.05346$, es un punto de interpolación, porque $h_{00} = 0.05346 < h_{\text{máx}} = 0.49829$. Los demás puntos, *b, c* y *d*, son todos de extrapolación, porque sus valores de h_{00} son mayores que $h_{\text{máx}}$. Esto se confirma con facilidad al inspeccionar la figura 3.11.

http://carlos2524.jimdo.com/

3.7 COEFICIENTES NORMALIZADOS DE REGRESIÓN

En general, es difícil comparar en forma directa coeficientes de regresión, porque la magnitud de $\hat{\beta}_j$ refleja las unidades de medida del regresor x_j. Por ejemplo, supongamos que el modelo de regresión es

$$\hat{y} = 5 + x_1 + 1\,000x_2$$

y que y se mide en litros, x_1 en mililitros y x_2 en litros. Nótese que aunque $\hat{\beta}_2$ es bastante mayor que $\hat{\beta}_1$, el efecto de ambos regresores sobre $\hat{y}$ es idéntico, porque un cambio de 1 litro en x_1 o en x_2, cuando la otra variable se mantiene constante, produce el mismo cambio de $\hat{y}$. Por lo general, las unidades del coeficiente de regresión β_j son y/unidades de x_j. Por esta razón a veces ayuda trabajar con regresores y variables de respuesta escalados, que produzcan **coeficientes de regresión adimensionales**. A esos coeficientes adimensionales se les suele llamar **coeficientes estandarizados de regresión**. A continuación se describirá cómo se calculan, aplicando dos técnicas frecuentes de escalamiento.

Escalamiento normal unitario

El primer método emplea el **escalamiento normal unitario** para los regresores y la variable de respuesta. Esto es,

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k, \tag{3.55}$$

y

$$y_i^* = \frac{y_i - \bar{y}}{s_y}, \quad i = 1, 2, \ldots, n, \tag{3.56}$$

en donde

$$s_j^2 = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n - 1}$$

es la varianza muestral del regresor x_j, y

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1}$$

es la varianza muestral de la respuesta. Nótese la semejanza con la estandarización de una variable normal aleatoria. Todos los regresores escalados y la respuesta escalada tienen su media muestral igual a cero, y varianza muestral igual a 1.

http://carlos2524.jimdo.com/

Con estas nuevas variables, el modelo de regresión se transforma en

$$y_i^* = b_1 z_{i1} + b_2 z_{i2} + \cdots + b_k z_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{3.57}$$

Al centrar las variables regresoras y de respuesta restando $\bar{x}_j$ y $\bar{y}$, se elimina la ordenada al origen del modelo (en realidad, el estimado de b_0 por mínimos cuadrados es $\hat{b} = \bar{y}^* = 0$). El estimador de **b** por mínimos cuadrados es

$$\hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}^* \tag{3.58}$$

Escalamiento de longitud unitaria

El segundo escalamiento que se usa con frecuencia es el **escalamiento de longitud unitaria**,

$$w_{ij} = \frac{x_{ij} - \bar{x}_j}{S_{jj}^{1/2}}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k, \tag{3.59}$$

y

$$y_i^0 = \frac{y_i - \bar{y}}{SS_T^{1/2}}, \quad i = 1, 2, \ldots, n, \tag{3.60}$$

en donde

$$S_{jj} = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

es la suma de cuadrados corregida, para el regresor x_j. En este escalamiento, cada nuevo regresor w_j tiene promedio $\bar{w}_j = 0$, y longitud $\sqrt{\sum_{i=1}^{n} (w_{ij} - \bar{w}_j)^2} = 1$. En función de esas variables, el modelo de regresión es

$$y_i^0 = b_q w_{i1} + b_2 w_{i2} + \cdots + b_k w_{ik} + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{3.61}$$

El vector de los coeficientes de regresión por mínimos cuadrados es

$$\hat{\mathbf{b}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}^0 \tag{3.62}$$

En el escalamiento de longitud unitaria, la matriz $\mathbf{W}'\mathbf{W}$ tiene la forma de una **matriz de correlación**, esto es,

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} 1 & r_{12} & r_{T3} & \cdots & r_{1k} \\ r_{12} & 1 & r_{23} & \cdots & r_{2k} \\ r_{13} & r_{23} & 1 & \cdots & r_{34} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{1k} & r_{2k} & r_{3k} & \cdots & 1 \end{bmatrix}$$

http://carlos2524.jimdo.com/

en donde

$$r_{ij} = \frac{\sum_{u=1}^{n} (x_{ui} - \bar{x}_i)(x_{uj} - \bar{x}_j)}{(S_{ii}S_{jj})^{1/2}} = \frac{S_{ij}}{(S_{ii}S_{jj})^{1/2}}$$

es la correlación simple entre el regresor x_i y el x_j. De igual modo,

$$\mathbf{W}'\mathbf{y}^0 = \begin{bmatrix} r_{1y} \\ r_{2y} \\ r_{3y} \\ \vdots \\ r_{ky} \end{bmatrix}$$

en la que

$$r_{jy} = \frac{\sum_{u=1}^{n} (x_{uj} - \bar{x}_j)(y_u - \bar{y})}{(S_{jj}SS_{\mathrm{T}})^{1/2}} = \frac{S_{jy}}{(S_{jj}SS_{\mathrm{T}})^{1/2}}$$

es la correlación simple[2] entre el regresor x_j y la respuesta y. Si se usa escalamiento normal unitario, la matriz $\mathbf{Z}'\mathbf{Z}$ se relaciona en forma estrecha con $\mathbf{W}'\mathbf{W}$; de hecho,

$$\mathbf{Z}'\mathbf{Z} = (n-1)\mathbf{W}'\mathbf{W} \tag{3.63}$$

En consecuencia, los estimados de los coeficientes de regresión en las ecuaciones (3.58) y (3.62) son idénticos, por lo que no importa qué escalamiento se use; ambos producen el mismo conjunto de coeficientes adimensionales de regresión $\hat{\mathbf{b}}$.

A los coeficientes de regresión $\hat{\mathbf{b}}$ se les suele llamar **coeficientes estandarizados de regresión.** La relación entre los coeficientes originales y los estandarizados de regresión es

$$\hat{\beta}_j = \hat{b}_j \left(\frac{SS_{\mathrm{T}}}{S_{jj}} \right)^{1/2}, \quad j = 1, 2, \ldots, k \tag{3.64}$$

y

$$\hat{\beta}_0 = \bar{y} - \sum_{j=1}^{k} \hat{\beta}_j \bar{x}_j \tag{3.65}$$

Muchos programas de cómputo para regresión múltiple usan este escalamiento para reducir los problemas que se originan en errores de redondeo, en la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Esos

[2] Se acostumbra llamar correlaciones a r_{iy} y r_{jj} aun cuando las de regresión no necesariamente sean variables aleatorias.

http://carlos2524.jimdo.com/

errores pueden ser graves, si las variables originales difieren mucho en su magnitud. La mayor parte de los programas de cómputo muestran también los coeficientes originales de regresión y los normalizados, a los que con frecuencia se les llama "coeficientes beta". Al interpretar los coeficientes estandarizados de regresión se debe recordar que siguen siendo **coeficientes de regresión parcial**, es decir, que b_j mide el efecto de x_j, dado que los demás regresores x_i, $i \neq j$, están en el modelo. Además, las b_j se afectan por el intervalo de valores de las variables regresoras. En consecuencia, puede ser peligroso usar la magnitud de las $\hat{b}_j$ como medida de la importancia relativa del regresor x_j.

Ejemplo 3.14 *Datos del tiempo de entrega*

Se calcularán los coeficientes estandarizados de regresión para los datos del tiempo de entrega en el ejemplo 3.1. Ya que

$$\begin{aligned} SS_T &= 5\,784.5426 & S_{11} &= 1\,136.5600 \\ S_{1y} &= 2\,473.3440 & S_{22} &= 2\,537\,935.0330 \\ S_{2y} &= 108\,038.6019 & S_{12} &= 44\,266.6800 \end{aligned}$$

al aplicar el escalamiento de longitud unitaria se determina que

$$r_{12} = \frac{S_{12}}{(S_{11}S_{22})^{1/2}} = \frac{44\,266.6800}{\sqrt{(1\,136.5600)(2\,537\,935.0303)}} = 0.824215$$

$$r_{1y} = \frac{S_{1y}}{(S_{11}SS_T)^{1/2}} = \frac{2\,473.3440}{\sqrt{(1\,136.5600)(5\,784.53426)}} = 0.964615$$

$$r_{2y} = \frac{S_{2y}}{(S_{22}SS_T)^{1/2}} = \frac{108{,}038.6019}{\sqrt{(2\,537\,935.0330)\,(5\,784.5426)}} = 0.891670$$

y la matriz de correlación para este problema es

$$\mathbf{W'W} = \begin{bmatrix} 1 & 0.824215 \\ 0.824215 & 1 \end{bmatrix}$$

Las ecuaciones normales, en función de los coeficientes estandarizados de regresión, son

$$\begin{bmatrix} 1 & 0.824215 \\ 0.824215 & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 0.964615 \\ 0.891670 \end{bmatrix}$$

En consecuencia, los coeficientes estandarizados de regresión son

http://carlos2524.jimdo.com/

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0.824215 \\ 0.824215 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.964615 \\ 0.891670 \end{bmatrix}$$

$$= \begin{bmatrix} 3.11841 & -2.57023 \\ -2.57023 & 3.11841 \end{bmatrix} \begin{bmatrix} 0.964615 \\ 0.891670 \end{bmatrix}$$

$$= \begin{bmatrix} 0.716267 \\ 0.301311 \end{bmatrix}$$

El modelo ajustado es

$$\hat{y}^0 = 0.716267w_1 + 0.301311w_2$$

Así, al aumentar una unidad el valor estandarizado de las cajas w_1 el valor del tiempo estandarizado $\hat{y}^0$ aumenta 0.716267. Además, al aumentar una unidad el valor de la distancia estandarizado w^2, $\hat{y}^0$ aumenta 0.301311 unidades. Por lo anterior, parece que el volumen del producto entregado es más importante que la distancia, porque tiene un efecto mayor sobre el tiempo de entrega en función de las variables estandarizadas. Sin embargo, se debe tener algo de precaución al llegar a esta conclusión, porque $\hat{b}_1$ y $\hat{b}_2$ siguen siendo coeficientes **parciales** de regresión, y $\hat{b}_1$ y $\hat{b}_2$ se afectan por la dispersión de los regresores. Así es que si se tomara otra muestra con distinto intervalo de valores de cajas y de distancia, se podría llegar a conclusiones distintas acerca de la importancia relativa de esos regresores.

3.8 MULTICOLINEALIDAD

Los modelos de regresión se usan en una gran diversidad de aplicaciones. Un problema serio que puede influir mucho sobre la utilidad de un modelo de regresión es la **multicolinealidad**, o **dependencia casi lineal** entre las variables de regresión. En esta sección se presentará en forma breve el problema, y se harán notar algunos de los efectos perjudiciales de la multicolinealidad. En el capítulo 10 se hará una presentación más amplia, que incluye más información sobre el diagnóstico y las medidas correctivas.

La multicolinealidad implica una dependencia casi lineal entre los regresores, los cuales son las columnas de la matriz **X**, por lo que es claro que una **dependencia lineal exacta** causaría una matriz **X′X singular**. La presencia de dependencias casi lineales puede influir en forma dramática sobre la capacidad de estimar coeficientes de regresión. Por ejemplo, véanse los datos de la figura 3.12

En la sección 3.7 se presentaron los coeficientes estandarizados de regresión. Supongamos que se usa escalamiento de **longitud unitaria** [Ecs. (3.59) y (3.60)] para los datos de la figura 3.12, por lo que la matriz **X′X** (llamada **W′W** en la Sec. 3.7) tendrá la forma de una matriz de correlación.

Esto da como resultado

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad (\mathbf{W}'\mathbf{W})^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

http://carlos2524.jimdo.com/

x_1	x_2
5	20
10	20
5	30
10	30
5	20
10	20
5	30
10	30

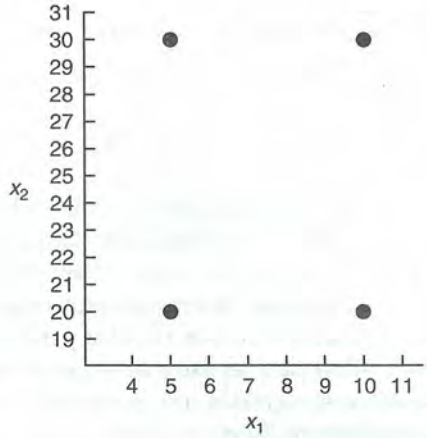


Figura 3.12 Datos de dos regresores.

Para los datos de tiempo de entrega de gaseosas, se demostró, en el ejemplo 3.14, que

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} 1.00000 & 0.824215 \\ 0.824215 & 1.00000 \end{bmatrix} \quad \text{y} \quad (\mathbf{W}'\mathbf{W})^{-1} = \begin{bmatrix} 3.11841 & -2.57023 \\ -2.57023 & 3.11841 \end{bmatrix}$$

Ahora se verán las varianzas de los coeficientes estandarizados de regresión $\hat{b}_1$ y $\hat{b}_2$ para los dos conjuntos de datos. Para el conjunto hipotético de la figura 3.12,

$$\frac{\text{Var}\left(\hat{b}_1\right)}{\sigma^2} = \frac{\text{Var}\left(\hat{b}_2\right)}{\sigma^2} = 1$$

mientras que para los datos de tiempo de entrega de gaseosas,

$$\frac{\text{Var}\left(\hat{b}_1\right)}{\sigma^2} = \frac{\text{Var}\left(\hat{b}_2\right)}{\sigma^2} = 3.11841$$

En los datos de tiempo de entrega de bebidas gaseosas, las varianzas de los coeficientes de regresión están **infladas** por la multicolinealidad. Esa multicolinealidad es evidente por los elementos fuera de la diagonal distintos de cero en $\mathbf{W}'\mathbf{W}$. Estos elementos fuera de la diagonal se llaman correlaciones simples entre los regresores, aunque puede ser que el término **correlación** no sea el adecuado, a menos de que las x sean variables aleatorias. Los elementos fuera de la diagonal sí proporcionan una medida de dependencia lineal entre los regresores. En esta forma, la multicolinealidad puede afectar en forma seria la precisión con la que se estiman los coeficientes de regresión.

Los elementos de la diagonal principal en la inversa de la matriz $\mathbf{X}'\mathbf{X}$ en forma de correlación [$(\mathbf{W}'\mathbf{W})^{-1}$ arriba] se llaman con frecuencia **factores de inflación de varianza** (VIF, de *variance inflation factors*), y son un diagnóstico importante de la multicolinealidad, para los datos de las bebidas gaseosas,

$$\text{VIF}_1 = \text{VIF}_2 = 3.11841$$

mientras que para los regresores hipotéticos anteriores,

$$\text{VIF}_1 = \text{VIF}_2 = 1$$

http://carlos2524.jimdo.com/

que implica que los dos regresores, x_1 y x_2, son **ortogonales**. Se puede demostrar que, en general, el factor de inflación de varianza para el j-ésimo coeficiente de regresión se puede escribir como sigue:

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2}$$

donde R_j^2 es el coeficiente de determinación múltiple obtenido haciendo la regresión de x_j sobre las demás variables regresoras. Es claro que si x_j depende casi linealmente de alguno de los demás regresores, entonces R_j^2 será casi la unidad, y que VIF_j será grande. Los factores VIF mayores que 10 implican problemas graves de multicolinealidad. La mayor parte de los paquetes de cómputo calculan y muestran el VIF_j. Por ejemplo, véase la salida de SAS PROC REG para los datos de tiempo de entrega en la figura 3.6.

Los modelos de regresión que se ajustan a los datos por el modelo de mínimos cuadrados, cuando hay una fuerte colinealidad son ecuaciones notoriamente malas de predicción, y los valores de los coeficientes de regresión suelen ser muy sensibles a los datos de la muestra tomada en particular. La formación de un modelo de regresión para los datos (x_1, x_2, y) de la figura 3.13a es análogo a hacer pasar un plano por los puntos. Es claro que ese plano será muy inestable, y sensible a cambios relativamente pequeños en los datos. Además, el modelo puede pronosticar en forma razonablemente buena valores de y en puntos parecidos a los observados en la muestra, pero es muy probable que cualquier extrapolación que se aleje de esa trayectoria produzca malas predicciones. En contraste, examínese el caso de los **regresores ortogonales** de la figura 3.13b. El plano ajustado a esos puntos será mucho más estable.

El diagnóstico y el tratamiento de la multicolinealidad es un aspecto importante del modelado de regresiones. Para conocer una descripción más profunda sobre este tema, véase el capítulo 10.

3.9 ¿POR QUÉ LOS COEFICIENTES DE REGRESIÓN TIENEN EL SIGNO INCORRECTO?

Al aplicar la regresión múltiple, a veces se llega a una contradicción respecto a la intuición o a la teoría, cuando uno o más de los coeficientes de regresión parece tener el **signo incorrecto**. Por ejemplo, el caso del problema puede implicar que un coeficiente determinado

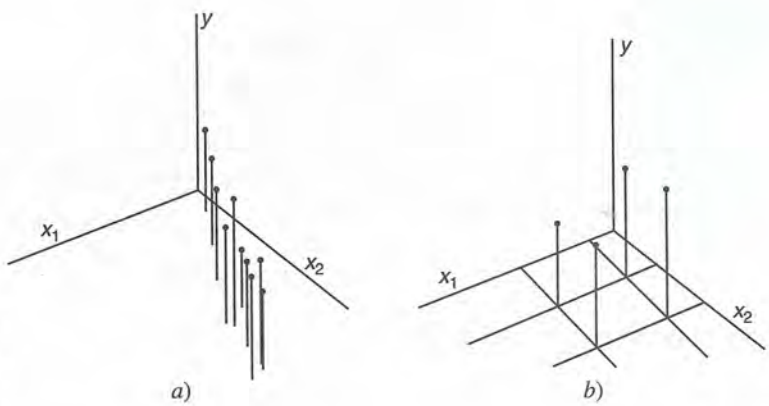


Figura 3.13 a) Un conjunto de datos con multicolinealidad. b) Regresores ortogonales.

http://carlos2524.jimdo.com/

de regresión debe ser positivo, mientras que el estimado real de ese parámetro es negativo. Este problema de signo "incorrecto" puede ser desconcertante, porque en general es difícil explicar un estimado negativo (por ejemplo) de un parámetro a quien usa el modelo, cuando ese usuario cree que el coeficiente debe ser positivo. Mullet [1976] hace notar que los coeficientes de regresión pueden tener el signo incorrecto por las siguientes razones:

1. El intervalo de alguno de los regresores es demasiado pequeño.
2. No se han incluido los regresores importantes en el modelo.
3. Hay multicolinealidad.
4. Se han cometido errores de cómputo.

Es fácil ver cómo el intervalo de valores de las x puede afectar el signo de los coeficientes de regresión. Examinemos el modelo de regresión lineal simple. La varianza del coeficiente de regresión $\hat{\beta}_1$ es $\text{Var}(\beta_1) = \sigma^2/S_{xx} = \sigma^2/\Sigma_{i=1}^{n}(x_i - \bar{x})^2$. Nótese que la varianza de $\hat{\beta}_1$ es inversamente proporcional a la "dispersión" del regresor. Así, si todos los niveles de x son cercanos entre sí, la varianza de $\hat{\beta}_1$ será grande. En algunos casos, la varianza de $\hat{\beta}_1$ podría ser tan grande que se obtenga un estimado negativo (por ejemplo) de un coeficiente de regresión que realmente es positivo. El caso se ilustra en la figura 3.14, donde se grafica la distribución de muestreo de $\hat{\beta}_1$. Al examinar esta figura se ve que la probabilidad de obtener un estimado negativo de $\hat{\beta}_1$ depende de lo cercano que el coeficiente real de regresión esté de cero, y de la varianza de $\hat{\beta}_1$, que está muy influida por la dispersión de las x.

En algunos casos, el analista puede controlar los niveles de los regresores. Aunque en estos casos es posible disminuir la varianza de los coeficientes de regresión aumentando el intervalo de las x, podría no ser lo deseable el dispersar demasiado los niveles de los regresores. Si las x abarcan un intervalo demasiado grande, y la función real de respuesta es no lineal, el analista debe desarrollar una ecuación mucho más complicada, para modelar en forma adecuada la curvatura en el sistema, además, en muchos problemas interviene una región del espacio x de interés específico para el experimentador, y podría ser impráctico o imposible repartir los regresores fuera de esa región de interés. En general, se debe balancear la precisión de la estimación, la complejidad probable del modelo y los valores de los regresores, con interés práctico, cuando se decida qué tanto dispersar las x.

También se pueden producir signos incorrectos cuando los regresores importantes no intervengan en el modelo; en esos casos, en realidad el signo no es incorrecto. La naturaleza parcial de los coeficientes de regresión causan la inversión del signo. Como ejemplo, se examinarán los siguientes datos:

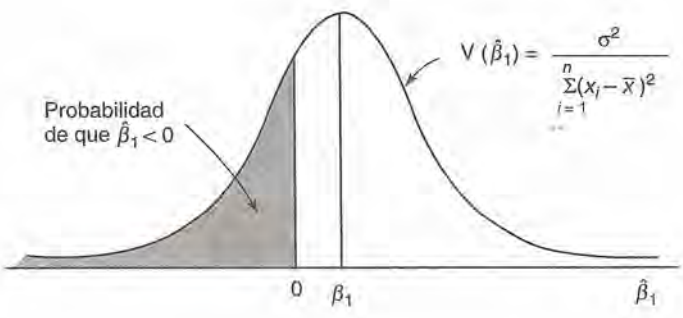


Figura 3.14 Distribución de muestreo de $\hat{\beta}_1$.

http://carlos2524.jimdo.com/

x_1	x_2	y	x_1	x_2	y
2	1	1	8	4	5
4	2	5	10	4	3
5	2	3	11	6	10
6	4	8	13	6	7

Si se está ajustando un modelo que sólo comprende a y y a x_1, la ecuación es

$$\hat{y} = 1.835 + 0.463x_1$$

donde $\hat{\beta} = 0.463$ es un coeficiente de regresión "total". Esto es, mide el efecto total de x_1, sin tener en cuenta la información contenida en x_2. El modelo donde intervienen x_1 y x_2 es

$$\bar{y} = 1.036 - 1.222x_1 + 3.649x_2$$

Nótese que ahora $\hat{\beta}_1 = -1.222$, se ha tenido una inversión de signo. La razón es que $\hat{\beta}_1 = -1.222$ en el modelo de regresión múltiple es un coeficiente de regresión parcial porque mide el efecto de x_1 **dado** que x_2 también está en el modelo.

Los datos de este ejemplo se ven en la figura 3.15. La razón de la diferencia de signos, entre los coeficientes de regresión parciales y totales, es obvia al revisar esta figura. Si no se tienen en cuenta los valores de x_2, la relación aparente entre y y x_1 tiene una pendiente positiva, sin embargo, si se tiene en cuenta la relación entre y y x_1 para valores constantes de x_2, se nota que en realidad esa relación tiene pendiente negativa. Así, un signo incorrecto en un modelo de regresión puede indicar que faltan regresores importantes, si el analista los puede identificar e incluirlos en el modelo, desaparecerán esos signos incorrectos.

La multicolinealidad puede ser causa de signos incorrectos de coeficientes de regresión. En efecto, una multicolinealidad severa infla las varianzas de los coeficientes de regresión, aumentando la probabilidad de que uno o más de los coeficientes de regresión tengan el signo incorrecto. Los métodos para diagnosticar y manejar la multicolinealidad se condensan en el capítulo 10.

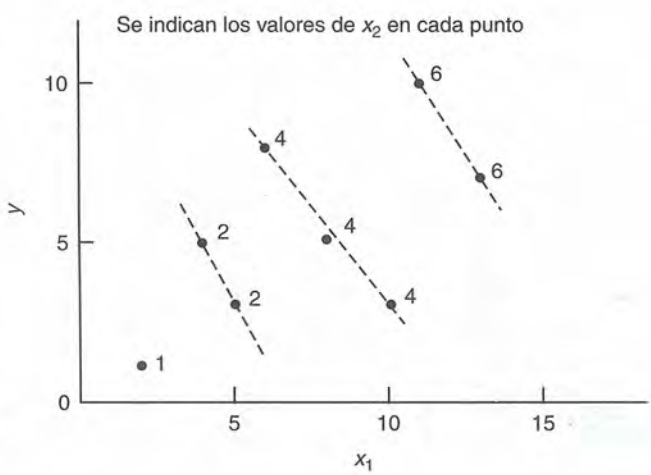


Figura 3.15 Gráfica de y en función de x_1.

http://carlos2524.jimdo.com/

Los errores de cómputo también son una fuente de signos incorrectos en los modelos de regresión. Los distintos programas de cómputo manejan los problemas de redondeo o de truncamiento en formas distintas, y unos son más efectivos que otros en este aspecto. Una gran multicolinealidad hace que la matriz $\mathbf{X}'\mathbf{X}$ esté mal acondicionada, lo cual también es una fuente de errores computacionales. Los errores de cómputo no sólo pueden causar inversiones de signo, sino hacer que los coeficientes de regresión difieran en varios órdenes de magnitud. Cuando se sospeche que haya problemas de signos incorrectos, se debe investigar la exactitud del programa de cómputo.

PROBLEMAS

3.1 Para los datos de la Liga Nacional de Futbol, en la tabla B.1 del apéndice:

a. Ajustar un modelo de regresión lineal múltiple que relacione la cantidad de juegos ganados con las yardas por aire del equipo (x_2), el porcentaje de jugadas por tierra (x_7) y las yardas por tierra del contrario (x_8).

b. Formar la tabla de análisis de varianza y probar el significado de la regresión.

c. Calcular el estadístico t para probar las hipótesis H_0: $\beta_2 = 0$, H_0: $\beta_7 = 0$ y H_0: $\beta_8 = 0$. ¿Qué conclusiones se pueden sacar acerca del papel de las variables x_2, x_7 y x_8 en el modelo?

d. Calcular R^2 y R^2_{Adj} para este modelo.

e. Con la prueba F parcial, determinar la contribución de x_7 al modelo. ¿Cómo se relaciona el estadístico F parcial con la prueba t para β_7 calculada en la parte c anterior?

3.2 Con los resultados del problema 3.1, demostrar en forma numérica que el cuadrado del coeficiente de correlación simple entre los valores observados y_i y los valores ajustados $\hat{y}_i$ es igual a R^2.

3.3 Véase el problema 3.1. Calcular:

a. Un intervalo de confianza de 95% para β_7.

b. Un intervalo de confianza de 95% para la cantidad media de juegos ganados por un equipo cuando $x_2 = 2\,300$, $x_7 = 56$ y $x_8 = 2\,100$.

3.4 Para los datos de la Liga Nacional de Futbol del problema 3.1, ajustar un modelo a esos datos, usando sólo x_7 y x_8 como regresores.

a. Probar la significacia de la regresión.

b. Calcular R^2 y R^2_{Adj}. ¿Cómo se comparan esas cantidades con las calculadas para el modelo del problema 3.1, que tenía un regresor más (x_2)?

c. Calcular un intervalo de confianza de 95% para β_7. También, un intervalo de confianza de 95% para la cantidad media de juegos ganados por un equipo cuando $x_7 = 56$ y $x_8 = 2\,100$. Comparar la longitud de esos intervalos de confianza con las longitudes de los correspondientes en el problema 3.3.

d. ¿Qué conclusiones se pueden sacar de este problema, acerca de las consecuencias de omitir un regresor importante de un modelo?

3.5 Véanse los datos de rendimiento de gasolina en la tabla B.3 del apéndice.

a. Ajustar un modelo de regresión lineal múltiple que relacione el rendimiento de la gasolina y, en millas por galón, la cilindrada del motor x_1 y la cantidad de gargantas del carburador, x_6.

b. Formar la tabla de análisis de varianza, y probar la significancia de la regresión.

c. Calcular R^2 y R^2_{Adj} para este modelo. Compararlas con la R^2 y R^2_{Adj} para el modelo de regresión lineal simple, que relacionaba las millas con la cilindrada en el problema 2.4.

http://carlos2524.jimdo.com/

d. Determinar un intervalo de confianza de 95% para β_1.
e. Calcular el estadístico t para probar H_0: $\beta_1 = 0$ y H_0: $\beta_6 = 0$. ¿Qué conclusiones se pueden sacar?
f. Determinar un intervalo de confianza de 95% para el rendimiento promedio de la gasolina, cuando $x_1 = 275$ pulg^3 y $x_6 = 2$ gargantas.
g. Determinar un intervalo de predicción de 95% para una nueva observación de rendimiento de gasolina cuando $x_1 = 275$ pulg^3 y $x_6 = 2$ gargantas.

3.6 En el problema 2.4 se pidió establecer un intervalo de confianza de 95% para el rendimiento promedio de gasolina, y un intervalo de predicción de 95% para el rendimiento cuando la cilindrada del motor es $x_1 = 275$ pulg^3. Comparar las longitudes de esos intervalos con las de los intervalos de confianza y de predicción, para el problema anterior, 3.5. ¿Dice algo acerca de las ventajas de agregar x_6 al modelo?

3.7 Véanse los datos sobre precios de viviendas en la tabla B.4 del apéndice.
a. Ajustar un modelo de regresión múltiple que relacione el precio de venta con los nueve regresores.
b. Probar la significancia de la regresión. ¿Qué conclusiones se pueden sacar?
c. Usar pruebas t para evaluar la contribución de cada regresor al modelo.
d. ¿Cuál es la contribución del tamaño de lote y el espacio vital para el modelo, dado que se incluyeron todos los demás regresores?
e. En este modelo, ¿la colinealidad es un problema potencial?

3.8 Los datos de la tabla B.5 del apéndice presentan la eficiencia de un proceso químico, en función de varias variables controlables del proceso.
a. Ajustar un modelo de regresión múltiple que relaciona el CO_2 del producto (y) con el solvente total (x_6) y el consumo de hidrógeno (x_7).
b. Probar la significancia de la regresión. Calcular R^2 y R^2_{Adj}.
c. Usar pruebas t para determinar la contribución de x_6 y x_7 al modelo.
d. Establecer intervalos de confianza de 95% para β_6 y β_7.
e. Volver a ajustar el modelo sólo con x_6 como regresor. Probar la significancia de la regresión y calcular R^2 y R^2_{Adj}. Comentar los resultados. Con base en estos estadísticos, ¿es satisfactorio el modelo?
f. Establecer un intervalo de confianza de 95% para β_6, con el modelo que se ajustó en el punto e. Comparar la longitud de este intervalo de confianza con la del determinado en la parte d. ¿Se deduce algo importante acerca de la contribución de x_7 al modelo?
g. Comparar los valores de MS_{Res} obtenidos con los dos modelos que se ajustaron (partes a y e). ¿Cómo cambió la MS_{Res} al quitar x_7 del modelo? ¿Indica lo anterior algo importante acerca de la contribución de x_7 al modelo?

3.9 En la tabla B.6 del apéndice muestra la concentración de $NbOCl_3$ en un reactor de tubo de flujo, en función de varias variables controlables.
a. Ajustar un modelo de regresión múltiple que relacione la concentración de $NbOCl_3(y)$ con la de $COCl_2(x_1)$ y la fracción mol (x_4).
b. Pruebe la significancia de la regresión.
c. Calcule R^2 y R^2_{Adj} para este modelo.
d. Con pruebas t, determinar la contribución de x_1 y x_4 al modelo. ¿Son necesarios los dos regresores, x_1 y x_4?
e. En este modelo, ¿es la colinealidad un problema potencial?

3.10 Se cree que la calidad del vino Pinot Noir se relaciona con sus propiedades de claridad, aroma cuerpo, sabor y fuerza. Los datos de 38 vinos se presentan en el apéndice B.11.

http://carlos2524.jimdo.com/

a. Ajustar un modelo de regresión lineal múltiple que relacione la calidad del vino con esos regresores.
b. Probar la significancia de la regresión. ¿A qué conclusiones se puede llegar?
c. Use pruebas t para evaluar la contribución de cada regresor al modelo. Comentar los resultados.
d. Calcular R^2 y R^2_{Adj} para este modelo. Comparar esos valores con R^2 y R^2_{Adj} para el modelo de regresión lineal que relacione la calidad del vino con su aroma y sabor. Comentar los resultados.
e. Determinar un intervalo de confianza de 95% para el coeficiente de regresión del sabor, para los dos modelos de la parte d. Comentar las diferencias encontradas.

3.11 Un ingeniero hizo un experimento para determinar la presión, temperatura y flujo de CO_2, la humedad y el tamaño de partícula de los cacahuates sobre el rendimiento total de aceite por lote de cacahuates. La tabla B.7 del apéndice resume sus resultados.
a. Ajustar un modelo de regresión lineal múltiple que relacione el rendimiento con esos regresores.
b. Probar la significancia de la regresión. ¿A qué conclusiones se puede llegar?
c. Hacer pruebas t para evaluar la contribución de cada regresor al modelo. Comentar los resultados.
d. Calcular R^2 y R^2_{Adj} para este modelo. Comparar esos valores con los de R^2 y R^2_{Adj} para el modelo de regresión lineal múltiple, que relacione el rendimiento con la temperatura y el tamaño de partícula. Comentar los resultados.
e. Establecer un intervalo de confianza de 95% para el coeficiente de regresión de la temperatura en ambos modelos de la parte d. Comentar las diferencias encontradas.

3.12. Un ingeniero químico estudió el efecto de la cantidad de surfactante y el tiempo sobre la formación de catrato. Los catratos se usan como medio de conservación en frío. La tabla B.8 del apéndice resume los resultados experimentales.
a. Ajustar un modelo de regresión lineal múltiple que relacione la formación de catrato con estos regresores.
b. Probar la significancia de la regresión. ¿A qué conclusiones se puede llegar?
c. Hacer pruebas t para evaluar la contribución de cada regresor al modelo. Comentar los resultados.
d. Calcular R^2 y R^2_{Adj} para este modelo. Comparar esos valores con los de R^2 y R^2_{Adj} para el modelo de regresión lineal simple que relaciona la formación de catrato con el tiempo. Comentar los resultados.
e. Determinar un intervalo de confianza de 95% para el coeficiente de regresión del tiempo para los dos modelos de la parte d. Comentar las diferencias encontradas.

3.13 Un ingeniero estudió el efecto de cuatro variables de un factor adimensional con el que se describen las caídas de presión en una columna de burbujeo de platos perforados. En la tabla B.9 del apéndice se resumen los resultados experimentales.
a. Ajustar un modelo de regresión lineal múltiple que relacione ese número adimensional con los cuatro regresores.
b. Probar la significancia de la regresión. ¿Qué conclusiones se pueden sacar?
c. Hacer pruebas t para evaluar la contribución de cada regresor al modelo. Comentar los resultados.
d. Calcular R^2 y R^2_{Adj} para este modelo. Comparar estos valores con los de R^2 y R^2_{Adj} para el modelo de regresión lineal múltiple que relacione el número adimensional con x_2 y x_3. Comentar los resultados.
e. Determinar un intervalo de confianza de 99% para el coeficiente de regresión de x_2 en ambos modelos de la parte d. Comentar las diferencias.

http://carlos2524.jimdo.com/

3.14 La viscosidad cinemática de cierto sistema de solventes depende de la relación entre los dos solventes y la temperatura. En la tabla B.10 del apéndice se resumen los resultados experimentales de una corrida.

a. Ajustar un modelo de regresión lineal múltiple que relacione la viscosidad con los dos regresores.

b. Probar la significancia de la regresión. ¿Qué conclusiones se pueden sacar?

c. Hacer pruebas t para evaluar la contribución de cada regresor al modelo. Comentar los resultados.

d. Calcular R^2 y R^2_{Adj} para este modelo. Comparar esos valores con los de R^2 y R^2_{Adj} para el modelo de regresión lineal simple que relacione la viscosidad sólo con la temperatura. Comentar los resultados obtenidos.

e. Establecer un intervalo de confianza de 99% para el coeficiente de regresión de la temperatura para los dos modelos de la parte d. Comentar las diferencias encontradas.

3.15 Un ingeniero químico estudia la forma en que la cantidad de conversión de un producto, a partir de una materia prima (y), depende de la temperatura de reacción (x_1) y del tiempo de reacción (x_2), y así llegó a los siguientes modelos de regresión:

1. $\hat{y} = 100 + 0.2x_1 + 4x_2$

2. $\hat{y} = 95 + 0.15x_1 + 3x_2 + 1x_1x_2$

Los dos modelos se han establecido para los intervalos de $20 \leq x_1 \leq 50$ (°C) y $0.5 \leq x_2 \leq 10$ (horas).

a. Al usar ambos modelos, ¿cuál es el valor calculado de conversión cuando $x_2 = 2$, en términos de x_1? Repetir este cálculo para $x_2 = 8$. Trazar una gráfica de los valores predichos en función de la temperatura para ambos modelos de conversión. Comentar sobre el efecto del término de interacción en el modelo 2.

b. Calcular el cambio esperado en la conversión promedio para un cambio unitario en la temperatura x_1 para el modelo 1, cuando $x_2 = 5$. ¿Depende esta cantidad del valor específico del tiempo de reacción seleccionado? ¿Por qué?

c. Determinar el cambio esperado en la conversión promedio para un cambio unitario de temperatura x_1 para el modelo 2, cuando $x_2 = 5$. Repetir este cálculo para $x_2 = 2$ y para $x_2 = 8$. ¿Depende el resultado del valor seleccionado para x_2? ¿Por qué?

3.16 Demostrar que una forma equivalente de hacer la prueba de la significancia de una regresión lineal múltiple es basar la prueba de R^2 como sigue: para probar H_0: $\beta_1 = \beta_2 = \cdots = \beta_k$ contra H_1: al menos una $\beta_j \neq 0$, calcular lo siguiente:

$$F_0 = \frac{R^2\,(n-p)}{k\,(1-R^2)}$$

y rechazar H_0 si el valor calculado de F_0 es mayor que $F_{\alpha,\,k,\,n-p}$, siendo $p = k + 1$.

3.17 Suponer que se ha ajustado un modelo de regresión lineal con $k = 2$ regresores, a $n = 25$ observaciones, y que $R^2 = 0.90$

a. Probar el significado de la regresión con $\alpha = 0.05$. Usar los resultados del problema anterior.

b. ¿Cuál es valor mínimo de R^2 que haría llegar a la conclusión de que la regresión es significativa si $\alpha = 0.05$? ¿Sorprende al lector la pequeñez de este valor de R^2?

3.18 Demostrar que una fórmula alternativa de cómputo de la suma de cuadrados de regresión, en un modelo de regresión lineal es

$$SS_{\text{R}} = \sum_{i=1}^{n} \hat{y}_i^2 - n\bar{y}^2$$

http://carlos2524.jimdo.com/

3.19 Se tiene el siguiente modelo de regresión lineal múltiple:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

De acuerdo con el procedimiento de prueba de una hipótesis lineal general, indicar cómo probar

a. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$

b. H_0: $\beta_1 = \beta_2, \beta_3 = \beta_4$

c. H_0: $\beta_1 - 2\,\beta_2 = 4\,\beta_3$

$\beta_1 + 2\,\beta_2 = 0$

3.20 Suponer que se cuenta con las dos muestras independientes siguientes:

$$\begin{array}{cc} y & x \\ \hline y_1 & x_1 \\ y_2 & x_2 \\ \vdots & \vdots \\ y_{n_1} & x_{n_1} \end{array} \Bigg\} \text{ Ejemplo 1} \qquad \begin{array}{cc} y & x \\ \hline y_{n_1+1} & x_{n_1+1} \\ \vdots & \vdots \\ y_{n_1+n_2} & x_{n_1+n_2} \end{array} \Bigg\} \text{ Ejemplo 2}$$

y que a esas muestras se pueden ajustar dos modelos:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n_2$$
$$y_i = \gamma_0 + \gamma_1 x_i + \varepsilon_i, \quad i = n_1 + 1,\ n_1 + 2, \ldots, n_1 + n_2$$

a. Indicar cómo se pueden escribir estos dos modelos separados en forma de un solo modelo.

b. Con el resultado de la parte a, indicar cómo se puede usar la hipótesis lineal general para probar la igualdad de las pendientes β_1 y γ_1.

c. Con el resultado de la parte a, indicar cómo se puede usar la hipótesis lineal general para probar la igualdad de las dos líneas de regresión.

d. Con el resultado de la parte a, indicar cómo se puede usar la hipótesis lineal general para probar que las dos pendientes son iguales a una constante c.

3.21 Demostrar que $\text{Var}(\hat{\mathbf{y}}) = \sigma^2\mathbf{H}$.

3.22 Demostrar que las matrices $\mathbf{H}$ e $\mathbf{I} - \mathbf{H}$ son idempotentes, es decir, que $\mathbf{HH} = \mathbf{H}$ e $(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) = \mathbf{I} - \mathbf{H}$.

3.23 Para el modelo de regresión lineal simple, demostrar que los elementos de la matriz de sombrero son

$$h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}} \quad \text{y} \quad h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}$$

Describir el comportamiento de estas cantidades cuando x_i se aleja de $\bar{x}$.

3.24 Se tiene el modelo de regresión lineal múltiple $\mathbf{y} = \mathbf{X}\beta + \varepsilon$. Demostrar que el estimador de mínimos cuadrados se puede escribir en la forma

$$\hat{\beta} = \beta + \mathbf{R}\varepsilon \quad \text{donde} \quad \mathbf{R} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

http://carlos2524.jimdo.com/

3.25 Demostrar que los residuales de un modelo de regresión lineal se pueden expresar en la forma $\mathbf{e} = (\mathbf{I} - \mathbf{H})\varepsilon$. [Sugerencia: véase la Ec. (3.15b).]

3.26 Para el modelo de regresión lineal múltiple, demostrar que $SS_R(\beta) = \mathbf{y}'\mathbf{H}\mathbf{y}$.

3.27 Demostrar que R^2 es el cuadrado de la correlación entre $\mathbf{y}$ y $\hat{\mathbf{y}}$.

3.28 **Mínimos cuadrados restringidos**. Se trata de determinar el estimador de β por mínimos cuadrados, en el modelo $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, sujeto a un conjunto de restricciones de igualdad de β, por ejemplo $\mathbf{T}\beta = \mathbf{c}$. Demostrar que el estimador es

$$\tilde{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}' \left[\mathbf{T}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}\right]^{-1} \left(\mathbf{c} - \mathbf{T}\hat{\beta}\right)$$

en donde $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Describir casos en los que podría ser adecuado este estimador restringido. Determinar la suma de cuadrados de residuales para el estimador restringido. ¿Es mayor o menor que la suma de cuadrados de residuales en el caso no restringido?

3.29 Sea $\mathbf{x}_j$ el j-ésimo renglón de $\mathbf{X}$, y $\mathbf{X}_{-j}$ la matriz $\mathbf{X}$ con el j-ésimo renglón eliminado. Demostrar que

$$\mathrm{Var}[\,\hat{\beta}_j\,] = \sigma^2\left[\,\mathbf{x}_j'\mathbf{x}_j - \mathbf{x}_j'\mathbf{X}_{-j}(\mathbf{X}_{-j}'\mathbf{X}_{-j})^{-1}\mathbf{X}_{-j}'\mathbf{x}_j\,\right]$$

3.30 Se tienen los siguientes dos modelos, donde $E(\varepsilon) = \mathbf{0}$ y $\mathrm{Var}(\varepsilon) = \sigma^2\mathbf{I}$:

Modelo A: $\mathbf{y} = \mathbf{X}_1\beta_1 + \varepsilon$

Modelo B: $\mathbf{y} = \mathbf{X}_1'\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$

Demostrar que $R_A^2 \leq \mathbf{R}_B^2$.

3.31 Suponer que se ajusta el modelo $\mathbf{y} = \mathbf{X}_1\beta_1 + \varepsilon$, pero que en realidad el modelo verdadero es $\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$. Para ambos modelos, suponer que $E(\varepsilon) = \mathbf{0}$ y que $\mathrm{Var}(\varepsilon) = \sigma^2\mathbf{I}$. Determinar el valor esperado y la varianza del estimado por mínimos cuadrados ordinarios, $\hat{\beta}_1$. ¿Bajo qué condiciones este estimador es insesgado?

3.32 Se tiene un modelo de regresión especificado correctamente, con p términos incluida la ordenada al origen. Hacer las hipótesis acostumbradas acerca de ε. Demostrar que

$$\sum_{i=1}^{n} \mathrm{Var}(\,\hat{y}_i\,) = p\sigma^2$$

3.33 Sea R_j^2 el coeficiente de determinación de la regresión del j-ésimo regresor respecto a los otros $k - 1$ regresores. Demostrar que el j-ésimo factor de inflación de la varianza se puede expresar como sigue:

$$\frac{1}{1 - R_j^2}$$

3.34 Considerar las hipótesis para el modelo lineal general, que son de la forma

$$H_0: \mathbf{T}\beta = \mathbf{c}$$

$$H_1: \mathbf{T}\beta \neq \mathbf{c}$$

siendo $\mathbf{T}$ una matriz de $q \times p$, de rango q. Deducir el estadístico F adecuado bajo las hipótesis nula y alternativa.

http://carlos2524.jimdo.com/

CAPÍTULO 4

COMPROBACIÓN DE LA ADECUACIÓN DEL MODELO

4.1 INTRODUCCIÓN

Las principales **premisas** que se han hecho hasta ahora al estudiar el análisis de regresión son las siguientes:

1. La relación entre la respuesta y y los regresores es lineal, al menos en forma aproximada.
2. El término de error ε tiene media cero.
3. El término de error ε tiene varianza σ^2 constante.
4. Los errores no están correlacionados.
5. Los errores tienen distribución normal.

En conjunto, las hipótesis 4 y 5 implican que los errores son variables aleatorias independientes. Se requiere la premisa 5 para probar hipótesis y para estimar intervalos.

Siempre se debe tener en cuenta que la validez de estas premisas es dudosa, y se deben hacer análisis para examinar la adecuación del modelo que se haya desarrollado en forma tentativa. Las clases de inadecuación de modelo que se describen aquí tienen consecuencias potenciales graves. Las grandes violaciones a las premisas pueden producir un modelo inestable, en el sentido que una muestra distinta podría conducir a un modelo totalmente diferente, y obtener conclusiones opuestas. En general, no se pueden detectar desviaciones respecto a las premisas básicas examinando los estadísticos estándar de resumen, como por ejemplo los estadísticos t, F o R^2. Éstas son propiedades "globales" del modelo, y como tal no aseguran la adecuación del mismo.

En este capítulo se presentarán varios métodos de utilidad para diagnosticar violaciones de las premisas básicas de regresión. Esos métodos de diagnóstico se basan principalmente en el estudio de los **residuales** del modelo. En los capítulos 5 y 6 se describirán métodos para manejar inadecuaciones de modelo, así como otros diagnósticos más complejos.

4.2 ANÁLISIS DE RESIDUALES

4.2.1 Definición de residuales

Ya se habían definido antes los residuales, en la siguiente forma:

$$e_i = y_i = \hat{y}_i, \qquad i = 1, 2, \ldots, n \tag{4.1}$$

http://carlos2524.jimdo.com/

siendo y_i una observación, y $\hat{y}_i$ su valor ajustado correspondiente. Como se puede considerar que un residual es la **desviación** entre los **datos** y el **ajuste**, también es una medida de la variabilidad de la variable de respuesta que no explica el modelo de regresión. También conviene imaginar que los residuales son los valores realizados, u observados, de los errores del modelo, por lo que toda desviación de las premisas de los errores se debe reflejar en los residuales. El análisis de los residuales es una forma eficaz de descubrir diversos tipos de inadecuación del modelo. Como se verá, una forma muy efectiva de investigar lo bien que se ajusta el modelo a los datos, y comprobar las premisas de la sección 4.1 es **graficar los residuales**.

Los residuales tienen varias propiedades importantes. Tienen media cero, y su varianza promedio aproximada se estima con:

$$\frac{\sum_{i=1}^{n}(e_i - \bar{e})^2}{n-p} = \frac{\sum_{i=1}^{n} e_i^2}{n-p} = \frac{SS_{\text{Res}}}{n-p} = MS_{\text{Res}}$$

Sin embargo, los residuales no son independientes, ya que los n residuales sólo tienen $n-p$ grados de libertad asociados a ellos. Esta no independencia de los residuales tiene poco efecto en su aplicación para comprobar la adecuación del modelo, siempre y cuando n no sea pequeña en relación con la cantidad de parámetros p.

4.2.2 Métodos para escalar residuales

A veces es mejor trabajar con **residuales escalados**. En esta sección se presentarán cuatro métodos frecuentes para escalar residuales. Los residuales escalados son útiles para determinar observaciones que sean **atípicas** o **valores extremos**, esto es, observaciones que en algún aspecto estén separadas del resto de los datos. Véanse ejemplos de valores atípicos y valores extremos en las figuras 2.6 a 2.8.

Residuales estandarizados

Ya que la varianza aproximada de un residual se estima con MS_{Res}, el cuadrado medio de los residuales, un escalamiento lógico de los residuales sería el de los **residuales estandarizados**

$$d_i = \frac{e_i}{\sqrt{MS_{\text{Res}}}}, \quad i = 1, 2, \ldots, n \tag{4.2}$$

Los residuales estandarizados tienen media cero y varianza aproximadamente unitaria, en consecuencia, un residual estandarizado grande (por ejemplo $d_i > 3$) indica que se trata de un valor atípico potencial.

Residuales estudentizados

Si se usa el MS_{Res} como la varianza del i-ésimo residual e_i sólo se tendrá una aproximación. Se puede mejorar el escalamiento de residuales dividiendo e_i entre la desviación estándar

http://carlos2524.jimdo.com/

exacta del i-ésimo residual. Recuérdese que, de acuerdo con la ecuación (3.15b), el vector de los residuales se puede escribir como sigue:

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{y} \tag{4.3}$$

en donde $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ es la **matriz de sombrero**. Esta matriz tiene varias propiedades útiles. Es **simétrica** ($\mathbf{H}' = \mathbf{H}$) y es **idempotente** ($\mathbf{HH} = \mathbf{H}$). De forma parecida, la matriz $\mathbf{I} - \mathbf{H}$ es simétrica e idempotente. Si se sustituye $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ en la ecuación (4.3) se obtiene

$$\begin{aligned} \mathbf{e} &= (\mathbf{I} - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta} - \mathbf{HX}\boldsymbol{\beta} + (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} \end{aligned} \tag{4.4}$$

Por lo anterior, los residuales son la misma transformación lineal de las observaciones $\mathbf{y}$ y los errores ε.

La matriz de covarianza de los residuales es

$$\mathrm{Var}(\mathbf{e}) = \mathrm{Var}[(\mathbf{I} - \mathbf{H})\varepsilon] = (\mathbf{I} - \mathbf{H})\mathrm{Var}(\varepsilon)(\mathbf{I} - \mathbf{H})' = \sigma^2(\mathbf{I} - \mathbf{H}) \tag{4.5}$$

porque $\mathrm{Var}(\varepsilon) = \sigma^2\mathbf{I}$, e $\mathbf{I} - \mathbf{H}$ es simétrica e idempotente. En general, la matriz $\mathbf{I} - \mathbf{H}$ no es diagonal, por lo que los residuales tienen distintas varianzas, y están correlacionados.

La varianza del i-ésimo residual es

$$\mathrm{Var}(e_i) = \sigma^2(1 - h_{ii}) \tag{4.6}$$

en donde h_{ii} es el i-ésimo elemento de la diagonal de la matriz de sombrero $\mathbf{H}$. La covarianza entre los residuales e_i y e_j es

$$\mathrm{Cov}(e_i, e_j) = -\sigma^2 h_{ij} \tag{4.7}$$

siendo h_{ij} el j-ésimo elemento de la matriz de sombrero. Ahora bien, ya que $0 \le h_{ii} \le 1$, si se usa el cuadrado medio de los residuales, MS_{Res}, para estimar la varianza de los residuales, en realidad se sobreestima la $\mathrm{Var}(e_i)$. Además, como h_{ii} es una medida del **lugar** o **ubicación** del i-ésimo punto en el espacio de x (recuérdese la descripción de la extrapolación oculta, en la Sec. 3.6), la varianza de e_i depende de dónde está el punto $\mathbf{x}_i$. En general, los puntos cercanos al centro del espacio de x tienen mayor varianza (un pobre ajuste por mínimos cuadrados) que los residuales en lugares más remotos. Las violaciones de las premisas del modelo están, con más probabilidad, en los puntos remotos, y pueden ser difíciles de detectar por inspección de los residuales ordinarios e_i (o de los residuales estandarizados d_i), porque en general, sus residuales serán menores.

Entonces, un procedimiento lógico es examinar los **residuales estudentizados**

$$r_i = \frac{e_i}{\sqrt{MS_{\mathrm{Res}}(1 - h_{ii})}}, \quad i = 1, 2, \ldots, n \tag{4.8}$$

en lugar de e_i (o d_i). Cuando la forma del modelo es correcta estos residuales estudentizados tiene varianza constante $\mathrm{Var}(r_i) = 1$, independiente del lugar de $\mathbf{x}_i$. En muchos casos, la varianza de los residuales se estabiliza, en especial para conjuntos grandes de datos. En esos casos podrá haber poca diferencia entre los residuales estandarizados y los estudentizados.

http://carlos2524.jimdo.com/

Así, los residuales estandarizados y los estudentizados aportan con frecuencia información equivalente. Sin embargo, ya que cualquier punto con un residual grande **y también** una h_{ii} grande tiene una influencia potencial muy grande sobre el ajuste por mínimos cuadrados, se recomienda por lo general examinar los residuales estudentizados.

Algunos de esos puntos son muy fáciles de apreciar, examinando los residuales estudentizados para un modelo de regresión lineal simple. Si sólo hay un regresor, es fácil demostrar que los residuales estudentizados son

$$r_i = \frac{e_i}{\sqrt{MS_{\text{Res}}\left[1 - \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)\right]}}, \quad i = 1, 2, \ldots, n \tag{4.9}$$

Nótese que cuando la observación x_i es cercana al punto medio de los datos de x, la diferencia $x_i - \bar{x}$ será pequeña, y la desviación estándar estimada de e_i [el denominador de la Ec. (4.9)] será grande. Al revés, cuando x_i está cerca de los extremos del intervalo de los datos de x, la diferencia $x_i - \bar{x}$ será grande, y la desviación estándar estimada de e_i será pequeña. También, cuando el tamaño n de la muestra es relativamente grande, el efecto de $(x_i - \bar{x})^2$ será relativamente pequeño, por lo que en conjuntos grandes de datos los residuales estudentizados no serán muy diferentes de los residuales estandarizados.

Residuales PRESS

Los residuales estandarizados y los estudentizados son efectivos para detectar valores atípicos. Otro método para hacer que los residuales sean útiles en la determinación de valores atípicos consiste en examinar la cantidad que se calcula partiendo de $y_i - \hat{y}_{(i)}$, siendo $\hat{y}_{(i)}$ el valor ajustado de la i-ésima respuesta, basado en todas las observaciones excepto esa i-ésima. La lógica de este método es que si la i-ésima observación y_i realmente es atípica, el modelo de regresión basado en todas las observaciones estará demasiado influido por esta observación. Esto puede conducir a un valor ajustado $\hat{y}_i$ muy parecido al valor observado y_i, y en consecuencia el residual ordinario e_i será pequeño. Por lo anterior, será difícil detectar el valor atípico. Sin embargo, si se elimina la i-ésima observación, entonces $\hat{y}_{(i)}$ no puede estar influido por esa observación, así que el residual que resulte probablemente indique la presencia del valor atípico.

Si se elimina la i-ésima observación, se ajusta el modelo de regresión a las $n - 1$ observaciones restantes, y se calcula el valor predicho de y_i, correspondiente a la observación omitida, el **error de predicción** correspondiente es

$$e_{(i)} = y_i - \hat{y}_{(i)} \tag{4.10}$$

Este cálculo de error de predicción se repite para cada observación $i = 1, 2, \ldots, n$. A esos errores de predicción se les suele llamar **residuales PRESS** (por su empleo para calcular la suma de cuadrados de error de predicción, en inglés *prediction error sum of squares*, que se describirá en la Sec. 4.3). Algunos autores llaman **residuales eliminados** a los $e_{(i)}$.

Al principio, parecería que el cálculo de PRESS residuales requiere el ajuste de n regresiones diferentes. Sin embargo, es posible calcularlos a partir de los resultados de un solo ajuste por mínimos cuadrados a todas las n observaciones. En el apéndice C.7 se indica cómo se procede. Sucede que el i-ésimo residual PRESS es

http://carlos2524.jimdo.com/

$$e_{(i)} = \frac{e_i}{1 - h_{ii}}, \quad i = 1, 2, \ldots, n \tag{4.11}$$

Es fácil de ver en esta ecuación que el residual PRESS no es más que el residual ordinario ponderado por los elementos diagonales de la matriz de sombrero h_{ii}. Los residuales asociados con puntos para los que h_{ii} es grande tendrán PRESS residuales grandes. Esos puntos, en general, serán puntos de **gran influencia**. En general, una gran diferencia entre el residual ordinario y el PRESS residual indica un punto donde el modelo se **ajusta** bien a los datos, pero un modelo formado sin ese punto hace malas **predicciones**. En el capítulo 6 se describirán algunas otras medidas de observaciones influyentes.

Por último, la varianza del i-ésimo PRESS residual es

$$\text{Var}[e_{(i)}] = \text{Var}\left[\frac{e_i}{1 - h_{ii}}\right] = \frac{1}{(1 - h_{ii})^2}[\sigma^2(1 - h_{ii})] = \frac{\sigma^2}{1 - h_{ii}}$$

por lo que un residual PRESS **estandarizado** es

$$\frac{e_{(i)}}{\sqrt{\text{Var}[e_{(i)}]}} = \frac{e/(1 - h_{ii})}{\sqrt{\sigma^2(1 - h_{ii})}} = \frac{e}{\sqrt{\sigma^2(1 - h_{ii})}}$$

que, si se usa MS_{Res} para estimar σ^2 resulta ser el **residual estudentizado** que se describió arriba.

R de Student

El residual estudentizado r_i que se describió arriba se considera con frecuencia como diagnóstico para valores atípicos. Se acostumbra usar MS_{Res} como estimado de σ^2 al calcular r_i. A esto se le llama **escalamiento interno** del residual, porque MS_{Res} es un estimado de σ^2 generado internamente, que se obtiene al ajustar el modelo a las n observaciones. Otro método sería usar un estimado de σ^2 basado en un conjunto de datos con la i-ésima observación eliminada. Al estimado de σ^2 obtenido así se le representa por $S^2_{(i)}$ Se puede demostrar (véase el Ap. C.8) que

$$S^2_{(i)} = \frac{(n - p)MS_{\text{Res}} - e_i^2/(1 - h_{ii})}{n - p - 1} \tag{4.12}$$

El estimado de σ^2 en la ecuación (4.12) se usa en lugar de MS_{Res}, para obtener un **residual estudentizado externamente**, que se suele llamar ***R* de Student** o ***R*-Student**, y se define como

$$t_i = \frac{e_i}{\sqrt{S^2_{(i)}(1 - h_{ii})}}, \quad i = 1, 2, \ldots, n \tag{4.13}$$

En muchos casos t_i será poco distinto del residual estudentizado r_i. Sin embargo, si la i-ésima observación es influyente, entonces $S^2_{(i)}$puede diferir bastante de MS_{Res}, y entonces el estadístico R de student será más sensible a este punto.

http://carlos2524.jimdo.com/

Sucede que bajo las premisas usuales para la regresión, t_i tendrá la distribución t_{n-p-1}. En el apéndice C.9 se establece un procedimiento formal de prueba de hipótesis para detectar valores atípicos, basado en la R de student. Se podría usar un método del tipo de Bonferroni y comparar todos los n valores de $|t_i|$ con $t_{(\alpha/2n),\, n-p-1}$ para tener una guía acerca de los valores atípicos. Sin embargo, los autores consideran que no es necesario un método formal, y que sólo se necesitan considerar valores aproximados de corte. En general, lo mejor es una **perspectiva de diagnóstico**, en contraposición de un punto de vista estricto de prueba de hipótesis. Además, con frecuencia se debe considerar la detección de valores atípicos en forma simultánea con la detección de valores influyentes, como se describe en el capítulo 6.

Ejemplo 4.1 *Datos del tiempo de entrega*

La tabla 4.1 presenta los residuales escalados que se describieron en esta sección, con el modelo de los datos de tiempo de entrega de bebidas gaseosas en el ejemplo 3.1. Al examinar la columna 1 de esa tabla (los residuales ordinarios, calculados antes en la Tabla 3.3) se nota que un residual, $e_9 = 7.4197$, parece anormalmente grande. La columna 2 muestra que el residual estandarizado es $d_9 = e_9 / \sqrt{MS_{\text{Res}}} = 7.4197 / \sqrt{10.6239} = 2.2763$. Todos los demás residuales estandarizados están dentro de los límites de ±2. La columna 3 de la tabla 4.1 contiene los residuales estudentizados. El correspondiente al punto 9 es $r_9 = e_9 / \sqrt{MS_{\text{Res}}(1 - h_{9,9})} = 7.4197 / \sqrt{10.6239(1 - 0.49829)} = 3.2138$, que es bastante mayor que el residual estandarizado. Como se dijo en el ejemplo 3.13, el punto 9 tiene el valor máximo de x_1 (30 cajas) y de x_2 (1 460 pies). Si se tiene en cuenta el lugar lejano del punto 9 al escalar su residual, se llega a la conclusión de que el modelo no se ajusta bien a este punto. Los elementos diagonales de la matriz de sombrero, que se usan en forma extensa para calcular los residuales escalados, están en la columna 4.

La columna 5 de la tabla 4.1 contiene los residuales PRESS. Esos residuales, para los puntos 9 y 22, son bastante más grandes que los residuales ordinarios correspondientes, e indican que es probable que en esos puntos el modelo ajuste en forma razonablemente bien, pero no produce buenas predicciones de datos frescos. Como se observó en el ejemplo 3.13, esos puntos están alejados del resto de la muestra.

La columna 6 contiene los valores de R de Student. Sólo hay un valor, t_9, que es anormalmente grande. Nótese que t_9 es mayor que el residual estudentizado correspondiente, r_9, lo que indica que cuando se aparta la corrida 9, $S^2_{(9)}$ es menor que MS_{Res}, por lo que es claro que esta observación es influyente. Nótese que $S^2_{(9)}$ se calcula a partir de la ecuación (4.12) como sigue:

$$
\begin{aligned}
S^2_{(9)} &= \frac{(n-p)MS_{\text{Res}} - e_9^2/(1 - h_{9,9})}{n-p-1} \\
&= \frac{(22)(10.6239) - (7.4197)^2/(1 - 0.49829)}{21} \\
&= 5.9046
\end{aligned}
$$

4.2.3 Gráficas de residuales

Como se dijo antes, el análisis gráfico de los residuales es una forma muy efectiva de investigar la adecuación del ajuste de un modelo de regresión, y para comprobar las premisas

http://carlos2524.jimdo.com/

TABLA 4.1 Residuales escalados para el ejemplo 4.1

Número de observación, i	$e_i = y_i - \hat{y}_i$ (1)	$d_i = e_i/\sqrt{MS_{\text{Res}}}$ (2)	$r_i = e_i/\sqrt{MS_{\text{Res}}(1-h_{ii})}$ (3)	h_{ii} (4)	$e_{(i)} = e_i/(1-h_{ii})$ (5)	$t_i = e_i/\sqrt{S^2_{(i)}(1-h_{ii})}$ (6)	$[e_i/(1-h_{ii})]^2$ (7)
1	−5.0281	−1.5426	−1.6277	0.10180	−5.5980	−1.6956	31.3373
2	1.1464	0.3517	0.349	0.07070	1.2336	0.3575	1.5218
3	−0.0498	−0.0153	−0.0161	0.09874	−0.0557	−0.0157	0.0031
4	4.9244	1.5108	1.5798	0.05838	5.2297	1.6392	27.3499
5	−0.4444	−0.1363	−0.1418	0.07501	−0.4804	−0.1386	0.2308
6	−0.2896	−0.0888	−0.0908	0.04287	−0.3025	−0.0887	0.0915
7	0.8446	0.2501	0.2704	0.08180	0.9198	0.2646	0.8461
8	1.1566	0.3548	0.3667	0.06373	1.2353	0.3594	1.5260
9	7.4197	2.2763	3.2138	0.49829	14.7888	4.3108	218.7093
10	2.3764	0.7291	0.8133	0.19630	2.9568	0.8068	8.728
11	2.2375	0.6865	0.7181	0.08613	2.4484	0.7099	5.9946
12	−0.5930	−0.1819	−0.1932	0.11366	−0.6690	−0.1890	0.4476
13	1.0270	0.3151	0.3252	0.06113	1.0938	0.3185	1.1965
14	1.0675	0.3275	0.3411	0.07824	1.1581	0.3342	1.3412
15	0.6712	0.2059	0.2103	0.04111	0.7000	0.2057	0.4900
16	−0.6629	−0.2034	−0.2227	0.16594	−0.7948	−0.2178	0.6317
17	0.4364	0.1339	0.1381	0.05943	0.4640	0.1349	0.2153
18	3.4486	1.0580	1.1130	0.09626	3.8159	1.1193	14.5612
19	1.7932	0.5502	0.5787	0.09645	1.9846	0.5698	3.9387
20	−57880	−1.7758	−1.8736	0.10169	−6.4432	−1.9967	41.5150
21	−2.6142	−0.8020	−0.8779	0.16528	−3.1318	−0.8731	9.8084
22	−3.6865	−1.1310	−1.4500	0.39158	−6.0591	−1.4896	36.7131
23	−4.6076	−1.4136	−1.4437	0.04126	−4.8059	−1.4825	23.0966
24	−4.5728	−1.4029	−1.4961	0.12061	−5.2000	−1.5422	27.0397
25	−0.2126	−0.0652	−0.0675	0.06664	−0.2278	−0.0660	0.0519
							PRESS = 457.4000

http://carlos2524.jimdo.com/

básicas. En esta sección se presentarán y se ilustrarán las gráficas básicas de residuales. Se deben examinar en forma rutinaria en todos los problemas de modelado con regresión. Con frecuencia se aconseja graficar tanto los residuales originales como uno o más de los residuales escalados. Normalmente se graficarán aquí los residuales estudentizados, porque tienen varianza constante.

Gráfica de probabilidad normal

Las pequeñas desviaciones respecto a la hipótesis de normalidad no afectan mucho al modelo, pero una no normalidad grande es potencialmente más seria, porque los estadísticos t o F y los intervalos de confianza y de predicción dependen de la suposición de normalidad. Además, si los errores provienen de una distribución con colas más gruesas que la normal, el ajuste por mínimos cuadrados será sensible a un subconjunto menor de datos. Las distribuciones de error con "colas" gruesas generan con frecuencia valores atípicos que "jalan" demasiado en su dirección el ajuste por mínimos cuadrados. En esos casos se deben considerar otras técnicas de estimación (como los métodos de **regresión robusta** del capítulo 11).

Un método muy sencillo de comprobar la suposición de normalidad es trazar una gráfica de **probabilidad normal** de los residuales. Es una gráfica diseñada para que al graficarse la distribución normal acumulada parezca una línea recta. Sean $e_{[1]} < e_{[2]} < \cdots < e_{[n]}$ los residuales ordenados en orden creciente. Si se grafican $e_{[i]}$ en función de la probabilidad acumulada $P_i = (i - \frac{1}{2})/n$, $i = 1, 2, \ldots, n$, en papel de probabilidad normal, los puntos que resulten deberían estar aproximadamente sobre una línea recta. Esa recta se suele determinar en forma visual, con énfasis en los valores centrales (por ejemplo, los puntos de probabilidad acumulada 0.33 y 0.67), y no en los extremos. Las diferencias apreciables respecto a la recta indican que la distribución no es normal. A veces, las gráficas de probabilidad normal se trazan graficando el residual clasificado $e_{[i]}$ en función del "valor normal esperado", $\Phi^{-1}[(i - \frac{1}{2})/n]$, donde Φ representa la distribución acumulada normal estándar. Esto es consecuencia de que $E(e_{[i]}) \simeq \Phi^{-1}[(i - \frac{1}{2})/n]$.

La figura 4.1a muestra una gráfica de probabilidad normal "idealizada". Obsérvese que los puntos caen aproximadamente sobre una recta. Las partes b a e de la gráfica muestran otros problemas característicos. La parte b muestra una curva que va bruscamente hacia arriba y hacia abajo en los dos extremos, lo que indica que las colas de esta distribución son demasiado gruesas para poder considerarla como normal. Al contrario, la parte c muestra un aplanamiento en los extremos, que es un comportamiento característico de las muestras tomadas de una distribución con colas más delgadas que la normal. Las partes d y e de la gráfica muestran patrones asociados con asimetría positiva y negativa, respectivamente.[1]

Ya que las muestras tomadas de una distribución normal no caen exactamente sobre una recta, se requiere cierta experiencia para interpretar las gráficas de probabilidad normal. Daniel y Wood [1980] presentan gráficas se probabilidad normal para tamaños de muestra de 8 a 384. El estudio de esas gráficas ayuda a adquirir un grado de percepción de cuánta desviación de la recta es aceptable. Con frecuencia, los tamaños pequeños de muestra ($n \leq 16$) producen gráficas de probabilidad normal que se desvían bastante de la linealidad. Para muestras mayores ($n \geq 32$), las gráficas se comportan mucho mejor. Por lo general se requieren unos 20 puntos para producir gráficas de probabilidad suficientemente estables como para poder interpretarse con facilidad.

---

[1] En estas interpretaciones se supone que se grafican los residuales clasificados en el eje horizontal. Si se grafican en el eje vertical, como hacen algunos sistemas de cómputo, se invierte la interpretación.

http://carlos2524.jimdo.com/

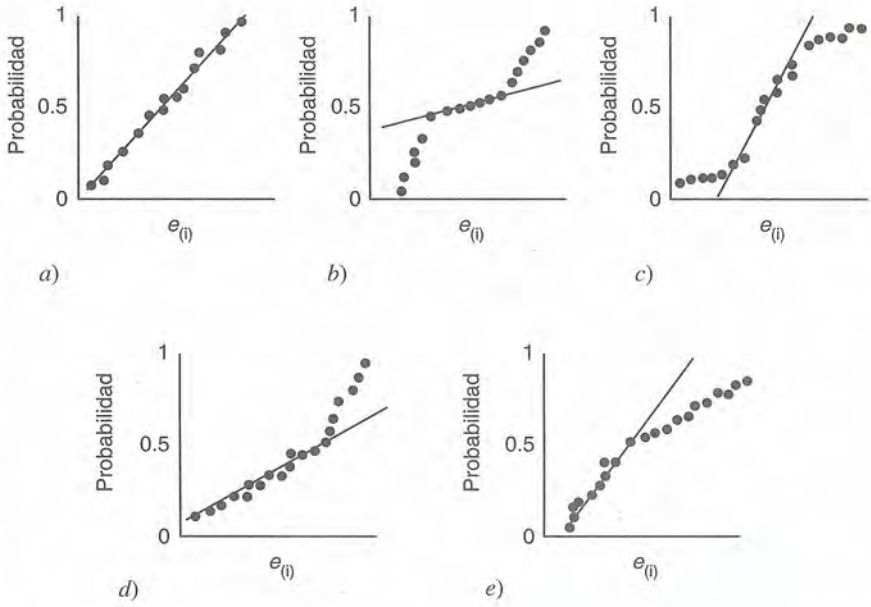


Figura 4.1 Gráficas de probabilidad normal: a) ideal; b) distribución con colas gruesas; c) distribución con colas delgadas; d) asimetría positiva; e) asimetría negativa.

Andrews [1979] y Gnanadesikan [1977] hacen notar que con frecuencia, las gráficas de probabilidad normal no presentan comportamiento anormal, aun cuando los errores ε_i no estén normalmente distribuidos. El problema se debe a que los residuales no son una muestra aleatoria simple; son los restos de un proceso de estimación de parámetros. En realidad, los residuales son combinaciones lineales de los errores del modelo (las ε_i). Así, al ajustar los parámetros se tiende a destruir la evidencia de la no normalidad en los residuales, y en consecuencia no siempre se puede confiar en la gráfica de probabilidad normal para detectar diferencias respecto a la normalidad.

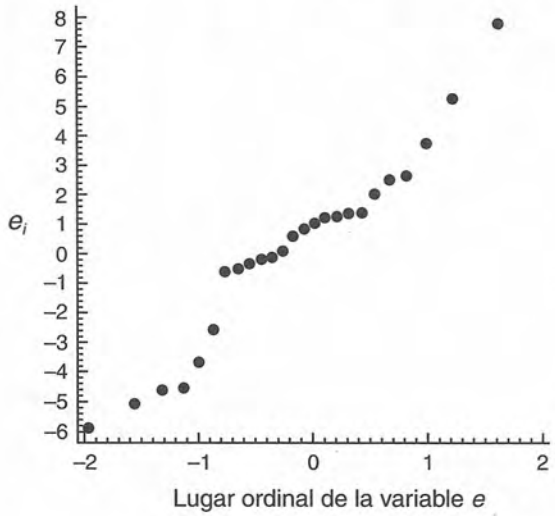
Un defecto común que aparece en la gráfica de probabilidad normal es la ocurrencia de uno o dos residuales grandes. A veces esto es indicativo de que las observaciones correspondientes son **atípicas**. Para una descripción adicional de los puntos atípicos, véase la sección 4.4.

Ejemplo 4.2 *Datos del tiempo de entrega*

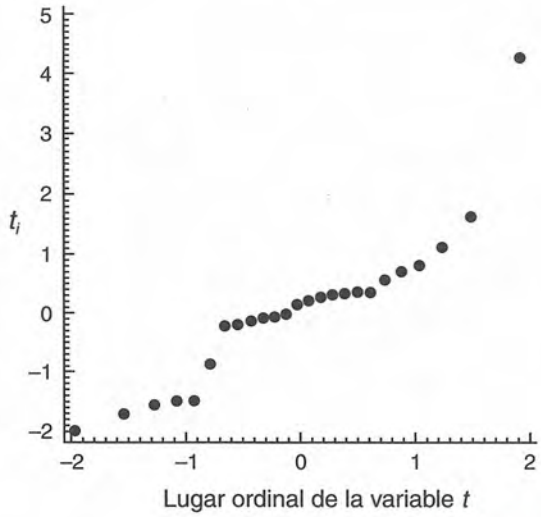
La figura 4.2 representa gráficas de probabilidad normal de los residuales del modelo de regresión para los datos de tiempo de entrega del ejemplo 3.1. Nótese que los residuales se grafican en el eje vertical. Los residuales originales de mínimos cuadrados se grafican en la figura 4.2a, y los residuales R de Student se grafican en la figura 4.2b. Estos residuales se ven en las columnas 1 y 2 de la tabla 4.1.

En ambas gráficas, los residuales no caen exactamente en una línea recta, lo que indica que puede haber algunos problemas con la suposición de normalidad, o que puede haber uno o más valores atípicos en los datos. De acuerdo con el ejemplo 4.1, se sabe que el residual estudentizado para la observación 9 es moderadamente grande ($r_9 = 3.2138$), al igual que el residual R de Student ($r_9 = 4.3108$). Sin embargo, no hay indicios de problemas graves en los datos de tiempo de entrega.

http://carlos2524.jimdo.com/



a)



b)

Figura 4.2 Gráficas de probabilidad normal de los residuales para los datos de tiempo de entrega: a) residuales ordinarios de mínimos cuadrados; b) residuales estudentizados.

Gráfica de residuales en función de los valores ajustados $\hat{y}_i$

Es útil una gráfica de los residuales e_i (o los residuales escalados d_i, r_i o t_i) en función de los valores ajustados correspondientes $\hat{y}_i$, para detectar algunos tipos frecuentes de inadecuaciones del modelo.[2] Si esta gráfica se parece a la de la figura 4.3a, que indica que los residuales se pueden encerrar en una banda horizontal, entonces no hay defectos obvios del modelo. Las gráficas de e_i en función de $\hat{y}_i$ que se parezcan a cualquiera de los patrones de las partes b a d son síntomas de deficiencias del modelo.

[2] Se deben graficar los residuales en función de los valores ajustados $\hat{y}_i$, y no los valores observados y_i, porque las e_i y las $\hat{y}_i$ no están correlacionadas, mientras que las e_i y las y_i suelen estar correlacionadas.

http://carlos2524.jimdo.com/

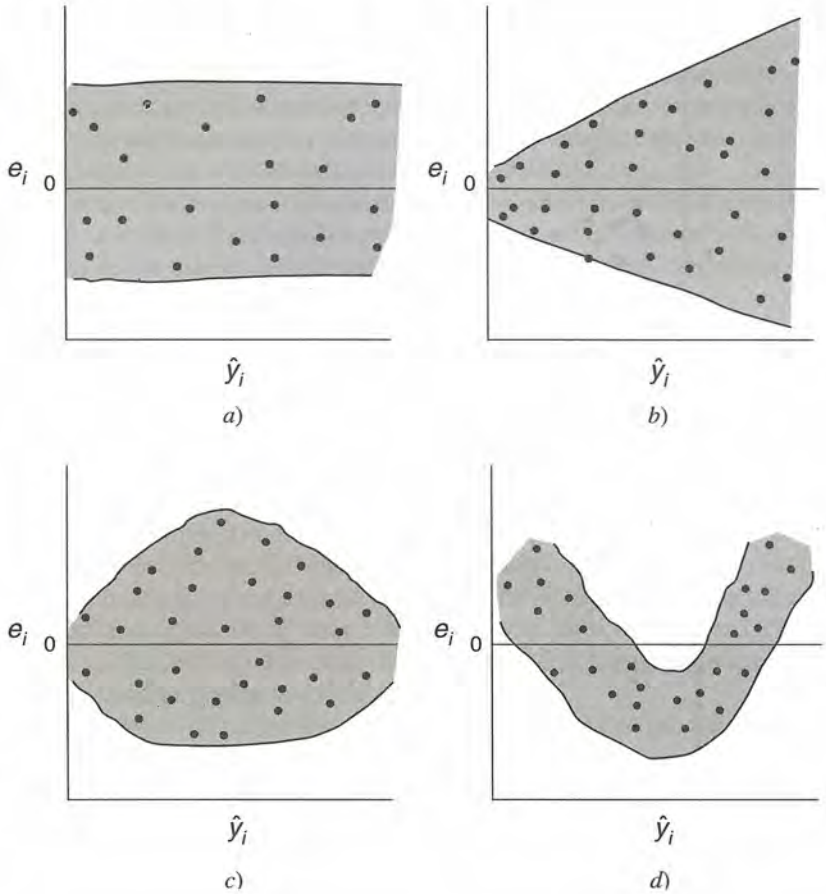


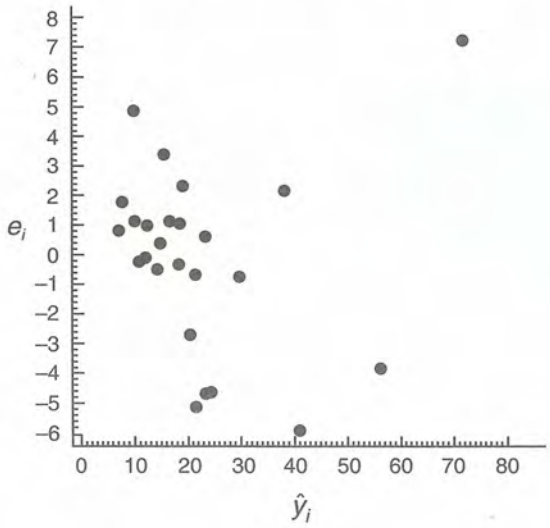
Figura 4.3 Patrones en las gráficas de residuales: a) satisfactorio; b) en embudo; c) en doble arco; d) no lineal.

Las distribuciones en las partes b y c indican que la varianza de los errores no es constante. La **figura de embudo abierto hacia afuera** en la parte b implica que la varianza es función creciente de y. También es posible un embudo abierto hacia dentro, que indica que $\text{Var}(\varepsilon)$ aumenta a medida que y disminuye. La distribución en doble arco en la parte c se presenta con frecuencia cuando y es una proporción entre 0 y 1. La varianza de una proporción binomial cercana a 0.5 es mayor que una cercana a 0 o a 1. El método común para manejar la no constancia de la varianza es aplicar una **transformación** adecuada ya sea a la variable regresora o a la de respuesta (véase el Cap. 5, Sec. 5.2 y 5.3), o usar el método de mínimos cuadrados ponderados (Sec. 5.5). En la práctica, se emplean en general las transformaciones de la variable de respuesta para estabilizar la varianza.

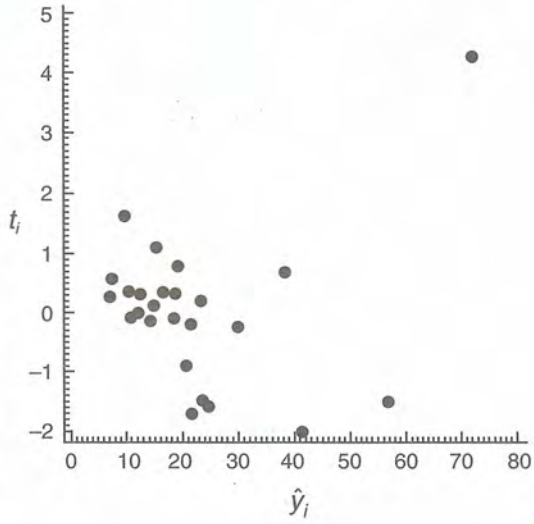
Una gráfica en curva, como la de la parte d, indica **no linealidad**. Esto podría indicar que se necesitan otras variables regresoras en el modelo. Por ejemplo, podría ser necesario un término al cuadrado. Las transformaciones de la variable regresora y/o la de respuesta también podrían ayudar en estos casos.

Una gráfica de los residuales en función de $\hat{y}_i$ también puede revelar uno o más residuales anormalmente grandes. Esos puntos son, es natural, valores atípicos potenciales. Los residuales grandes que están en los valores extremos de $\hat{y}_i$ también podrían indicar que la varianza no es constante, o bien que la relación real entre y y x no es lineal. Se deben investigar esas posibilidades antes de considerar los puntos como atípicos.

http://carlos2524.jimdo.com/



a)



b)

Figura 4.4 Gráfica de residuales en función de valores predichos para los datos de tiempo de entrega: a) residuales originales; b) valores R de Student.

Ejemplo 4.3 *Datos del tiempo de entrega*

La figura 4.4 presenta gráficas de los residuales originales y los valores de R de Student en función de los valores ajustados del tiempo de entrega. Esas gráficas no muestran un patrón fuerte, aunque el residual grande e_9 (o t_9) se destaca con claridad. Parece haber una tendencia ligera del modelo a predecir poco con tiempos de entrega cortos, y a sobreestimar con tiempos de entrega largos.

Gráfica de residuales en función del regresor

También puede ayudar el graficar los residuales en función de los valores correspondientes de cada variable regresora. Esas gráficas suelen presentar patrones como los de la figura

http://carlos2524.jimdo.com/

4.3, excepto que la escala horizontal es x_{ij} para el j-ésimo regresor, y no es $\hat{y}_i$. Una vez más, es deseable la impresión de una banda horizontal que contiene a los residuales. Los patrones en embudo y en doble arco de las partes b y c indican varianza no constante. La banda curva en la parte d, o una distribución no lineal implican, en general, que no es correcta la relación supuesta entre y y el regresor x_j. En ese caso se deberían considerar términos de orden superior en x_j (como x_j^2), o una transformación.

En el caso del regresor lineal simple, no es necesario graficar residuales en función de $\hat{y}_i$ o de la variable regresora. La razón es que los valores ajustados $\hat{y}_i$ son combinaciones lineales de los valores x_i de la variable regresora, por lo que las gráficas sólo serían distintas en la escala de las abscisas.

Ejemplo 4.4 *Datos del tiempo de entrega*

La figura 4.5 presenta las gráficas de valores de R de Student para el problema del tiempo de entrega, ejemplo 3.1, en función de los dos regresores. La parte a es una gráfica de valores de R de Student en función de cajas, y la parte b la gráfica de valores de R de Student en función de la distancia. Ninguna de esas gráficas muestra alguna indicación clara de que hay algún problema, sea de mala especificación del regresor (que implica la necesidad de una transformación del regresor o de términos de orden superior en las cajas y/o la distancia), o sea de no constancia de la varianza, aunque en ambas gráficas es notable el residual moderadamente grande asociado con el punto 9.

También es útil la grafica de residuales en función de las variables regresoras que **no estén ya en el modelo**, pero que potencialmente se pudieran incluir. Todo patrón en la gráfica en función de una variable omitida indicaría que el modelo puede mejorar con la incorporación de esa variable.

El graficar residuales en función de un regresor no siempre es el método más efectivo para revelar si se requiere un efecto de curvatura (o una transformación) para esa variable en el modelo. En la sección 4.2.4 se describirán otras dos gráficas de residuales que son más efectivas, para investigar la relación entre la variable de respuesta y los regresores.

Gráfica de residuales en el tiempo

Si se conoce la secuencia temporal de recolección de los datos, se aconseja graficar los residuales en función de su orden en el tiempo. En el caso ideal, esa gráfica se parecerá a la de la figura 4.3a, esto es, una banda horizontal que abarca todos los residuales, y los residuales varían en forma más o menos aleatoria dentro de esa banda. Sin embargo, si la gráfica se asemeja a las pautas de la figura 4.3b-d, puede indicar que la varianza está cambiando con el tiempo, o que se deben agregar al modelo términos lineales o cuadráticos.

La gráfica de residuales en secuencia temporal puede indicar que los errores en un periodo se correlacionan con los de otros periodos. La correlación entre los errores del modelo en distintos periodos se llama **autocorrelación**. Una gráfica como la de la figura 4.6a indica una autocorrelación positiva, mientras que la figura 4.6b es característica de una autocorrelación negativa. La presencia de autocorrelación es una violación potencialmente grave a las premisas básicas de la regresión. En el capítulo 14 se presentarán más descripciones de métodos para detectar la autocorrelación y las medidas correctivas.

http://carlos2524.jimdo.com/

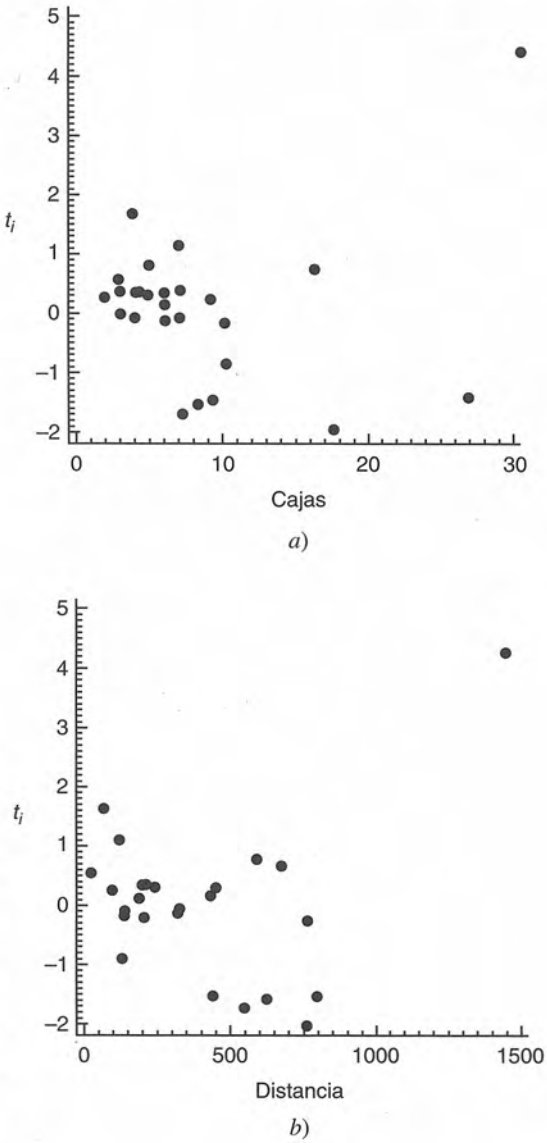


Figura 4.5 Gráfica de valores R de Student en función de los regresores, para los datos de tiempo de entrega: a) valores R de Student en función de las cajas; b) valores R de Student en función de la distancia.

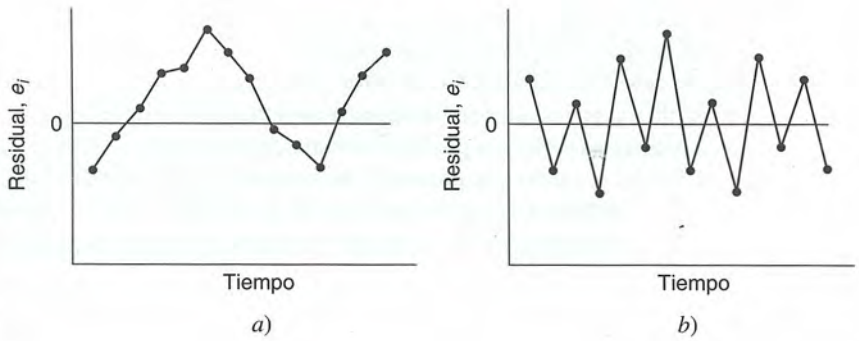


Figura 4.6 Gráficas prototipo de residuales que muestran autocorrelación en los errores: a) autocorrelación positiva; b) autocorrelación negativa.

http://carlos2524.jimdo.com/

4.2.4 Gráficas de regresión parcial y de residuales parciales

Se hizo notar, en la sección 4.2.3, que una gráfica de residuales en función de una variable regresora tiene utilidad para determinar si en el modelo se necesita un efecto de curvatura para ese regresor. Una limitación de esas gráficas es que pueden no mostrar en su totalidad el efecto marginal correcto o completo de un regresor, dados los demás regresores en el modelo. Una **gráfica de regresión parcial** es una variación de la gráfica de residuales en función del predictor, que es una forma ampliada para estudiar la relación marginal de un regresor, dadas las demás variables que están en el modelo. Esa gráfica puede ser muy útil para evaluar si se ha especificado en forma correcta la relación entre la respuesta y las variables regresoras. A veces, a la gráfica de residuales parciales se le llama **gráfica de variable agregada** o **gráfica de variable ajustada**. Las gráficas de regresión parcial también se pueden usar para obtener información acerca de la utilidad marginal de una variable que no esté ya en el modelo.

Las gráficas de regresión parcial consideran el papel marginal del regresor x_j dados los demás regresores que ya están en el modelo. En esta gráfica, se determina la regresión de la variable de respuesta y y el regresor x_j respecto a los demás regresores en el modelo, y de los residuales obtenidos para cada regresión. La gráfica de esos residuales en función de ellos mismos proporciona información sobre la naturaleza de la relación marginal para el regresor x_j que se esté considerando.

Para ilustrar lo anterior, supóngase que se examina un modelo de regresión múltiple de primer orden, con dos variables regresoras: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. Se trata de determinar la naturaleza de la relación marginal para el regresor x_1, en otras palabras, ¿se especificó en forma correcta la relación entre y y x_1?

Primero se haría la regresión de y sobre x_2, y se obtendrían los valores ajustados y los residuales:

$$\hat{y}_i(x_2) = \hat{\theta}_0 + \hat{\theta}_1 x_{i2}$$
$$e_i(y|x_2) = y_i - \hat{y}_i(x_2), \quad i = 1, 2, \ldots, n \tag{4.14}$$

A continuación se hace la regresión de x_1 sobre x_2 y se calculan los residuales:

$$\hat{x}_{i1}(x_2) = \hat{\alpha}_0 + \hat{\alpha}_1 x_{i2}$$
$$e_i(x_1|x_2) = x_{i1} - \hat{x}_{i1}(x_2), \quad i = 1, 2, \ldots, n \tag{4.15}$$

La gráfica de regresión parcial para la variable regresora x_1 se obtiene con los residuales de y, que son $e_i(y|x_2)$ en función de los residuales de x_1: $e_i(x_1|x_2)$. Si el regresor x_1 entra en el modelo en forma lineal, entonces la gráfica de regresión parcial debe mostrar una relación lineal, ya que los residuales parciales estarán a lo largo de una recta con pendiente distinta de cero. La pendiente de esta recta será el coeficiente de regresión de x_1 en el modelo de regresión lineal múltiple. Si la gráfica de regresión parcial muestra una banda curva, quiere decir que términos de orden superior en x_1, podrían ser de utilidad o una transformación (por ejemplo, reemplazar x_1 por $1/x_1$). Cuando x_1 es una variable **candidato** que se esté considerando para incluirla en el modelo, una banda horizontal en la gráfica de regresión parcial indica que no se obtendrá información adicional útil en x_1 para predecir y.

http://carlos2524.jimdo.com/

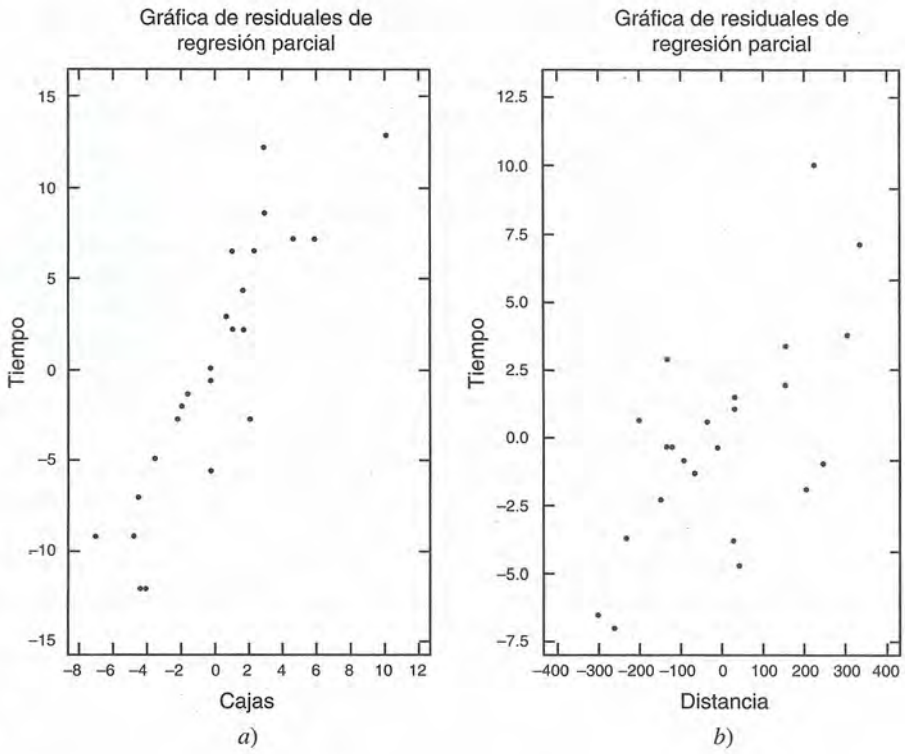


Figura 4.7 Gráficas de residuales parciales para los datos de tiempo de entrega.

Ejemplo 4.5 *Datos del tiempo de entrega*

La figura 4-7 muestra las gráficas de regresión parcial para los datos de tiempo de entrega; la gráfica de x_1 se ve en la figura 4.7a y la gráfica de x_2 en la figura 4.7b. La relación lineal entre ambos casos y la distancia se aprecia con claridad en las dos gráficas, aunque, una vez más, la observación 9 está algo fuera de la recta que describe aparentemente bien el resto de los datos. Eso es otro indicio de que el punto 9 merece mayor investigación.

Algunos comentarios acerca de las gráficas de regresión parcial

1. Las gráficas de regresión parcial deben usarse con precaución, porque sólo **sugieren posibles** relaciones entre el regresor y la respuesta. Esas gráficas podrán no dar información acerca de la forma adecuada de la relación, si ya hay varias variables en el modelo que están mal especificadas. En general, será necesario investigar varias formas alternativas, o varias transformaciones, de la relación entre el regresor y y. Se deben examinar gráficas de residuales para esos modelos siguientes, para identificar la mejor relación o transformación.
2. En general, las gráficas de regresión parcial no detectarán efectos de **interacción** entre los regresores.
3. La presencia de una fuerte multicolinealidad (véase la Sec. 3.8 y el Cap. 10) puede hacer que las gráficas de regresión parcial den como resultado información incorrecta entre la respuesta y las variables regresoras.

http://carlos2524.jimdo.com/

4. Es bastante fácil presentar un desarrollo general del concepto de gráficas de regresión parcial que muestre con claridad por qué la pendiente de la gráfica debe ser el coeficiente de regresión para la variable de interés, por ejemplo x_j.

La gráfica de regresión parcial es de residuales, de los cuales se ha eliminado la dependencia lineal de y en todos los regresores, excepto x_j, en función del regresor x_j. En forma matricial, se pueden escribir esas cantidades en la forma $\mathbf{e}[y \mid \mathbf{X}_{(j)}]$ y $\mathbf{e}[x_j \mid \mathbf{X}_{(j)}]$, respectivamente, siendo $\mathbf{X}_{(j)}$ la matriz $\mathbf{X}$ original, eliminado de ella el j-ésimo regresor (x_j). Para indicar cómo se definen esas cantidades, considérese el modelo

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon = \mathbf{X}_{(j)}\beta + \beta_j \mathbf{x}_j + \varepsilon \tag{4.16}$$

Se premultiplica esta ecuación por $\mathbf{I} - \mathbf{H}_{(j)}$ y se obtiene

$$(\mathbf{I} - \mathbf{H}_{(j)})\mathbf{y} = (\mathbf{I} - \mathbf{H}_{(j)})\mathbf{X}_{(j)}\,\beta + \beta_j(\mathbf{I} - \mathbf{H}_{(j)})\mathbf{x}_j + (\mathbf{I} - \mathbf{H}_{(j)})\varepsilon$$

y se observa que $(\mathbf{I} - \mathbf{H}_{(j)})\mathbf{X}_{(j)} = \mathbf{0}$, así que

$$(\mathbf{I} - \mathbf{H}_{(j)})\mathbf{y} = \beta_j(\mathbf{I} - \mathbf{H}_{(j)})\mathbf{x}_j + (\mathbf{I} - \mathbf{H}_{(j)})\,\varepsilon,$$

o

$$\mathbf{e}[\mathbf{y} \mid \mathbf{X}_{(j)}] = \beta_j \mathbf{e}[\mathbf{x}_j \mid \mathbf{X}_{(j)}] + \varepsilon^*$$

siendo $\varepsilon^* = (\mathbf{I} - \mathbf{H}_{(j)})\varepsilon$. Esto sugiere que una gráfica de regresión parcial debería tener la pendiente β_j. Así, si x_j entra a la regresión en forma lineal, la gráfica de regresión parcial debe mostrar una relación lineal que pase por el origen. Muchos programas de cómputo (como SAS y Minitab) generan gráficas de regresión parcial.

Gráficas de residuales parciales

Una gráfica de residuales que se relaciona en forma estrecha con la gráfica de regresión parcial es la **gráfica de residuales parciales**. Su diseño es tal que debe mostrar la relación entre la variable de respuesta y los regresores. Supóngase que el modelo contiene los regresores $x_1, x_2, \ldots, x_k$. Los **residuales parciales** para el regresor x_j se definen como sigue:

$$e_i^*(y \mid x_j) = e_i + \hat{\beta}_j x_{ij}, \qquad i = 1, 2, \ldots, n$$

donde las e_i son los residuales del modelo, incluyendo todos los k regresores. Cuando se grafican los residuales parciales en función de x_{ij}, la figura que resulta tiene la pendiente $\hat{\beta}_j$, que es el coeficiente de regresión asociado con x_j en el modelo. La interpretación de la gráfica de residuales parciales se parece mucho a la de regresión parcial. Para conocer más detalles y ejemplos, véase Larsen y McCleary [1972], Daniel y Wood [1980], Wood [1973], Mallows [1986], Mansfield y Conerly [1987] y Cook [1993].

4.2.5 Otras gráficas de residuales y métodos de análisis

Además de las gráficas básicas de residuales que se describieron en las secciones 4.2.3 y 4.2.4, hay otras que en ocasiones pueden ser de utilidad. Por ejemplo, podría ser muy útil trazar una gráfica de dispersión del regresor x_i en función del regresor x_j. Esta gráfica podría

http://carlos2524.jimdo.com/

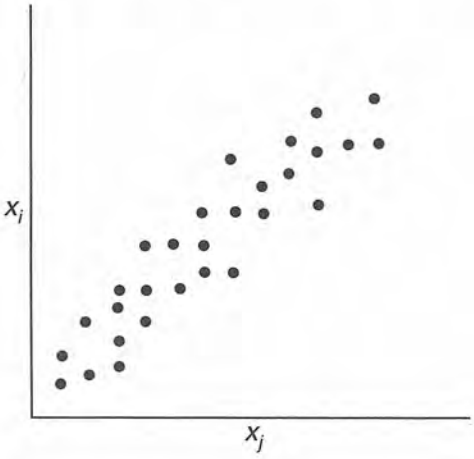


Figura 4.8 Gráfica de x_i en función de x_j.

ayudar en el estudio de la relación entre las variables regresoras y la distribución de los datos en el espacio de x. Se considerará la gráfica de x_i en función de x_j de la figura 4.8. Esta distribución indica que x_i y x_j están muy correlacionadas positivamente. En consecuencia, podrá no ser necesario incluir ambos regresores en el modelo. Si dos o más regresores están muy correlacionados, es posible que haya multicolinealidad en los datos. Como se observó en el capítulo 3 (Sec. 3.8), la multicolinealidad puede perturbar mucho el ajuste por mínimos cuadrados, y en algunos casos, hacer que el modelo de regresión sea casi inútil. Las gráficas de x_i en función de x_j también pueden usarse para descubrir puntos alejados del resto de los datos, que en forma potencial influyen sobre propiedades clave del modelo. Anscombe [1973] presenta algunos otros tipos de gráficas entre regresores. Cook y Weisberg [1994] describen un tratamiento muy moderno de las gráficas de regresión, que incluye muchas técnicas avanzadas que no se mencionan en este libro.

La figura 4.9 es un diagrama de dispersión de x_1 (cajas) en función de x_2 (distancia) para los datos tiempo de entrega del ejemplo 3.1 (Tabla 3.2). Al comparar la figura 4.9 con la figura 4.8, se ve que las cajas y la distancia tienen correlación positiva. De hecho, la correlación simple entre x_1 y x_2 es $r_{12} = 0.82$. Si bien los regresores muy correlacionados pueden causar varios problemas graves en la regresión, en este ejemplo no hay síntomas

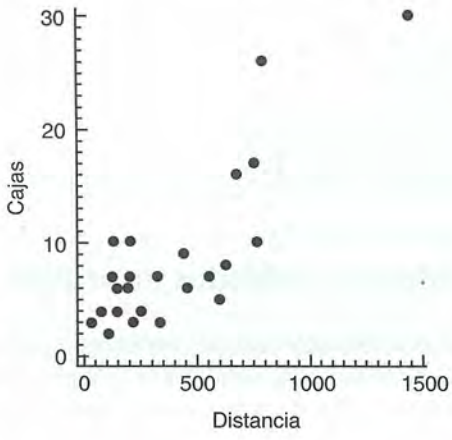


Figura 4.9 Gráfica del regresor x_1 (cajas) en función del regresor x_2 (distancia) para los datos de tiempo de entrega en la tabla 3.2.

http://carlos2524.jimdo.com/

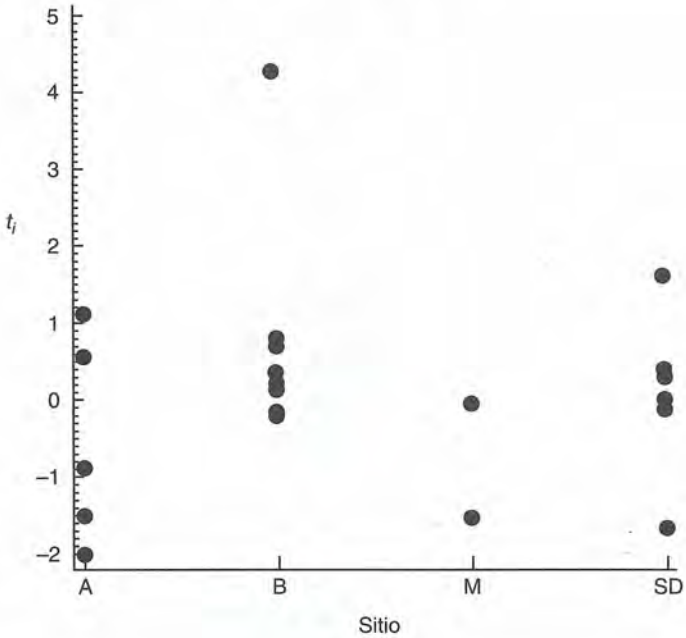


Figura 4.10 Gráfica de valores R de Student por sitio (ciudad), para los datos de tiempo de entrega en la tabla 3.2.

de que se hayan presentado problemas. El diagrama de dispersión muestra con claridad que la observación 9 es atípica con respecto a las cajas y a la distancia ($x_1 = 30$, $x_2 = 1\,460$); de hecho, está bastante alejado, en el espacio de x, del resto de los datos. La observación 22 ($x_1 = 26$, $x_2 = 810$) también está bastante alejada del resto de los datos. Los puntos alejados en el espacio de x pueden controlar, en forma potencial, algunas de las propiedades del modelo de regresión. En el capítulo 6 se describirán otros modelos formales para estudiar lo anterior.

El **caso** o la **situación del problema** parece indicar, con frecuencia, el uso de otros tipos de gráficas de residuales. Por ejemplo, para los tiempos de entrega del ejemplo 3.1, las 25 observaciones de la tabla 3.2 se recolectaron en rutas de camión y en cuatro ciudades distintas. Las observaciones 1 a 7 se hicieron en San Diego, las observaciones 8 a 17 en Boston, las observaciones 18 a 23 en Austin, y las observaciones 24 y 25 en Minneapolis. Se podría creer que hay una diferencia, de ciudad a ciudad, en las operaciones de entrega, debido a factores tales como equipos distintos, grados distintos de adiestramiento y de experiencia de cuadrillas, o factores motivacionales influidos por las políticas administrativas. Esos factores podrían dar como resultado un efecto de "sitio" que no se ha incorporado en la ecuación que se obtuvo. Para investigar esta posibilidad, se graficaron los residuales, identificados por sitio, en la figura 4.10. Se ve allí que hay algo de desequilibrio en la distribución de residuales positivos y negativos en cada sitio. En forma específica, hay una tendencia aparente para que el modelo estime mayores tiempos de entrega en Austin, y que prediga menores tiempos en Boston. Esto podría deberse a los factores dependientes de sitio mencionados arriba, o a que se hayan omitido uno o más regresores importantes del modelo.

http://carlos2524.jimdo.com/

Pruebas estadísticas con los residuales

Se pueden aplicar pruebas estadísticas a los residuales, para obtener medidas cuantitativas de algunas de las inadecuaciones del modelo que se describieron arriba. Véanse, por ejemplo, Anscombe [1961, 1967], Anscombe y Tukey [1963], Andrews [1971], Looney y Gulledge [1985], Levene [1960] y Cook y Weisberg [1983]. Se describen varios procedimientos formales de prueba estadística para residuales en Draper y Smith [1998] y en Neter, Kutner, Nachtsheim y Wasserman [1996].

De acuerdo con nuestra experiencia, las pruebas estadísticas con residuales del modelo de regresión no se usan mucho. En la mayor parte de los casos prácticos, las gráficas de residuales son más informativas que las pruebas correspondientes. Sin embargo, como las gráficas de residuales sí requieren destreza y experiencia para su interpretación, a veces las pruebas estadísticas pueden ser más útiles. Para conocer un buen ejemplo del empleo de pruebas estadísticas junto con gráficas, véase Feder [1974].

4.3 LA ESTADÍSTICA PRESS

En la sección 4.2.2 se definieron los residuales PRESS como $e_{(i)} = y_i - \hat{y}_{(i)}$, siendo $\hat{y}_{(i)}$ el valor predicho de la i-ésima respuesta observada, basado en un ajuste de modelo con los $n - 1$ puntos de muestra. Se observó que los residuales PRESS grandes tienen una utilidad potencial para identificar observaciones donde el modelo no se ajusta bien a los datos, o para observaciones en las que es probable que el modelo produzca malas predicciones futuras.

Allen [1971, 1974] ha sugerido el uso de la Suma de Cuadrados de Error de Predicción (es decir, la estadística PRESS), definida como la suma de los residuales PRESS al cuadrado, como medida de la calidad del modelo. La estadística PRESS es

$$\begin{aligned} \textbf{PRESS} &= \sum_{i=1}^{n} \left[y_i - \hat{y}_{(i)} \right]^2 \\ &= \sum_{i=1}^{n} \left(\frac{e_i}{1 - h_{ii}} \right)^2 \end{aligned} \tag{4.17}$$

Se considera que PRESS es una medida de lo bien que funciona un modelo de regresión para **predecir nuevos datos**. Lo deseable es tener un modelo con valor pequeño de PRESS.

Ejemplo 4.6 *Datos del tiempo de entrega*

La columna 5 de la tabla 4.1 muestra los cálculos de los residuales PRESS para los datos de tiempo de entrega del ejemplo 3.1. La columna 7 de esa tabla contiene los residuales PRESS al cuadrado, y al pie de esa columna se ve la estadística PRESS. El valor de PRESS = 457.4000 es casi el doble que el de la suma de cuadrados de residuales para este modelo, $SS_{\text{Res}} = 233.7260$. Obsérvese que casi la mitad de la estadística PRESS se debe al punto 9, que está relativamente alejado en el espacio de x, y tiene un residual moderadamente grande. Esto indica que es posible que el modelo no prediga muy bien nuevas observaciones con cantidades grandes de cajas, y distancias grandes.

http://carlos2524.jimdo.com/

R^2 para predicción basado en PRESS

Con la estadística PRESS se puede calcular un estadístico parecido a la R^2 para predicción, por ejemplo

$$R^2_{\text{predicción}} = 1 - \frac{\textbf{PRESS}}{SS_T} \tag{4.18}$$

Este estadístico da cierta indicación de la capacidad predictiva del modelo de regresión. Para el modelo del tiempo de entrega de bebidas gaseosas se calcula:

$$\begin{aligned} R^2_{\text{predicción}} &= 1 - \frac{\textbf{PRESS}}{SS_T} \\ &= 1 - \frac{457.4000}{5\,784.5426} \\ &= 0.9209 \end{aligned}$$

Por consiguiente, cabría esperar que este modelo "explicara" un 92.09% de la variabilidad cuando se predigan nuevas observaciones, en comparación con el 95.96% de variabilidad en los datos originales, explicado por el ajuste de mínimos cuadrados. Parece satisfactoria la capacidad predictiva del modelo, en general. Sin embargo, recuérdese que los residuales PRESS individuales indicaron que puede ser que las observaciones parecidas al punto 9 no sean bien pronosticadas.

Uso de PRESS para comparar modelos

Una aplicación muy importante de la estadística PRESS es comparar modelos de regresión. En general, un modelo con pequeño valor de PRESS es preferible a uno con PRESS grande. Por ejemplo, cuando se agregó x_2 = distancia, al modelo de regresión para los datos de tiempo de entrega, que ya contenía x_1 = cajas, el valor de PRESS disminuyó de 733.55 a 457.40. Es una indicación que es probable que el modelo con dos regresores sea mejor predictor que el modelo que sólo contiene x_1 = cajas.

4.4 DETECCIÓN Y TRATAMIENTO DE PUNTOS ATÍPICOS

Un valor atípico es una observación extrema. Los residuales cuyo valor absoluto es bastante mayor que los demás, digamos de tres a cuatro desviaciones estándar respecto a la media, indican que hay valores atípicos potenciales en el espacio de y. Los valores atípicos son puntos que no son representativos del resto de los datos. De acuerdo con su ubicación en el espacio de x, los valores atípicos pueden tener efectos de moderados a graves sobre el modelo de regresión (por ejemplo, véase las Figs. 2.6 a 2.8). Las gráficas de residuales en función de $\hat{y}_i$ y la gráfica de probabilidad normal son útiles para identificar puntos atípicos. El examen de los **residuales escalados**, como por ejemplo los residuales estudentizados y los R de Student es una forma excelente de identificar puntos atípicos potenciales.

http://carlos2524.jimdo.com/

Una excelente descripción general de los problemas con valores atípicos se encuentra en Barnett y Lewis [1994]. También véase Myers [1990], para una buena discusión del tema.

Los valores atípicos se deben investigar con cuidado, para ver si se puede encontrar una razón de su comportamiento extraordinario. A veces, los valores atípicos son "malos" y se deben a eventos desacostumbrados, pero explicables. Entre los ejemplos están la medición o el análisis incorrectos, el registro incorrecto de los datos y la falla de un instrumento de medición. Si éste es el caso, el valor atípico se debería corregir (si es posible) o eliminar del conjunto de datos. Es claro que el eliminar valores malos es conveniente, porque los mínimos cuadrados jalan la ecuación ajustada hacia el valor atípico, ya que eso minimiza la suma de cuadrados de residuales, sin embargo, se hace notar que debe contarse con una fuerte evidencia no estadística de que el valor atípico es malo, para entonces descartarlo.

A veces se encuentra que el valor atípico es una observación extraordinaria, pero perfectamente factible. Puede ser peligroso eliminar estos puntos para "mejorar el ajuste de la ecuación", porque puede dar al usuario una sensación falsa de precisión de la estimación o la predicción. A veces se ve que el valor atípico es más importante que el resto de los datos, porque puede controlar muchas propiedades clave del modelo. También, los valores atípicos pueden hacer resaltar inadecuaciones en el modelo, como la falla de tener un buen ajuste con los datos en cierta región del espacio de x. Si el valor atípico es un punto de respuesta especialmente deseable (por ejemplo, bajo costo o alto rendimiento), sería en extremo valioso conocer los valores de los regresores, cuando se observó esa respuesta. Los análisis de identificación y de seguimiento de los valores atípicos con frecuencia dar como resultado mejoras en el proceso, o nuevos conocimientos acerca de factores cuyo efecto sobre la respuesta se desconocía antes.

Se han propuesto diversas pruebas estadísticas para detectar y rechazar los valores atípicos. Por ejemplo, véase Barnett y Lewis [1994]. Stefansky [1971, 1972] ha propuesto una prueba aproximada para identificar puntos atípicos, basada en el residual máximo normado $|e_i|/\sqrt{\Sigma_{i=1}^{n} e_i^2}$, cuya aplicación es bastante fácil. Ejemplos de esta prueba y de otras referencias relacionadas se encuentran en Cook y Prescott [1981], Daniel [1976] y Williams [1973], además véase también el apéndice C.9. Si bien esas pruebas pueden tener utilidad para identificar valores atípicos, no se debe creer que los puntos identificados con ellas deben eliminarse en forma automática. Como se ha dicho, esos puntos pueden ser claves importantes que contienen información valiosa.

El efecto de los valores atípicos sobre el modelo de regresión se puede comprobar con facilidad eliminándolos y volviendo a ajustar la ecuación de regresión. Se podrá encontrar que los valores de los coeficientes de regresión, o de los estadísticos de resumen como t, F o R^2, y que el cuadrado medio de residuales pueden ser muy sensibles a los valores atípicos. Los casos en los que un porcentaje relativamente pequeño de los datos tiene un gran impacto sobre el modelo podrán no ser aceptables para el usuario de la ecuación de regresión. En general, uno se siente más cómodo suponiendo que una ecuación de regresión es válida si no es muy sensible a unas pocas observaciones. Se preferiría que la relación de regresión estuviera embebida en todas las observaciones, y no sólo fuera un artificio de unos pocos puntos.

Ejemplo 4.7 *Datos del propelente de reacción*

La figura 4.11 muestra la gráfica de probabilidad normal y la gráfica de valores R de Student, en función de las $\hat{y}_i$ predichas para los datos del propelente de reacción presentados en el ejemplo 2.1. Se ve que hay dos residuales negativos grandes, muy apartados del resto (observaciones 5 y 6, en la Tabla 2.1). Esos puntos son atípicos potenciales; tienden a producir

http://carlos2524.jimdo.com/

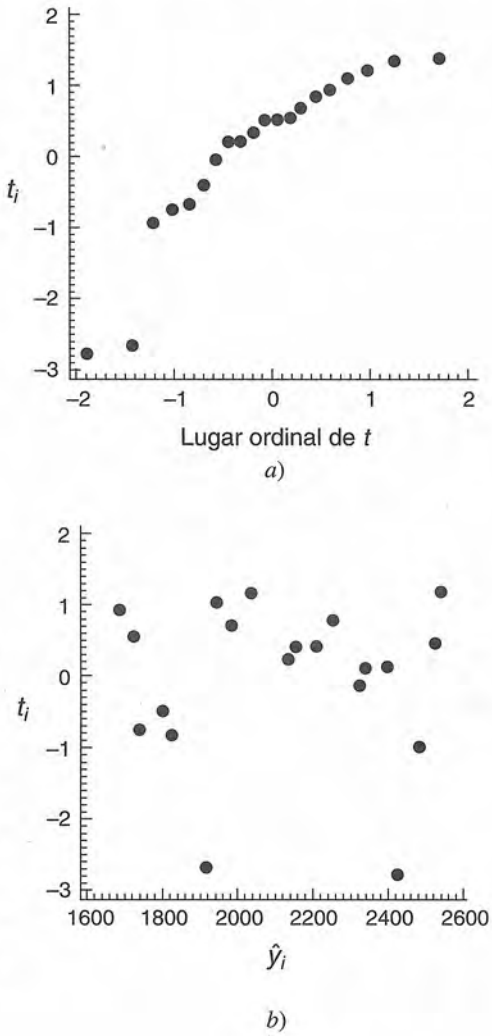


Figura 4.11 Gráficas de residuales para los datos del propelente de reacción: a) gráfica de probabilidad normal; b) gráfica de valores R de Student en función de las $\hat{y}_i$ pronosticadas.

una gráfica de probabilidad normal con aspecto asimétrico. Nótese que la observación 5 se presenta a un valor relativamente bajo de la edad (5.5 semanas) y que la observación 6 está en un valor relativamente grande de la edad (19 semanas). Por lo anterior, esos dos puntos están muy separados en el espacio de x, y se presentan cerca de los valores extremos de x, y pueden ser influyentes para determinar las propiedades del modelo. Aunque ninguno de esos residuales es demasiado grande, la impresión general que se obtiene de las gráficas de residuales (Fig. 4.11) es que esas dos observaciones son marcadamente distintas de las demás.

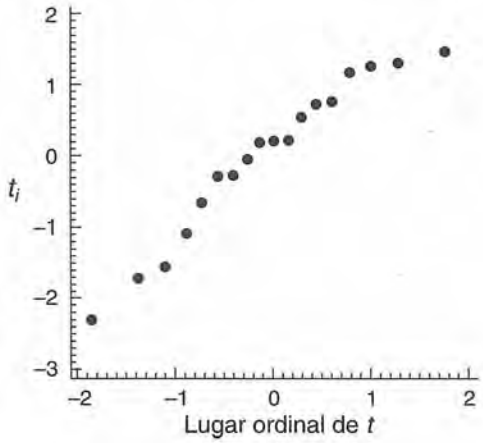
Para investigar la influencia de esos dos puntos sobre el modelo, se obtiene una nueva ecuación de regresión sin las observaciones 5 y 6. A continuación se muestra una comparación de los estadísticos de resumen para esos dos modelos:

http://carlos2524.jimdo.com/

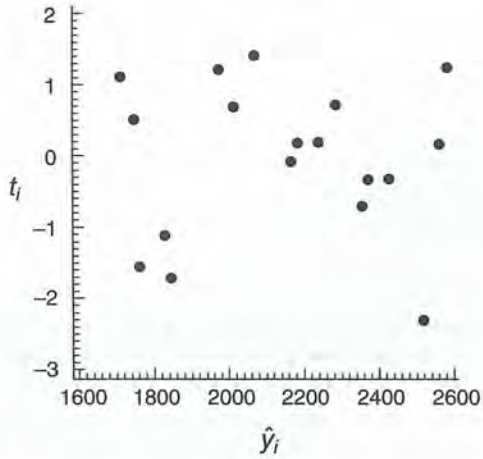
	Con observaciones 5 y 6	Sin observaciones 5 y 6
$\hat{\beta}_0$	2 627.82	2 658.97
$\hat{\beta}_1$	– 37.15	– 37.69
R^2	0.9018	0.9578
MS_{Res}	9 244.59	3 964.63
se($\hat{\beta}_1$)	2.89	1.98

Al quitar los puntos 5 y 6 casi no se observa efecto alguno sobre los estimados de los coeficientes de regresión. Sin embargo, sí hay una gran reducción en el cuadrado medio de residuales, un aumento moderado de R^2 y una reducción aproximada de la tercera parte en el error estándar de $\hat{\beta}_1$.

Como los estimados de los parámetros no cambiaron mucho, se llega a la conclusión de que los puntos 5 y 6 no son demasiado influyentes. Están algo fuera de la recta que pa-



a)



b)

Figura 4.12 Gráficas de residuales para los datos de propelente de reacción, eliminadas las observaciones 5 y 6: a) gráfica de probabilidad normal; b) gráfica de valores R de Student en función de las $\hat{y}_i$ pronosticadas.

http://carlos2524.jimdo.com/

sa por los otros 18 puntos, pero no controlan ni la pendiente ni la ordenada al origen. Sin embargo, esos dos residuales forman el 56%, aproximadamente, de la suma de cuadrados de residuales. Así, si esos puntos fueran realmente valores malos, y se debieran omitir, mejoraría la precisión de los estimados de parámetro, y se reduciría bastante el ancho de los intervalos de confianza y de predicción. La figura 4.12 muestra la gráfica de probabilidad normal de los residuales, y la de los valores de R de Student en función de $\hat{y}_i$ para el modelo con los puntos 5 y 6 omitidos. Esas gráficas no indican diferencias graves respecto a las hipótesis.

Un examen posterior de los puntos 5 y 6 no puede revelar razón alguna de las resistencias al cortante tan extraordinariamente bajas que se obtuvieron. En consecuencia, no se deben desechar esos dos puntos. Sin embargo, se tiene la confianza relativa que al incluirlos no se limitará mucho el uso del modelo.

4.5 FALTA DE AJUSTE DEL MODELO DE REGRESIÓN

4.5.1 Una prueba formal de la falta de ajuste

Ahora se describirá una prueba estadística formal de la falta de ajuste de un modelo de regresión. En el procedimiento se supone que se cumplen los requisitos de normalidad, independencia y varianza constante, y que sólo es dudoso el carácter de primer orden, o de línea recta, de la relación. Por ejemplo, véase los datos de la figura 4.13. Hay cierta indicación que el ajuste a la línea recta no es muy satisfactorio, y sería bueno contar con un procedimiento de prueba para determinar si hay una curvatura sistemática.

En la prueba de falta de ajuste se requiere contar con observaciones duplicadas, o *replicadas*, de la respuesta de y cuando menos para un nivel de x. Se subraya que deben ser réplicas verdaderas y no sólo indicaciones o mediciones duplicadas de y. Por ejemplo, su-

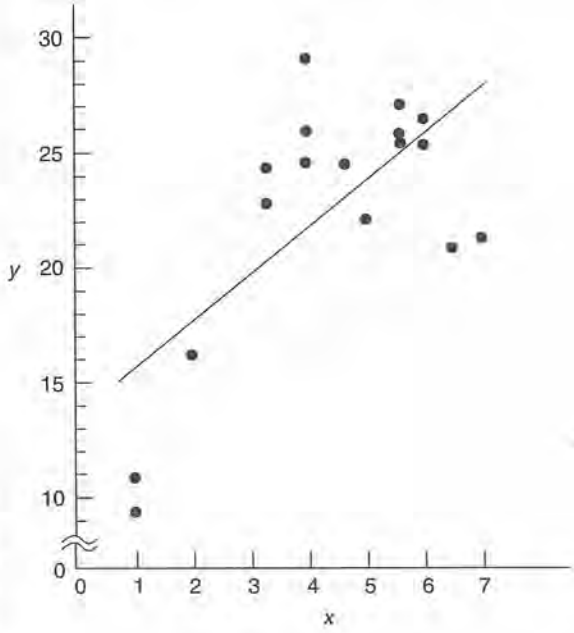


Figura 4.13 Datos que indican falta de ajuste del modelo de línea recta.

http://carlos2524.jimdo.com/

póngase que y es la viscosidad del producto y que x es la temperatura. La réplica verdadera consiste en correr n_i experimentos separados con $x = x_i$, y observar la viscosidad, y no tan sólo correr un solo experimento en x_i y medir la viscosidad n_i veces. Los valores obtenidos con este último procedimiento sólo dan información sobre la variabilidad del método de medición de viscosidad. La varianza del error, σ^2, incluye este error de medición, y la variabilidad asociada con el logro y el mantenimiento del mismo valor de temperatura en distintos experimentos. Esas observaciones *replicadas* se usan para obtener un **estimado de σ^2 independiente del modelo**.

Supóngase que se cuenta con n_i observaciones de la respuesta al i-ésimo nivel del regresor x_i, $i = 1, 2, \ldots, m$. Sea y_{ij} la j-ésima observación de la respuesta en x_i, $i = 1, 2, \ldots, m$, y $j = 1, 2, \ldots, n_i$. Hay $n = \sum_{i=1}^{m} n_i$ observaciones en total. El procedimiento de prueba implica agrupar la suma de cuadrados de residuales en dos componentes, por ejemplo

$$SS_{\text{Res}} = SS_{\text{EP}} + SS_{\text{FA}},$$

en donde SS_{EP} es la suma de cuadrados debida al **error puro** y SS_{FA} es la suma de cuadrados debidos a la **falta de ajuste**.

Para desarrollar esta partición de SS_{Res}, obsérvese que el i, j-ésimo residual es

$$y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i) \tag{4.19}$$

siendo $\bar{y}_i$ el promedio de las n_i observaciones en x_i. Al elevar ambos lados de esta ecuación y sumar sobre i y j se obtiene

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{m} n_i (\bar{y}_i - \hat{y}_i)^2 \tag{4.20}$$

porque el término de productos cruzados es igual a cero.

El lado derecho de la ecuación (4.20) es la suma acostumbrada de residuales. Los dos componentes del lado derecho miden el error puro y la falta de ajuste. Se ve que la suma de cuadrados por error puro

$$SS_{\text{EP}} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \tag{4.21}$$

se obtiene calculando la suma corregida de cuadrados de las observaciones repetidas en cada nivel de x, y a continuación agrupando sobre los m niveles de x. Si se satisface la hipótesis de varianza constante, ésta es una **medida del error puro independiente del modelo**, porque sólo se usa la variabilidad de las y en cada nivel de x para calcular SS_{EP}. Como hay $n_i - 1$ grados de libertad para el error puro en cada nivel de x_i, la cantidad total de grados de libertad asociada con la suma de cuadrados de error puro es

$$\sum_{i=1}^{m} (n_i - 1) = n - m \tag{4.22}$$

La suma de cuadrados de falta de ajuste,

$$SS_{\text{FA}} = \sum_{i=1}^{m} n_i (\bar{y}_i - \hat{y}_i)^2 \tag{4.23}$$

http://carlos2524.jimdo.com/

es una suma ponderada de desviaciones elevadas al cuadrado, de la respuesta promedio $\hat{y}_i$, en cada nivel de x, y el valor ajustado correspondiente. Si los valores ajustados $\hat{y}_i$ son cercanos a las respuestas promedio correspondientes $\bar{y}_i$, se tendrá una fuerte indicación de que la función de regresión es lineal. Si las $\hat{y}_i$ se desvían mucho de las $\bar{y}_i$, es probable que la función de regresión no sea lineal. Hay $m - 2$ grados de libertad asociados con SS_{FA}, porque hay m niveles de x y se pierden dos grados de libertad, porque se deben estimar dos parámetros para obtener las $\bar{y}_i$. En los cálculos, se suele obtener SS_{FA} restando SS_{EP} de SS_{Res}.

El estadístico de prueba para la falta de ajuste es

$$F_0 = \frac{SS_{FA} \;/(m-2)}{SS_{EP}\,/(n-m)} = \frac{MS_{FA}}{MS_{EP}} \tag{4.24}$$

El valor esperado de MS_{EP} es σ^2, y el valor esperado de MS_{FA} es

$$E(MS_{FA}) = \sigma^2 + \frac{\sum_{i=1}^{m} n_i \left[E(y_i) - \beta_0 - \beta_1 x_i\right]^2}{m-2} \tag{4.25}$$

Si la función verdadera de regresión es lineal, entonces $E(y_i) = \beta_0 + \beta_1 x_i$, y el segundo término de la ecuación (4.25) es cero; esto da como resultado $E(MS_{FA}) = \sigma^2$. Sin embargo, si la función verdadera de regresión no es lineal, entonces $E(y_i) \neq \beta_0 + \beta_1 x_i$, y $E(MS_{FA})$ es $> \sigma^2$. Además, si la verdadera función de regresión es lineal, entonces el estadístico F_0 sigue la distribución $F_{m-2,\,n-m}$. Por consiguiente, para probar la falta de ajuste se calcularía el estadístico de prueba F_0 y se llegaría a la conclusión de que la función de regresión no es lineal si $F_0 > F_{\alpha,\,m-2,\,n-m}$.

Este procedimiento de prueba se puede introducir con facilidad en el análisis de varianza para determinar el significado de la regresión. Si se concluye que la función de regresión no es lineal, se debe abandonar el modelo tentativo y se debe tratar de formular una ecuación más adecuada. En forma alternativa, si F_0 no es mayor que $F_{\alpha,\,m-2,\,n-m}$, no hay pruebas suficientes de falta de ajuste, y con frecuencia se combinan MS_{EP} y MS_{FA} para estimar σ^2.

En el caso ideal se ve que la relación F de la falta de ajuste no es significativa, y se rechaza la hipótesis de la significancia de la regresión (H_0: $\beta_1 = 0$). Desafortunadamente eso no garantiza que el modelo sea satisfactorio como ecuación de predicción. A menos que la variación de los valores predichos sea grande en relación con el error aleatorio, el modelo no se estima con la precisión suficiente para producir pronósticos satisfactorios. Esto es, el modelo puede estar ajustado para los errores solamente. Se han hecho trabajos analíticos para desarrollar criterios para juzgar la adecuación del modelo de regresión desde un punto de vista predictivo. Véase Box y Wetz [1973], Ellerton [1978], Gunst y Mason [1979], Hill, Judge y Fomby [1978], y Suich y Derringer [1977]. El trabajo de Box y Wetz parece indicar que la relación F debe ser cuando menos cuatro o cinco veces mayor que el valor crítico de la tabla F para que el modelo de regresión sea útil como predictor, esto es, si la dispersión de los valores pronosticados debe ser grande con respecto al ruido.

Una medida relativamente sencilla de la eficiencia potencial de predicción se determina comparando el recorrido de los valores ajustados $\hat{y}_i$ (es decir, $\hat{y}_{máx} - \hat{y}_{mín}$), con su error estándar promedio. Se puede demostrar que, independientemente de la forma del modelo, la varianza promedio de los valores ajustados es

$$\overline{\mathrm{Var}(\hat{y})} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(\hat{y}_i) = \frac{p\sigma^2}{n} \tag{4.26}$$

http://carlos2524.jimdo.com/

TABLA 4.2 Análisis de varianza para el ejemplo 4.8

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Regresión	234.7087	1	234.7087		
Residuales	252.9039	15	16.8603		
(Falta de ajuste)	237.3407	8	29.6676	13.34	0.0013
(Error puro)	15.5632	7	2.2233		
Total	487.6126	16			

en la que p es la cantidad de parámetros en el modelo. En general, no es probable que el modelo sea un predictor satisfactorio, a menos que el rango de los valores ajustados $\hat{y}_1$ sea grande en relación con su error estándar promedio estimado, $\sqrt{(p\hat{\sigma}^2)/n}$, donde $\hat{\sigma}^2$ es un estimado de la varianza de error independiente del modelo.

Ejemplo 4.8 *Prueba de falta de ajuste*

Los datos de la figura 4.13 se ven a continuación:

x	1.0	1.0	2.0	3.3	3.3	4.0	4.0	4.0	4.7	5.0
y	10.84	9.30	16.35	22.88	24.35	24.56	25.86	29.16	24.59	22.25
x	5.6	5.6	5.6	6.0	6.0	6.5	6.9			
y	25.90	27.20	25.61	25.45	26.56	21.03	21.46			

El ajuste de línea recta es $\hat{y} = 13.301 + 2.108x$, con $SS_T = 487.6126$, $SS_R = 234.7087$ y $SS_{Res} = 252.9039$. Nótese que hay 10 niveles distintos de x, con puntos duplicados en $x = 1.0$, $x = 3.3$, $x = 4.0$, $x = 5.6$ y $x = 6.0$. La suma de cuadrados de error puro se calcula usando los puntos duplicados como sigue:

Nivel de x	$\sum_j (y_{ij} - \bar{y}_i)^2$	Grados de libertad
1.0	1.1858	1
3.3	1.0805	1
4.0	11.2467	2
5.6	1.4341	2
6.0	0.6161	1
Total	15.5632	7

La suma de cuadrados de falta de ajuste se calcula restando, como sigue:

$$\begin{aligned} SS_{FA} &= SS_{Res} - SS_{EP} \\ &= 252.9039 - 15.5632 = 237.3407, \end{aligned}$$

con $m - 2 = 10 - 2 = 8$ grados de libertad. El análisis de varianza que incorpora la prueba de falta de ajuste se ve en la tabla 4.2. El estadístico de prueba de falta de ajuste es $F_0 = 13.34$,

http://carlos2524.jimdo.com/

y como el valor P es muy pequeño, se rechaza la hipótesis de que el modelo tentativo describe a los datos en forma adecuada.

4.5.2 Estimación del error puro a partir de vecinos próximos

En la sección 4.5.1 se describió una prueba de falta de ajuste para el modelo de regresión lineal. El procedimiento implicaba agrupar la suma de cuadrados de error, o de residuales, en un componente debido al error "puro" y otro debido a la falta de ajuste:

$$SS_{\text{Res}} = SS_{\text{EP}} + SS_{\text{FA}}$$

La suma de cuadrados de error puro, SS_{EP}, se calcula con respuestas en observaciones *replicadas* o repetidas en el mismo nivel de x. Es un **estimado de σ^2 independiente del modelo.**

Este procedimiento general se puede, en principio, aplicar a cualquier modelo de regresión. En el cálculo de SS_{EP} se requieren observaciones repetidas de la respuesta y al mismo conjunto de niveles de las variables regresoras $x_1, x_2, \ldots, x_k$, lo que significa que algunos de los **renglones** de la matriz $\mathbf{X}$ deben ser iguales. Sin embargo, en regresión múltiple no se presentan con frecuencia observaciones repetidas, y no es útil el procedimiento descrito en la sección 4.5.1.

Daniel y Wood [1980], y Joglekar, Schuenemeyer y La Riccia [1989] han investigado métodos para obtener un estimado del error independiente del modelo cuando no hay puntos repetidos exactos. En estos procedimientos se buscan puntos en el espacio de x que sean **vecinos cercanos**, esto es, conjuntos de observaciones que se hayan tomado con niveles casi idénticos de $x_1, x_2, \ldots, x_k$. Se considera que las respuestas y_i a partir de esos vecinos cercanos son puntos de repetición, y se usan para obtener un estimado del error puro. Como medida de la distancia entre dos puntos cualesquiera, por ejemplo, entre $x_{i1}, x_{i2}, \ldots, x_{ik}$ y $x_{i'1}, x_{i'2}, \ldots, x_{i'k}$ se usará la suma ponderada de la distancia al cuadrado (WSSD, de *weighted sum of squared distance*):

$$D_{ii'}^2 = \sum_{j=1}^{k} \left[\frac{\hat{\beta}_j (x_{ij} - x_{i'j})}{\sqrt{MS_{\text{Res}}}} \right]^2 \tag{4.27}$$

Los pares de datos que tienen valores pequeños de D_{ii}^2 son "vecinos cercanos", por lo que, están relativamente cerca entre sí en el espacio de x. Los pares de puntos para los que D_{ii}^2 es grande (es decir, $D_{ii}^2 \gg 1$), están muy separados en el espacio de x. Los residuales en dos puntos con un valor pequeño de D_{ii}^2 se pueden usar para obtener un estimado del error puro. El estimado se obtiene a partir del recorrido de los residuales en los puntos i e i', por ejemplo

$$E_i = |e_i - e_{i'}|$$

Hay una relación entre el rango de una muestra de una población normal, y la desviación estándar de la población. Para muestras de tamaño 2, esta relación es

$$\hat{\sigma} = (1.128)^{-1} E = 0.886\, E$$

La cantidad $\hat{\sigma}$ así obtenida es un estimado de la desviación estándar del error puro.

http://carlos2524.jimdo.com/

TABLA 4.3 Cálculo de $D_{ii'}^2$, para el ejemplo 4.9

			Cálculos de vecinos cercanos											
			Delta entre residuales y distancias normalizadas ponderadas al cuadrado entre vecinos cercanos											
			Adyacente			Separados por 1			Separados por 2			Separados por 3		
Observación	y ordenada ajustada	Residual	Delta	$D_{ii'}^2$	R[a]	Delta	$D_{ii'}^2$	R	Delta	$D_{ii'}^2$	R	Delta	$D_{ii'}^2$	R
7	7.155	.845	.949	.3524E + 00		4.080	.1001E + 01		.302	.4814E + 00		1.057	.1014E + 01	
19	7.707	1.793	3.131	.2835E + 00	12	.647	.6594E + 00		2.006	.4989E + 00		1.843	.1800E + 01	
4	9.956	4.924	3.778	.6275E + 00		5.137	.9544E − 01	3	4.974	.1562E + 01		3.897	.5965E + 00	
2	10.354	1.146	1.359	.3412E + 00	15	1.196	.2805E + 00	11	.119	.2696E + 00	9	1.591	.2307E + 01	
25	10.963	− .213	.163	.9489E + 00		1.240	.2147E + 00	6	.232	.9831E + 00		.649	.1032E + 01	
3	12.080	− .050	1.077	.3865E + 00		.395	.2915E + 01		.486	.2594E + 01		3.498	.4775E + 01	
13	12.473	1.027	1.471	.1198E + 01		.591	.1042E + 01		2.422	.2507E + 01		.130	.2251E + 01	
5	14.194	− .444	.881	.4869E − 01	2	3.893	.2521E + 00	8	1.601	.3159E + 00	13	.155	.8768E + 00	
17	14.914	.436	3.012	.3358E + 00	14	.720	.2477E + 00	7	.726	.5749E + 00		.631	.1337E + 01	
18	15.551	3.449	2.292	.1185E + 00	5	3.738	.7636E + 00		2.381	.2367E + 01		1.072	.5341E + 01	
8	16.673	1.157	1.446	.2805E + 00	10	.089	.1483E + 01		1.220	.4022E + 01		3.771	.2307E + 01	
6	18.400	− .290	1.357	.5851E + 00		2.666	.2456E + 01		2.325	.2915E + 01		.303	.2470E + 01	
14	18.682	1.068	1.309	.6441E + 00		3.682	.5952E + 01		1.661	.5121E + 01		6.096	.4328E + 00	
10	19.124	2.376	4.991	.1036E + 02		2.969	.9107E + 01		7.404	.1023E + 01		1.705	.4412E + 01	
21	20.514	− 2.614	2.021	.1096E + 00	4	2.414	.5648E + 01		3.285	.2093E + 01		1.993	.2117E + 01	
12	21.593	− .593	4.435	.4530E + 01		1.264	.1303E + 01		4.015	.1321E + 01		3.980	.4419E + 01	
1	21.708	− 5.028	5.699	.1227E + 01		.421	.1219E + 01		.455	.3553E + 00		4.365	.3121E + 01	
15	23.329	.671	5.279	.7791E − 04	1	5.244	.9269E + 00		1.334	.2341E + 01		1.566	.1316E + 02	
23	23.358	− 4.608	.035	.9124E + 00		3.945	.2316E + 01		6.845	.1315E + 02		1.180	.1772E + 02	
24	24.403	− 4.573	3.910	.1370E + 01		6.810	.1578E + 02		1.215	.2026E + 02		.886	.8023E + 02	
16	29.663	− .663	2.900	.8999E + 01		5.125	.1204E + 02		3.024	.6294E + 02		8.083	.1074E + 03	
11	38.093	2.237	8.025	.3767E + 00		5.924	.2487E + 02		5.182	.5978E + 02				
20	40.888	− 5.788	2.101	.1994E + 02		13.208	.5081E + 02							
22	56.007	− 3.687	11.106	.1216E + 02										
9	71.820	7.420												

[a] La columna R muestra el lugar ordinal de los 15 valores menores de $D_{ii'}^2$.

http://carlos2524.jimdo.com/

TABLA 4.4 Cálculo de $\hat{\sigma}$ para el ejemplo 4.9

	Desviación estándar estimada a partir de residuales de observaciones vecinas				
		Ordenadas por $D_{ii'}^2$			
Número	Desviación estándar acumulada	$D_{ii'}^2$	Observación	Observación	Delta entre residuales
1	.4677E + 01	.7791E −04	15	23	5.2788
2	.2729E + 01	.4859E −01	5	17	0.8807
3	.3336E + 01	.9544E −01	4	25	5.1369
4	.2950E + 01	.1096E +00	21	12	2.0211
5	.2766E + 01	.1185E +00	18	8	2.2920
6	.2488E + 01	.2147E +00	25	13	1.2396
7	.2224E + 01	.2477E +00	17	8	0.7203
8	.2377E + 01	.2521E +00	5	18	3.8930
9	.2125E + 01	.2696E +00	2	13	0.1194
10	.2040E + 01	.2805E +00	8	6	1.4462
11	.1951E + 01	.2805E +00	2	3	1.1962
12	.2020E + 01	.2835E +00	19	4	3.1312
13	.1973E + 01	.3159E +00	5	8	1.6010
14	.2023E + 01	.3358E +00	17	18	3.0123
15	.1969E + 01	.3412E +00	2	25	1.3590
16	.1898E + 01	.3524E +00	7	19	0.9486
17	.1810E + 01	.3553E +00	1	24	0.4552
18	.2105E + 01	.3767E +00	11	20	8.0255
19	.2044E + 01	.3865E +00	3	13	1.0768
20	.2212E + 01	.4328E +00	14	1	6.0956
21	.2119E + 01	.4814E +00	7	2	0.3018
22	.2104E + 01	.4989E +00	19	25	2.0058
23	.2040E + 01	.5749E +00	17	6	0.7259
24	.2005E + 01	.5851E +00	6	14	1.3571
25	.2063E + 01	.5965E +00	4	13	3.8973
26	.2113E + 01	.6275E +00	4	2	3.7780
27	.2077E + 01	.6441E +00	14	10	1.3089
28	.2024E + 01	.6594E +00	19	2	0.6468
29	.2068E + 01	.7636E +00	18	6	3.7382
30	.2004E + 01	.8768E +00	5	6	0.1548
31	.1940E + 01	.9124E +00	23	24	0.0347
32	.2025E + 01	.9269E +00	15	24	5.2441
33	.1968E + 01	.9489E +00	25	3	0.1628
34	.1916E + 01	.9831E +00	25	5	0.2318
35	.1964E + 01	.1001E +01	7	4	4.0797
36	.1936E + 01	.1014E +01	7	25	1.0572
37	.2061E + 01	.1023E +01	10	1	7.4045
38	.2022E + 01	.1032E +01	25	17	0.6489
39	.1983E + 01	.1042E +01	13	17	0.5907
40	.1966E + 01	.1198E +01	13	5	1.4714

http://carlos2524.jimdo.com/

Para calcular este estimado se puede usar un algoritmo eficiente. En Montgomery, Martin y Peck [1980] se presenta un programa de cómputo para este algoritmo. Primero se ordenan los datos $x_{i1}, x_{i2}, \ldots, x_{ik}$ en orden creciente de $\hat{y}_i$. Nótese que los puntos con valores muy distintos de $\hat{y}_i$ no pueden ser vecinos cercanos. Pero los que tienen valores parecidos de $\hat{y}_i$ podrían ser vecinos cercanos (o podrían estar cerca de la misma curva de nivel de $\hat{y}$ constante, pero estar muy apartados en algunas coordenadas x). Entonces:

1. Se calculan los valores de D_{ii}^2 para todos los $n - 1$ pares de puntos con valores adyacentes de $\hat{y}$. Se repite este cálculo para pares de puntos separados por uno, dos y tres valores intermedios de $\hat{y}$, esto producirá $4n - 10$ valores de D_{ii}^2.
2. Se promedian los $4n - 10$ valores de D_{ii}^2 calculados en el paso 1, en orden creciente. Sea E_u, $u = 1, 2, \ldots, 4n - 10$ el recorrido de los residuales en esos puntos.
3. Para los primeros m valores de E_u se calcula un estimado de la desviación estándar del error puro, como sigue:

$$\hat{\sigma} = \frac{0.886}{m} \sum_{u=1}^{m} E_u \tag{4.28}$$

Nótese que $\hat{\sigma}$ se basa en el recorrido promedio de los residuales asociados con los m valores más pequeños de D_{ii}^2; m se debe escoger después de inspeccionar los valores de D_{ii}^2. No se deben incluir en el cálculo los valores de E_u para los que la suma ponderada de la distancia al cuadrado sea demasiado grande.

Ejemplo 4.9 *Datos del tiempo de entrega*

Se usará el procedimiento que se acaba de describir para calcular un estimado de la desviación estándar del error puro para los datos de entrega de gaseosas, del ejemplo 3.1. La tabla 4.3 muestra el cálculo de D_{ii}^2 para pares de puntos que, en términos de $\hat{y}$, son adyacentes, separados por un valor, por dos valores y por tres. Las R columnas de esta tabla identifican los 15 valores menores de D_{ii}^2. Con los residuales de esos 15 pares de puntos se estima σ. Esos cálculos dan como resultado $\hat{\sigma} = 1.969$, y se resumen en la tabla 4.4. Según la tabla 3.5, se ve que $\sqrt{MS_{\text{Res}}} = \sqrt{10.6239} = 3.259$. Ahora, si no hay falta de ajuste apreciable, cabría esperar que $\hat{\sigma} = \sqrt{MS_{\text{Res}}}$. En este caso $\sqrt{MS_{\text{Res}}}$ es más o menos 65% mayor que $\hat{\sigma}$, lo cual indica cierta falta de ajuste. Eso se podría deber a los efectos de regresores que no estén en el modelo, o a la presencia de uno o más datos atípicos.

PROBLEMAS

4.1 Se tiene el modelo de ajuste de regresión simple para el desempeño de los equipos de la Liga Nacional de Futbol, en el problema 2.1.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la hipótesis de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.

c. Graficar los residuales en función de las yardas por aire del equipo, x_2. ¿Indica esta gráfica que mejorará el modelo agregándole x_2?

http://carlos2524.jimdo.com/

4.2 Para el modelo de ajuste de regresión múltiple, de los datos de desempeño de los equipos de la Liga Nacional de Futbol en el problema 3.1:

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.

c. Trazar gráficas de los residuales en función de cada una de las variables regresoras. ¿Implican esas gráficas que se especificó en forma correcta el regresor?

d. Trazar las gráficas de regresión parcial para este modelo. Compararlas con las gráficas de residuales en función de los regresores obtenidas en la parte c anterior. Describir la clase de información que proporcionan esas gráficas.

e. Calcular los residuales estudentizados y los residuales R de Student para este modelo. ¿Qué información aportan esos residuales escalados?

4.3 Se tiene el modelo de ajuste con regresión lineal simple para los datos de energía solar en el problema 2.3.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.

4.4 Para el ajuste con regresión múltiple de los datos de rendimiento de gasolina en el problema 3.5.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta pronosticada.

c. Trazar e interpretar las gráficas de regresión parcial para este modelo.

d. Calcular los residuales estudentizados y los residuales R de Student para este modelo. ¿Qué información aportan esos residuales escalados?

4.5 Se tiene el modelo de ajuste por regresión múltiple de los datos de precios de casas en el problema 3.7.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de residuales en función de la respuesta pronosticada.

c. Trazar las gráficas de regresión parcial para este modelo. ¿Parece que no son necesarias algunas de las variables que ya están en este modelo?

d. Calcular los residuales estudentizados y los residuales R de Student para este modelo. ¿Qué información aportan esos residuales escalados?

4.6 Para el modelo de ajuste por regresión lineal simple para los datos de pureza de oxígeno en el problema 2.7.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.

4.7 Se tiene el modelo de ajuste con regresión lineal simple para los datos de peso y presión sanguínea en el problema 2.10.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.

c. Suponer que se reunieron los datos en el orden que muestra la tabla. Graficar los residuales en función del orden temporal, y comentar la gráfica.

http://carlos2524.jimdo.com/

4.8 Se tiene el modelo de ajuste por regresión lineal simple para los datos de la planta de vapor en el problema 2.12.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta pronosticada.

c. Suponer que los datos se reunieron en el orden que muestra la tabla. Graficar los residuales en función del orden temporal y comentar la gráfica.

4.9 Se tiene el modelo de ajuste por regresión lineal simple para los datos de ozono en el problema 2.13.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.

c. Graficar los residuales en función de su orden en el tiempo y comentar la gráfica.

4.10 Para el modelo de ajuste por regresión lineal simple, de los datos de viscosidad de copoliéster en el problema 2.14.

a. Trazar una gráfica de probabilidad normal de los residuales no escalados. ¿Parece haber algún problema con la suposición de normalidad?

b. Repetir la parte a, usando los residuales estudentizados. ¿Hay alguna diferencia apreciable entre las dos gráficas?

c. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.

4.11 Se tiene el modelo de ajuste por regresión lineal simple para los datos de viscosidad de mezclas de tolueno y tetralina, en el problema 2.15.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta pronosticada.

4.12 Considérese el modelo de ajuste por regresión lineal simple para los datos de presión y volumen de tanque, en el problema 2.16.

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?

b. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.

c. Suponer que los datos se tomaron en el orden que muestra la tabla. Graficar los residuales en función de su orden en el tiempo, y comentar la gráfica.

4.13 En el problema 3.8 se pidió ajustar dos modelos distintos a los datos del proceso químico de la tabla B.5 del apéndice. Hacer los análisis de residuales adecuados para ambos modelos. Comentar los resultados de esos análisis. Calcular la estadística PRESS para ambos modelos. ¿Las gráficas de residuales y de PRESS proporcionan algún indicio acerca de la mejor opción de modelo para los datos?

4.14 En los problemas 2.4 y 3.5 se pidió ajustar dos modelos distintos a los datos de rendimiento de gasolina de la tabla B.3 del apéndice. Calcular la estadística PRESS para esos dos modelos. Con base en esa estadística, ¿cuál modelo producirá mejores pronósticos con datos nuevos?

4.15 En el problema 3.9 se pidió ajustar un modelo para los datos del reactor de flujo tubular en la tabla B.6 del apéndice.

http://carlos2524.jimdo.com/

a. Trazar una gráfica de probabilidad normal de los residuales. ¿Parece haber algún problema con la suposición de normalidad?
b. Trazar e interpretar una gráfica de los residuales en función de la respuesta pronosticada.
c. Trazar las gráficas de regresión parcial para este modelo. ¿Parece que no son necesarias algunas de las variables que incluye el modelo?

4.16 En el problema 3.12 se pidió ajustar un modelo para los datos de formación de catrato de la tabla B.8 del apéndice.
a. Trazar una gráfica de probabilidad normal de los residuales del modelo completo. ¿Parece haber algún problema con la suposición de normalidad?
b. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.
c. En ese problema (3.12) se pidió ajustar un segundo modelo. Calcular la estadística PRESS para ambos modelos. Con base en esa estadística, ¿cuál modelo producirá, probablemente, mejores predicciones con nuevos datos?

4.17 En el problema 3.14 se pidió ajustar un modelo para los datos de viscosidad cinemática en la tabla B.10 del apéndice.
a. Trazar una gráfica de probabilidad normal de los residuales para el modelo completo.
b. Trazar e interpretar una gráfica de los residuales en función de la respuesta predicha.
c. En ese problema (3.14) se pidió ajustar un segundo modelo. Calcular la estadística PRESS para ambos modelos. Con base en esta estadística, ¿qué modelo producirá, probablemente, mejores predicciones?

4.18 Coteron, Sánchez, Martínez y Aracil ("Optimization of the Synthesis of an Analogue of Jojoba Oil Using a Fully Central Composite Design", *Canadian Journal of Chemical Engineering,* 1993) estudiaron la relación entre la temperatura de reacción, x_1, la cantidad inicial de catalizador, x_2 y la presión, x_3, sobre el rendimiento de un análogo sintético del aceite de jojoba. La tabla siguiente resume los resultados experimentales.

x_1	x_2	x_3	y
−1	−1	−1	17
1	−1	−1	44
−1	1	−1	19
1	1	−1	46
−1	−1	1	7
1	−1	1	55
−1	1	1	15
1	1	1	41
0	0	0	29
0	0	0	28.5
0	0	0	30
0	0	0	27
0	0	0	28

a. Hacer un análisis minucioso de los resultados, que incluya gráficas de residuales.
b. Hacer la prueba adecuada de falta de ajuste.

http://carlos2524.jimdo.com/

4.19 Derringer y Suich ("Simultaneous Optimization of Several Response Variables", *Journal of Quality Technology*, 1980) estudiaron la relación de un índice de abrasión para una lona de neumático, en función de tres factores: x_1, concentración de sílice hidratada; x_2, concentración de agente acoplante de silano, y x_3, concentración de azufre. La siguiente tabla muestra los resultados obtenidos.

x_1	x_2	x_3	y
−1	−1	1	102
1	−1	−1	120
−1	1	−1	117
1	1	1	198
−1	−1	−1	103
1	−1	1	132
−1	1	1	132
1	1	−1	139
0	0	0	133
0	0	0	133
0	0	0	140
0	0	0	142
0	0	0	145
0	0	0	142

a. Hacer un análisis detallado de los resultados, que incluya gráficas de residuales.
b. Hacer la prueba adecuada de falta de ajuste.

4.20 Myers y Montgomery (*Response Surface Methodology*, 1955, pp. 267-268) describen un experimento para determinar la influencia de

x_1 – temperatura del baño ácido

x_2 – concentración del ácido de cascada

x_3 – temperatura del agua

x_4 – concentración de sulfuro

x_5 – cantidad de blanqueador de cloro

sobre una medida adecuada de la blancura del rayón (y). Los ingenieros que hicieron este experimento desean minimizar la medida. Los resultados experimentales se muestran a continuación.

http://carlos2524.jimdo.com/

Temperatura del ácido	Concentración del ácido	Temperatura del agua	Concentración de sulfuro	Cantidad de blanqueador	y
35	0.3	82	0.2	0.3	76.5
35	0.3	82	0.3	0.5	76.0
35	0.3	88	0.2	0.5	79.9
35	0.3	88	0.3	0.3	83.5
35	0.7	82	0.2	0.5	89.5
35	0.7	82	0.3	0.3	84.2
35	0.7	88	0.2	0.3	85.7
35	0.7	88	0.3	0.5	99.5
55	0.3	82	0.2	0.5	89.4
55	0.3	82	0.3	0.3	97.5
55	0.3	88	0.2	0.3	103.2
55	0.3	88	0.3	0.5	108.7
55	0.7	82	0.2	0.3	115.2
55	0.7	82	0.3	0.5	111.5
55	0.7	88	0.2	0.5	102.3
55	0.7	88	0.3	0.3	108.1
25	0.5	85	0.25	0.4	80.2
65	0.5	85	0.25	0.4	89.1
45	0.1	85	0.25	0.4	77.2
45	0.9	85	0.25	0.4	85.1
45	0.5	79	0.25	0.4	71.5
45	0.5	91	0.25	0.4	84.5
45	0.5	85	0.15	0.4	77.5
45	0.5	85	0.35	0.4	79.2
45	0.5	85	0.25	0.2	71.0
45	0.5	85	0.25	0.6	90.2

a. Hacer un análisis detallado de los resultados, que incluya gráficas de residuales.
b. Hacer la prueba adecuada de falta de ajuste.

4.21 Para la prueba de falta de ajuste, calcular $E(MS_{EP})$ y $E(MS_{FA})$.

4.22 La tabla B.14 del apéndice contiene datos sobre los puntos transitorios de un inversor electrónico. Usar sólo los regresores x_1 a x_4 para ajustarles un modelo de regresión múltiple.
a. Investigar la adecuación del modelo.
b. Suponer que la observación 2 se anotó mal. Eliminar esta observación, volver a ajustar el modelo y hacer un análisis detallado de residuales. Comentar la diferencia entre los resultados obtenidos.

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

CAPÍTULO 5

TRANSFORMACIONES Y PONDERACIÓN PARA CORREGIR INADECUACIONES DEL MODELO

5.1 INTRODUCCIÓN

El capítulo 4 presentó algunas técnicas para comprobar la adecuación del modelo de regresión lineal. Recuérdese que el ajuste de modelos de regresión tiene varias suposiciones implícitas, que incluyen las siguientes:

1. Los errores del modelo tienen media cero y varianza constante, y no están correlacionados.
2. Los errores del modelo tienen distribución normal; esta premisa es para poder hacer pruebas de hipótesis, y determinar intervalos de confianza. Bajo esta hipótesis, los errores son independientes.
3. La forma del modelo, incluyendo la especificación de los regresores, es correcta.

Las gráficas de residuales son poderosos métodos para detectar violaciones de esas premisas básicas de la regresión. Esta forma de comprobar la adecuación del modelo se debería hacer para cada modelo de regresión que se esté considerando usar en la práctica.

En este capítulo se detallarán los métodos y procedimientos para construir modelos de regresión cuando se violan algunas de las suposiciones anteriores. Se dará mucho énfasis a la **transformación de datos**. No es raro ver que cuando las variables de respuesta y/o regresoras se expresan en la escala o métrica correcta de regresión, ya no se presentan ciertas violaciones de las premisas, como la no constancia de la varianza. En el caso ideal, la elección de la métrica la debe hacer el ingeniero o el científico que tenga **conocimientos sobre la materia**, pero hay muchos casos en los que esta información no está disponible. En esos casos, se puede escoger una transformación de datos en forma heurística (*por tanteos*) o con algún otro procedimiento analítico.

El método de cuadrados mínimos ponderados también es útil para construir modelos de regresión en los casos en que se violan algunas de las premisas básicas. Se ilustrará cómo se aplican los mínimos cuadrados cuando no es adecuada la suposición de varianza constante. Esta técnica también demostrará ser esencial en los capítulos siguientes, cuando se consideren otros métodos para manejar variables de respuesta no normales.

http://carlos2524.jimdo.com/

5.2 TRANSFORMACIONES ESTABILIZADORAS DE LA VARIANZA

La suposición de **varianza constante** es un requisito básico del análisis de regresión. Una razón frecuente de violación de esta premisa es que la variable de respuesta y sigue una distribución de probabilidades en la que la varianza se relaciona en forma funcional con la media. Por ejemplo, si y es una variable aleatoria de Poisson en un modelo de regresión lineal simple, la varianza de y es igual a la media. Como la media de y está relacionada con la variable regresora x, la varianza de y será proporcional a x. Con frecuencia, en estos casos son útiles las transformaciones estabilizadoras de varianza. Así, si la distribución de y es de Poisson, se podría hacer la regresión $y' = \sqrt{y}$ en función de x, porque la varianza de la raíz cuadrada de una variable aleatoria de Poisson es independiente de la media. Para citar otro ejemplo, si la variable de respuesta es una proporción ($0 \leq y_i \leq 1$), y la gráfica de los residuales en función de $\hat{y}_i$ tiene el patrón de doble arco como en la figura 4.3c, entonces sería adecuada la transformación arco seno: $y' = \text{sen}^{-1}(\sqrt{y})$.

En la tabla 5.1 se resumen algunas transformaciones estabilizadoras de varianza de uso frecuente. La **potencia** de una transformación depende de la cantidad de curvatura que induce. Las transformaciones de la tabla 5.1 van desde la raíz cuadrada, relativamente benigna, hasta el recíproco, relativamente fuerte. Hablando en general, una transformación benigna aplicada dentro de un margen de valores relativamente angosto (por ejemplo, $y_{\text{máx}}/y_{\text{mín}}) < 2$ a 3), tiene poco efecto. Por otra parte, una transformación fuerte dentro de un gran intervalo de valores tendrá un efecto dramático sobre el análisis.

A veces se puede recurrir a la experiencia o a consideraciones teóricas como guía para seleccionar una transformación adecuada. Sin embargo, en muchos casos no se tendrá un conocimiento *a priori* para creer que la varianza del error no sea constante. La primera indicación del problema se tiene al inspeccionar los diagramas de dispersión o con el análisis de residuales. En esos casos, la transformación adecuada se debe seleccionar **empíricamente**.

Es importante descubrir y corregir una varianza no constante de error. Si no se elimina este problema, los estimadores de mínimos cuadrados seguirán siendo insesgados, pero ya no tendrán la propiedad de la varianza mínima. Esto quiere decir que los coeficientes de regresión tendrán errores estándar mayores que lo necesario. El efecto de la transformación suele proporcionar estimados más precisos de los parámetros del modelo y mayor sensibilidad para las pruebas estadísticas.

Cuando la variable de respuesta se ha reexpresado, las variables predichas están en la escala transformada. Con frecuencia se hace necesario reconvertir los valores predichos a las unidades originales. Desafortunadamente, si se aplica la transformación inversa en

TABLA 5.1 Transformaciones útiles estabilizadoras de varianza

Relación entre σ^2 y $E(y)$	Transformación
$\sigma^2 \propto$ constante	$y' = y$ (sin transformación)
$\sigma^2 \propto E(y)$	$y' = \sqrt{y}$ (raíz cuadrada; datos de Poisson)
$\sigma^2 \propto E(y)[1 - E(y)]$	$y' = \sin^{-1}(\sqrt{y})$ (arco seno; proporciones binomiales $0 \leq y_i \leq 1$)
$\sigma^2 \propto [E(y)]^2$	$y' = \ln(y)$ (logarítmica)
$\sigma^2 \propto [E(y)]^3$	$y' = y^{-1/2}$ (raíz cuadrada recíproca)
$\sigma^2 \propto [E(y)]^4$	$y' = y^{-1}$ (recíproca)

http://carlos2524.jimdo.com/

forma directa a los valores predichos se obtiene un estimado de la mediana de la distribución de la respuesta, y no de la media. Con frecuencia es posible adaptar un método para obtener predicciones insesgadas en las unidades originales. Los procedimientos para producir estimados insesgados puntuales para varias transformaciones estándar aparecen en Neyman y Scott [1960]. También Miller [1984] sugiere algunas soluciones sencillas de este problema. Los intervalos de confianza o de predicción se pueden convertir en forma directa de una a otra métrica, porque esos estimados de intervalo son percentiles de una distribución, y los percentiles no se afectan por la transformación. Sin embargo, no hay seguridad de que los intervalos resultantes en las unidades originales sean los más estrechos posibles. Para conocer más detalles, véase Land [1974].

Ejemplo 5.1 *Datos de la compañía eléctrica*

Una empresa eléctrica se interesa en desarrollar un modelo que relacione la demanda horaria pico (y) al uso total de energía durante el mes (x). Es un problema importante de planeación, porque la mayor parte de los clientes pagan en forma directa por el uso de la energía (en kilowatt horas), el sistema de generación debe ser lo bastante grande como para cumplir con la demanda máxima impuesta. En la tabla 5.2 se ven datos de 53 clientes

TABLA 5.2 Datos de demanda (y) y consumo de energía (x) para 53 clientes residenciales, agosto de 1979

Cliente	x (KWH)	y (KW)	Cliente	x (KWH)	y (KW)
1	679	0.79	27	837	4.20
2	292	0.44	28	1748	4.88
3	1012	0.56	29	1381	3.48
4	493	0.79	30	1428	7.58
5	582	2.70	31	1255	2.63
6	1156	3.64	32	1777	4.99
7	997	4.73	33	370	0.59
8	2189	9.50	34	2316	8.19
9	1097	5.34	35	1130	4.79
10	2078	6.85	36	463	0.51
11	1818	5.84	37	770	1.74
12	1700	5.21	38	724	4.10
13	747	3.25	39	808	3.94
14	2030	4.43	40	790	0.96
15	1643	3.16	41	783	3.29
16	414	0.50	42	406	0.44
17	354	0.17	43	1242	3.24
18	1276	1.88	44	658	2.14
19	745	0.77	45	1746	5.71
20	435	1.39	46	468	0.64
21	540	0.56	47	1114	1.90
22	874	1.56	48	413	0.51
23	1543	5.28	49	1787	8.33
24	1029	0.64	50	3560	14.94
25	710	4.00	51	1495	5.11
26	1434	0.31	52	2221	3.85
			53	1526	3.93

http://carlos2524.jimdo.com/

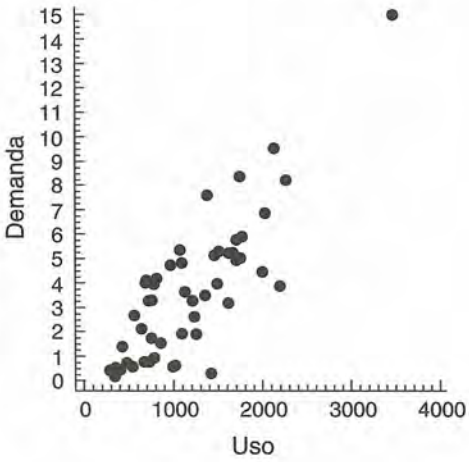


Figura 5.1 Diagrama de dispersión de la demanda de potencia (kW) en función del uso de la energía (kWh), ejemplo 5.1.

residenciales para el mes de agosto de 1979, y en la figura 5.1 se ve un diagrama de dispersión. Como punto de partida se supone un modelo de regresión simple, y el ajuste por mínimos cuadrados es

$$\hat{y} = -0.8313 + 0.00368x$$

El análisis de varianza se ve en la tabla 5.3. Para este modelo, $R^2 = 0.7046$; esto es, un 70% de la variabilidad de la demanda se explica con un ajuste rectilíneo al uso de la energía. Los estadísticos de resumen no indican que hay algún problema obvio en este modelo.

En la figura 5.2 se ve una gráfica de los residuales R de Student en función de los valores ajustados $\hat{y}_i$. Estos residuales forman un embudo que se abre hacia afuera, lo que indica que la varianza de error aumenta a medida que aumenta el consumo de energía. Podría ser útil tener una transformación para corregir esta inadecuación de modelo. Para seleccionar la forma de la transformación obsérvese que la variable y de respuesta se puede considerar como un "conteo" de la cantidad de kilowatts usados por un cliente durante determinada hora. El modelo probabilista más sencillo para datos de conteo es la distribución de Poisson. Esto sugiere usar $y^* = \sqrt{y}$ como transformación estabilizadora de varianza al hacer la regresión respecto a x. El ajuste resultante por mínimos cuadrados es

$$\hat{y}^* = 0.5822 + 0.0009529x$$

Los valores de R de Student para este ajuste por mínimos cuadrados se grafican en función de $\hat{y}_i^*$ en la figura 5.3. La impresión, al examinar esta gráfica, es que la varianza es estable. En consecuencia, se llega a la conclusión de que el modelo transformado es adecuado.

TABLA 5.3 Análisis de varianza para la regresión de y con respecto a x para el ejemplo 5.1

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Regresión	303.6331	1	302.6331	121.66	< 0.0001
Residuales	126.8660	51	2.4876		
Total	429.4991	52			

http://carlos2524.jimdo.com/

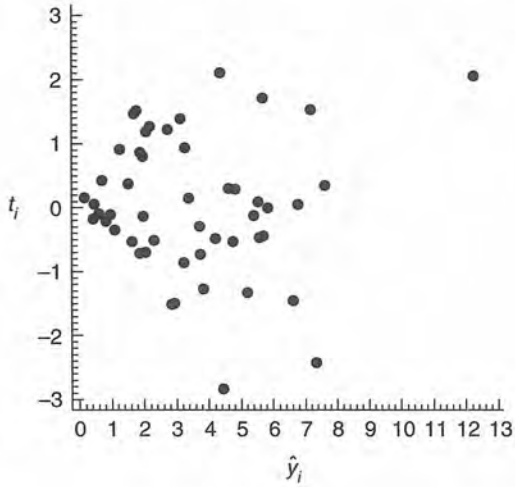


Figura 5.2 Gráfica de valores t_i de R de Student en función de los valores ajustados $\hat{y}_i$, ejemplo 5.1.

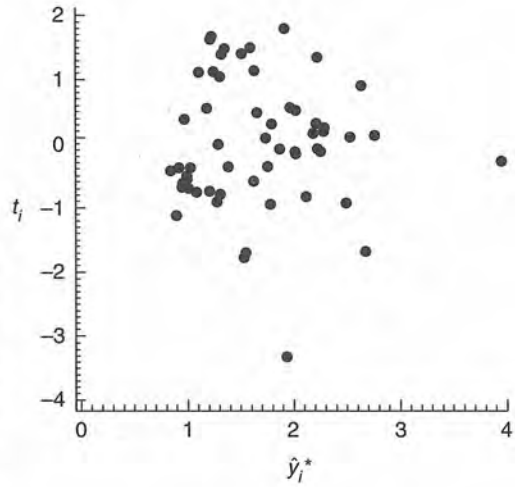


Figura 5.3 Gráfica de valores t_i de R de Student en función de los valores ajustados y_i^* para los datos transformados, ejemplo 5.1.

Nótese que hay un residual sospechosamente grande (cliente 26) y un cliente cuyo uso de energía es algo grande (cliente 50). Se debería estudiar más el efecto de esos dos puntos sobre el ajuste, antes de liberar el modelo para su empleo.

5.3 TRANSFORMACIONES PARA LINEALIZAR EL MODELO

La premisa de una relación lineal entre y y los regresores es el punto acostumbrado de partida en el análisis de regresión. A veces se ve que esta hipótesis es inadecuada. Se puede descubrir la no linealidad a través de la prueba de falta de ajuste que se describió en la sección 4.5, o con diagramas de dispersión, la matriz de diagramas de dispersión o con gráficas de residuales, como la gráfica de regresión parcial. A veces son la experiencia o las consideraciones teóricas las que pueden indicar que la relación entre y y los regresores no es lineal. En algunos casos, una función no lineal se puede linealizar con una transformación adecuada. A esos modelos no lineales se les llama **intrínsecamente** o **transformablemente lineales**.

http://carlos2524.jimdo.com/

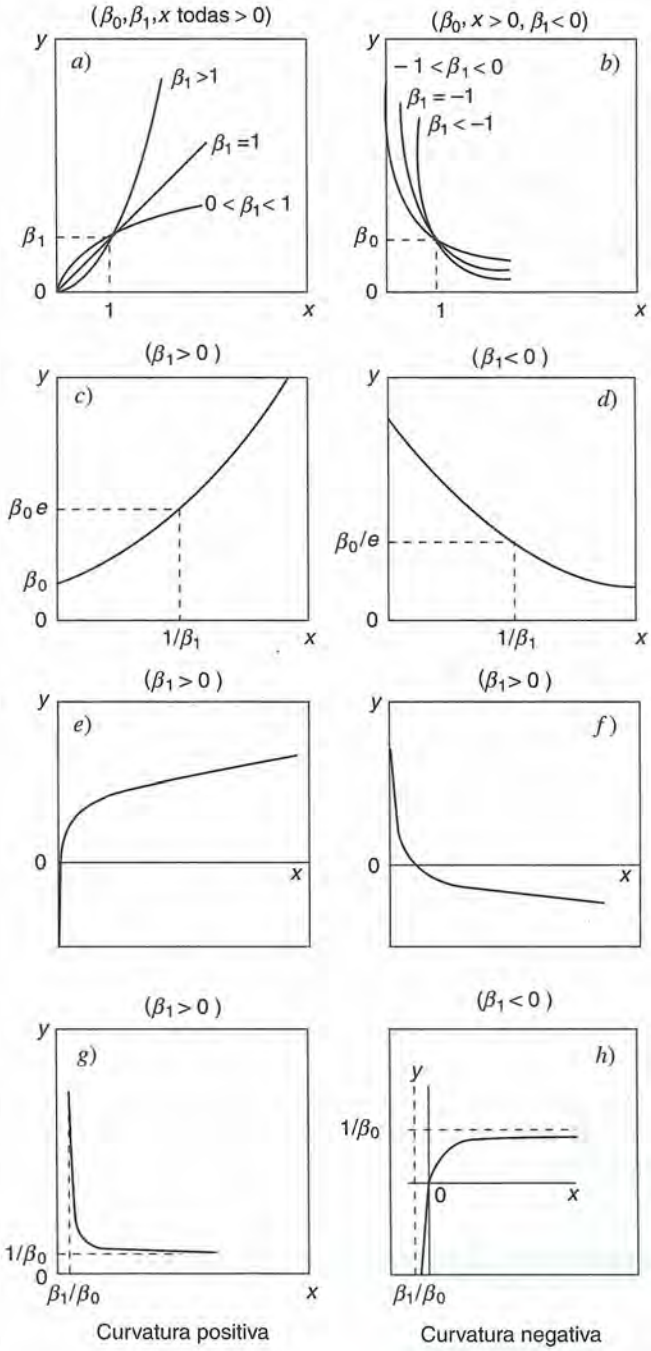


Figura 5.4 Funciones linealizables. (Tomado de Daniel y Wood [1980]; se usa con autorización de la editorial.)

Algunas funciones linealizables se ven en la figura 5.4. Las funciones no lineales, las transformaciones y las formas lineales que resultan se ven en la tabla 5.4. Cuando el diagrama de dispersión de y en función de x indica que hay curvatura, se podrá amoldar el comportamiento observado en la gráfica con alguna de las curvas de la figura 5.4 y usar la forma linealizada de la función, para representar los datos.

http://carlos2524.jimdo.com/

TABLA 5.4 Funciones linealizables y su forma lineal correspondiente

Figura	Función linealizable	Transformación	Forma lineal
5.4a, b	$y = \beta_0 x^{\beta_1}$	$y' = \log y,\ x' = \log x$	$y' = \log \beta_0 + \beta_1 x'$
5.4c, d	$y = \beta_0 e^{\beta_1 x}$	$y' = \ln y,$	$y' = \ln \beta_0 + \beta_1 x$
5.4e, f	$y = \beta_0 + \beta_1 \log x$	$x' = \log x$	$y' = \beta_0 + \beta_1 x'$
5.4g, h	$y = \dfrac{x}{\beta_0 x - \beta_1}$	$y' = \dfrac{1}{y},\ x' = \dfrac{1}{x}$	$y' = \beta_0 - \beta_1 x'$

Para ilustrar un modelo no lineal que es intrínsecamente lineal, se tiene la función exponencial

$$y = \beta_0 e^{\beta_1 x} \varepsilon$$

Esta función es intrínsecamente lineal, porque se puede transformar a una ecuación lineal mediante la **transformación logarítmica**

$$\ln y = \ln \beta_0 + \beta_1 x + \ln \varepsilon$$

o sea

$$y' = \beta_0' + \beta_1 x + \varepsilon'$$

como se ve en la tabla 5.4. Para esta transformación se requiere que los términos transformados de error, $\varepsilon' = \ln \varepsilon$ estén distribuidos en forma normal e independiente, con media cero y varianza σ^2. Eso implica que el error multiplicativo ε en el modelo original tenga distribución log normal. Se deben examinar los residuales del modelo transformado, para ver si las suposiciones son válidas. En general si x y/o y están en la métrica adecuada, es más probable que se satisfagan las suposiciones acostumbradas, aunque no es raro descubrir en esta etapa que es preferible un modelo no lineal (véase el Cap. 12).

También son útiles varios tipos de transformaciones recíprocas. Por ejemplo, el modelo

$$y = \beta_0 + \beta_1 \left(\frac{1}{x}\right) + \varepsilon$$

se puede linealizar usando la **transformación recíproca** $x' = 1/x$. El modelo linealizado que resulta es

$$y = \beta_0 + \beta_1 x' + \varepsilon$$

Otros modelos que se pueden linealizar mediante transformaciones recíprocas son

$$\frac{1}{y} = \beta_0 + \beta_1 x + \varepsilon$$

y

$$y = \frac{x}{\beta_0 x - \beta_1 + \varepsilon}$$

Este último modelo se ilustra en la figura 5.4g, h.

http://carlos2524.jimdo.com/

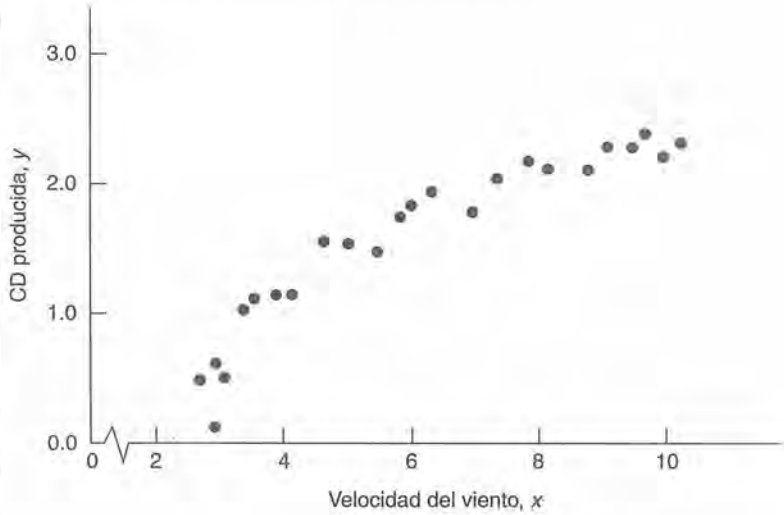


Figura 5.5 Gráfica de la corriente directa (CD) producida, y, en función de la velocidad del viento, x, para los datos del molino de viento.

Cuando se emplean transformaciones como las anteriores, el estimador de mínimos cuadrados tiene propiedades de mínimos cuadrados con respecto a los datos transformados, y no a los datos originales. Como lecturas adicionales acerca de transformaciones, véase Atkinson [1983, 1985], Box, Hunter y Hunter [1978], Carroll y Ruppert [1985], Dolby [1963], Mosteller y Tukey [1977, Caps. 4 a 6], Myers [1990], Smith [1972] y Tukey [1957].

Ejemplo 5.2 *Datos de molinos de viento*

Un ingeniero investiga el uso de un molino de viento para generar electricidad. Ha reunido datos sobre la corriente directa (CD) producida con su molino y la velocidad correspondiente del viento. Los datos se ven en la figura 5.5, y aparecen en la tabla 5.5.

Al inspeccionar el diagrama de dispersión parece que la relación entre la CD producida (y) y la velocidad del viento (x) puede ser no lineal. Sin embargo, para iniciar se ajusta un modelo recitilíneo a los datos. Este modelo de regresión es

$$\hat{y} = 0.1309 + 0.2411x$$

Los estadísticos de resumen para este modelo son $R^2 = 0.8745$, $MS_{\text{Res}} = 0.0557$ y $F_0 = 160.26$ (el valor P es < 0.0001). La columna A de la tabla 5.6 muestra los valores ajustados y los residuales obtenidos de este modelo, en esta tabla, las observaciones están ordenadas por velocidad creciente de viento. Los residuales muestran una distribución definida; esto es, se mueven en forma sistemática de negativos a positivos y regresan a negativos a medida que aumenta la velocidad del viento.

En la figura 5.6 se presenta una gráfica de los residuales en función de $\hat{y}_i$. Esta gráfica de residuales indica que hay inadecuación de modelo, e implica que la relación lineal no ha plasmado toda la información de la variable velocidad del viento. Nótese que la curvatura que ya se notaba en el diagrama de dispersión de la figura 5.5 se amplifica

http://carlos2524.jimdo.com/

TABLA 5.5 Valores observados y_i y variable regresora x_i para el ejemplo 5.2

Observación número, i	Velocidad del viento (mph), x_i	CD producida, y_i
1	5.00	1.582
2	6.00	1.822
3	3.40	1.057
4	2.70	0.500
5	10.00	2.236
6	9.70	2.386
7	9.55	2.294
8	3.05	0.558
9	8.15	2.166
10	6.20	1.866
11	2.90	0.653
12	6.35	1.930
13	4.60	1.562
14	5.80	1.737
15	7.40	2.088
16	3.60	1.137
17	7.85	2.179
18	8.80	2.112
19	7.00	1.800
20	5.45	1.501
21	9.10	2.303
22	10.20	2.310
23	4.10	1.194
24	3.95	1.144
25	2.45	0.723

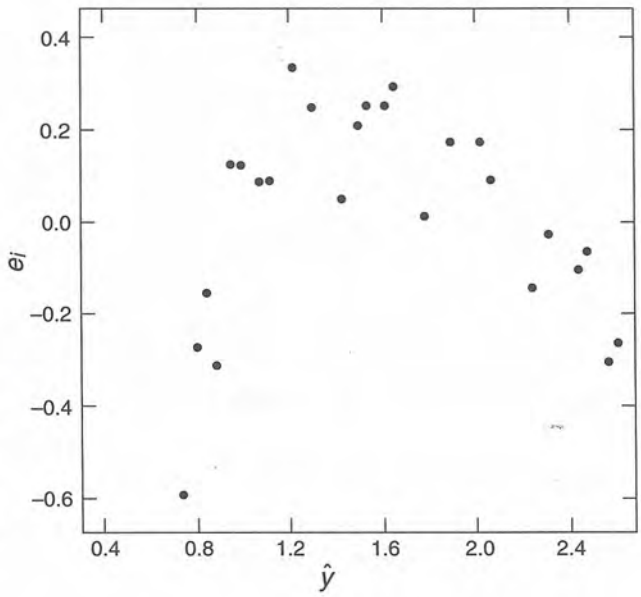


Figura 5.6 Gráfica de residuales e_i en función de los valores ajustados $\hat{y}_i$, para los datos del molino de viento.

http://carlos2524.jimdo.com/

TABLA 5.6 Observaciones y_i por velocidad de viento creciente, valores ajustados $\hat{y}_i$ y residuales ei para los dos modelos del ejemplo 5.2

Velocidad del viento, x_i	CD producida, y_i	A. Modelo rectilíneo $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x$		B. Modelo transformado $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(1/x)$	
		$\hat{y}_i$	e_i	y_i	e_i
2.45	0.123	0.7217	−0.5987	0.1484	−0.0254
2.70	0.500	0.7820	−0.2820	0.4105	0.0895
2.90	0.653	0.8302	−0.1772	0.5876	0.0654
3.05	0.558	0.8664	−0.3084	0.7052	−0.1472
3.40	1.057	0.9508	0.1062	0.9393	0.1177
3.60	1.137	0.9990	0.1380	1.0526	0.0844
3.95	1.144	1.0834	0.0606	1.2233	−0.0793
4.10	1.194	1.1196	0.0744	1.2875	−0.0935
4.60	1.562	1.2402	0.3218	1.4713	0.0907
5.00	1.582	1.3366	0.2454	1.5920	−0.0100
5.45	1.501	1.4451	0.0559	1.7065	−0.2055
5.80	1.737	1.5295	0.2075	1.7832	−0.0462
6.00	1.822	1.5778	0.2442	1.8231	−0.0011
6.20	1.866	1.6260	0.2400	1.8604	0.0056
6.35	1.930	1.6622	0.2678	1.8868	0.0432
7.00	1.800	1.8189	−0.0189	1.9882	−0.1882
7.40	2.088	1.9154	0.1726	2.0418	0.0462
7.85	2.179	2.0239	0.1551	2.0955	0.0835
8.15	2.166	2.0962	0.0698	2.1280	0.0380
8.80	2.112	2.2530	−0.1410	2.1908	−0.0788
9.10	2.303	2.3252	−0.0223	2.2168	0.0862
9.55	2.294	2.4338	−0.1398	2.2527	−0.1472
9.70	2.386	2.4700	−0.0840	2.2640	0.1220
10.00	2.236	2.5424	−0.3064	2.2854	−0.0494
10.20	2.310	2.5906	−0.2906	2.2990	0.0110

mucho en las gráficas de residuales. Es claro que se debe considerar alguna otra forma de modelo.

Se podría comenzar con un modelo cuadrático como el siguiente:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

para explicar la curvatura aparente. Sin embargo, el diagrama de dispersión de la figura 5.5 parece indicar que a medida que la velocidad del viento aumenta, la CD producida

http://carlos2524.jimdo.com/

tiende a un límite aproximado de 2.5. Esto es consistente con la teoría del funcionamiento de los molinos de viento. Como el modelo cuadrático al final se doblaría hacia abajo, a medida que aumenta la velocidad del viento, no sería el adecuado para estos datos. Un modelo más razonable, que establece una asíntota superior, sería

$$y = \beta_0 + \beta_1\left(\frac{1}{x}\right) + \varepsilon$$

La figura 5.7 es un diagrama de dispersión con la variable transformada $x' = 1/x$. Esta gráfica parece lineal, lo cual indica que la transformación recíproca es adecuada. El modelo ajustado de regresión es

$$\hat{y} = 2.9789 - 6.9345x'$$

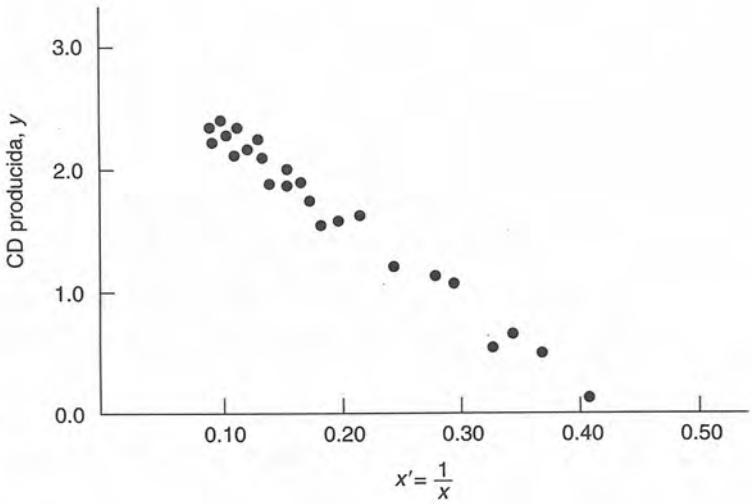


Figura 5.7 Gráfica de la corriente directa (CD) producida en función de $x' = 1/x$ para los datos del molino de viento.

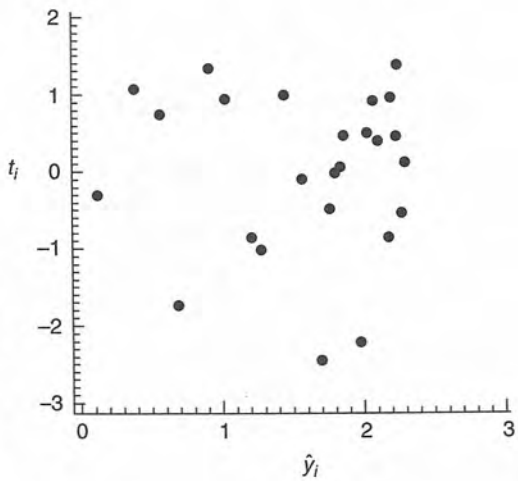


Figura 5.8 Gráfica de los valores de R de Student, t_i, en función de los valores ajustados $\hat{y}_i$, para el modelo transformado, con los datos del molino de viento.

http://carlos2524.jimdo.com/

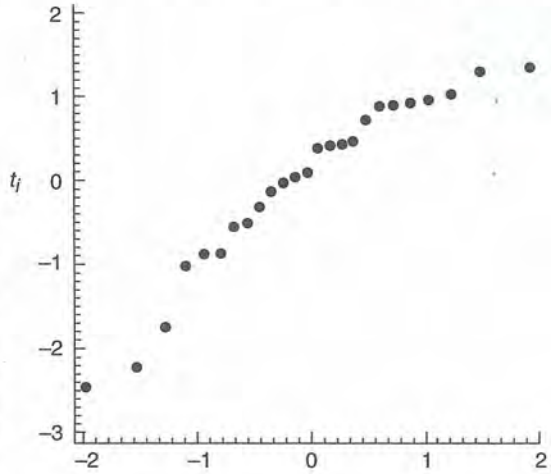


Figura 5.9 Gráfica de probabilidad normal de los residuales del modelo transformado para los datos del molino de viento.

Los estadísticos de resumen para este modelo son $R^2 = 0.9800$, $MS_{\text{Res}} = 0.0089$ y $F_0 = 1\,128.43$ (el valor P es < 0.0001).

Los valores ajustados y los residuales correspondientes al modelo transformado están en la columna B de la tabla 5.6. La figura 5.8 muestra una gráfica de los residuales R de Student del modelo transformado, en función de $\hat{y}$. En esta gráfica no se advierte problema grave alguno de desigualdad de varianza. La gráfica de probabilidad normal, que muestra la figura 5.9, parece indicar que los errores provienen de una distribución de colas más gruesas que la normal (nótense las ligeras curvaturas, hacia arriba y hacia abajo, en los extremos). Como no hay fuerte indicación de inadecuación del modelo, se llega a la conclusión que el modelo transformado es satisfactorio.

5.4 MÉTODOS ANALITICOS PARA SELECCIONAR UNA TRANSFORMACIÓN

Si bien en muchos casos las transformaciones se seleccionan en forma empírica, se pueden aplicar técnicas más formales y objetivas para ayudar a especificar una transformación adecuada. Esta sección describirá e ilustrará procedimientos analíticos para seleccionar transformaciones, tanto de la variable de respuesta como de las variables regresoras.

5.4.1 Transformaciones de y: el método de Box-Cox

Supóngase que se debe transformar y para corregir la no normalidad y/o la varianza no constante. Una clase útil de transformaciones es la de la **transformación de potencia** y^λ, donde λ es un parámetro que se debe determinar. Por ejemplo, $\lambda = \frac{1}{2}$ quiere decir usar y como respuesta. Box y Cox [1964] indican cómo se pueden estimar en forma simultánea los parámetros del modelo de regresión y λ, con el método de la máxima posibilidad.

Al imaginarse la transformación de potencia y^λ surge una dificultad cuando $\lambda = 0$: cuando λ tiende a cero, y^λ tiende a la unidad. Es obvio que es un problema, porque no tiene sentido tener todos los valores de respuesta iguales a uno. Un método para resolver esta dificultad (lo llamaremos discontinuidad en $\lambda = 0$) es usar $(y^\lambda - 1)/\lambda$ como variable de respuesta. Con esto se resuelve el problema de discontinuidad, porque cuando λ tiende a cero

http://carlos2524.jimdo.com/

$(y^{\lambda} - 1)/\lambda$ tiende al límite $\ln y$. Sin embargo, sigue habiendo un problema, porque cuando cambia λ, los valores de $(y^{\lambda} - 1)/\lambda$ cambian en forma dramática, por lo que sería difícil comparar los estadísticos de resumen de modelos con distintos valores de λ.

El procedimiento correcto es usar

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda - 1}}, & \lambda \neq 0 \\ \dot{y} \ln y, & \lambda = 0 \end{cases} \tag{5.1}$$

en las que $\dot{y} = \ln^{-1}[(1/n)\Sigma_{i=1}^{n} \ln y_i]$ es el promedio geométrico de las observaciones, y ajustar el modelo

$$\mathbf{y}^{(\lambda)} = \mathbf{X}\beta + \varepsilon \tag{5.2}$$

por mínimos cuadrados (o por máxima posibilidad). Sucede que el divisor $\dot{y}^{\lambda-1}$ se relaciona con el **jacobiano** de la transformación que convierte la variable de respuesta y en $y^{(\lambda)}$. Es, de hecho, un factor de escala que asegura que las sumas de cuadrados de residuales sean comparables para modelos con distintos valores de λ.

Procedimiento de cálculo

El estimado de λ por máxima posibilidad corresponde al valor de λ para el cual es mínima la suma de cuadrados de residuales del modelo ajustado, $SS_{\text{Res}}(\lambda)$. Este valor de λ se suele determinar ajustando un modelo a $y^{(\lambda)}$ para diversos valores de λ, graficando la suma de cuadrados de residuales $SS_{\text{Res}}(\lambda)$ en función de λ y viendo el valor de λ que minimiza $SS_{\text{Res}}(\lambda)$ en la gráfica. En general, son suficientes de 10 a 20 valores de λ para estimar el valor óptimo. Se puede hacer una segunda iteración con una malla más fina, si se desea. Como se vio arriba, **no** se puede seleccionar λ sólo comparando **en forma directa** las sumas de cuadrados de residuales de las regresiones de y^{λ} respecto a x, porque para cada λ, la suma de cuadrados de residuales se mide en una escala distinta. La ecuación 5.1 escala las respuestas de tal modo que las sumas de cuadrados de residuales se pueden comparar en forma directa. Se recomienda al analista usar valores sencillos de λ, porque es probable que la diferencia práctica en los ajustes para $\lambda = 0.5$ y $\lambda = 0.596$ sea pequeña, pero es más fácil interpretar el primer valor.

Una vez seleccionado un valor de λ, el analista queda libre para ajustar el modelo usando a y^{λ} como variable de respuesta si $\lambda \neq 0$. Si $\lambda = 0$, se usa en y como variable de respuesta. Es totalmente admisible usar $y^{(\lambda)}$ como respuesta para el modelo final; este modelo tendrá una escala diferente y un origen trasladado en comparación del que usa y^{λ} (o que usa $\ln y$). De acuerdo con nuestra experiencia, la mayor parte de los ingenieros y los científicos prefieren usar y^{λ} (o $\ln y$) como respuesta.

Un intervalo de confianza aproximado para λ

También se puede determinar un intervalo de confianza aproximado para el parámetro de transformación λ. Este intervalo de confianza puede servir para seleccionar el valor definitivo de λ. Por ejemplo, si $\hat{\lambda} = 0.596$ es el valor que minimiza la suma de cuadrados de residuales y si $\lambda = 0.5$ está en el intervalo de confianza, sería preferible usar la transformación

http://carlos2524.jimdo.com/

de raíz cuadrada, con base en que es más fácil su explicación. Además, si $\lambda = 1$ está en el intervalo de confianza, no habrá necesidad de transformación alguna.

Al aplicar el método de la máxima posibilidad al modelo de regresión, lo que en esencia se está maximizando es

$$L(\lambda) = -\tfrac{1}{2} n \ln[SS_{\text{Res}}(\lambda)] \tag{5.3}$$

o bien, lo que es igual, se está minimizando la función suma de cuadrados de residuales $SS_{\text{Res}}(\lambda)$. Un intervalo adecuado de confianza de $100(1-\alpha)$ por ciento para λ es el de todos aquellos valores que satisfacen la desigualdad

$$L(\hat{\lambda}) - L(\lambda) \leq \tfrac{1}{2}\chi^2_{\alpha,1}/n \tag{5.4}$$

en donde $\lambda^2_{\alpha,1}$ es el punto porcentual superior de la distribución ji cuadrada con un grado de libertad. Para trazar el intervalo de confianza en realidad se trazaría, sobre una gráfica de $L(\lambda)$ en función de λ, una recta horizontal a la altura

$$L(\hat{\lambda}) - \tfrac{1}{2}\chi^2_{\alpha,1}$$

en la escala vertical. Esa línea cortaría a la curva de $L(\lambda)$ en dos puntos, cuyos lugares en el eje de λ definen los dos extremos del intervalo aproximado de confianza. Si se está minimizando la suma de cuadrados de residuales y graficando $SS_{\text{Res}}(\lambda)$ en función de λ, entonces la línea se debe graficar a la altura

$$SS^* = SS_{\text{Res}}(\hat{\lambda})\, e^{\chi^2_{\alpha,1}/n} \tag{5.5}$$

Recuérdese que $\hat{\lambda}$ es el valor de λ que minimiza la suma de cuadrados de residuales.

Al aplicar en realidad el procedimiento del intervalo de confianza es probable que se determine que el factor $\exp(\chi^2_{\alpha,1}/n)$ del lado derecho de la ecuación (5.5) está sustituido ya sea por $1 + z^2_{\alpha/2}/n$, o por $1 + t^2_{\alpha/2,\nu}/n$, o también $1 + \chi^2_{\alpha,1}/n$ o hasta por ya sea $1 + z^2_{\alpha/2}/\nu$ o $1 + t^2_{\alpha/2,n}/\nu$ o por $1 + \chi^2_{\alpha,1}/\nu$, siendo ν la cantidad de grados de libertad de los residuales. Éstos se basan en el desarrollo de $\exp(x) = 1 + x + x^2/2! + x^3/3! + \cdots = 1 + x$, y en el hecho de que $\chi^2_1 = z^2 \approx t^2_\nu$, a menos que la cantidad de grados de libertad de los residuales ν sea pequeña. Quizá se pueda debatir si se debe usar n o ν, pero en la mayor parte de los casos habrá poca diferencia entre los intervalos de confianza que resultan.

Ejemplo 5.3 *Datos de la compañía eléctrica*

Con los datos de la empresa eléctrica que se presentaron en el ejemplo 5.1, se aplicará el procedimiento de Box-Cox para seleccionar una transformación estabilizadora de varianza. En la tabla 5.7 se ven los valores de $SS_{\text{Res}}(\lambda)$ correspondientes a diversas λ. En esa tabla se ve que $\lambda = 0.5$ (la transformación de raíz cuadrada) se acerca mucho al valor óptimo. Nótese que se ha usado una "malla" más fina de λ en la proximidad del valor óptimo. Esto ayuda para ubicar a λ con más precisión, y para graficar la función de suma de cuadrados de residuales.

En la figura 5.10 se ve una gráfica de la suma de cuadrados de residuales en función de λ. Si se toma a $\lambda = 0.5$ como valor óptimo, se puede determinar un intervalo aproximado

http://carlos2524.jimdo.com/

TABLA 5.7 Valores de la suma de cuadrados de residuales para diversos valores de λ, ejemplo 5.3

λ	$SS_{\text{Res}}(\lambda)$
− 2	34 101.0381
− 1	986.0423
− 0.5	291.5834
0	134.0940
0.125	118.1982
0.25	107.2057
0.375	100.2561
0.5	96.9495
0.625	97.2889
0.75	101.6869
1	126.8660
2	1 275.5555

de confianza de 95% para λ, calculando la suma crítica de cuadrados SS^* con la ecuación (5.5) como sigue:

$$\begin{aligned} SS^* &= SS_{\text{Res}}(\hat{\lambda})\, e^{\chi^2_{0.05,1}/n} \\ &= 96.9495 e^{3.84/53} \\ &= 96.9495\,(1.0751) \\ &= 104.23 \end{aligned}$$

La recta horizontal a esta altura se ve en la figura 5.10. Los valores correspondientes de $\lambda^- = 0.26$ y $\lambda^+ = 0.80$ determinados con la curva son los límites inferior y superior para λ,

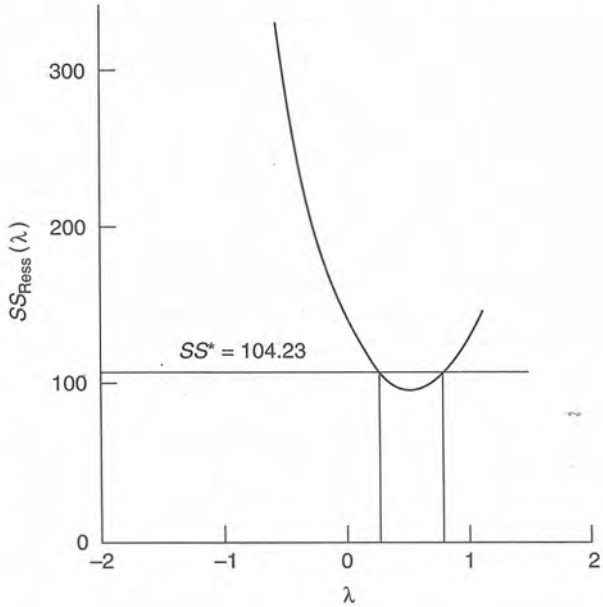


Figura 5.10 Gráfica de la suma de cuadrados de residuales, $SS_{\text{Res}}(\lambda)$, en función de λ.

http://carlos2524.jimdo.com/

respectivamente. Como en esos límites no se incluye el valor 1 (que implica que no hay transformación), la conclusión es que es útil una transformación. Además, la transformación de raíz cuadrada que se usó en el ejemplo 5.1 tiene justificación analítica.

5.4.2 Transformaciones de las variables regresoras

Supóngase que la relación entre y y una o más de las variables regresoras es no lineal, pero que las premisas acostumbradas de respuestas distribuidas en forma normal e independiente, con varianza constante, se satisfacen al menos aproximadamente. Se debe seleccionar una transformación adecuada, de las variables regresoras, tal que la relación entre y y el regresor transformado sea tan simple como fuere posible. Box y Tidwell [1962] describen un procedimiento analítico para determinar la forma de la transformación de x. Si bien se puede usar su procedimiento en el caso general de la regresión, presentaremos e ilustraremos esa aplicación al modelo de regresión lineal simple.

Supóngase que la variable y de respuesta se relaciona con una potencia del regresor, por ejemplo $\xi = x^{\alpha}$, como sigue:

$$E(y) = f(\xi, \beta_0, \beta_1) = \beta_0 + \beta_1 \xi$$

donde

$$\xi = \begin{cases} x^{\alpha}, & \alpha \neq 0 \\ \ln x, & \alpha = 0 \end{cases}$$

y β_0, β_1 y α son parámetros desconocidos. Supóngase que α_0 es un tanteo inicial de la constante α. Por lo general, este primer tanteo es $\alpha_0 = 1$, por lo que $\xi_0 = x^{\alpha 0} = x$, es decir, en la primera iteración no se aplica transformación alguna. Al desarrollar en una serie de Taylor respecto al tanteo inicial, e ignorar los términos de orden mayor que uno, se obtiene

$$\begin{aligned} E(y) &= f(\xi_0, \beta_0, \beta_1) + (\alpha - \alpha_0)\left\{\frac{df(\xi, \beta_0, \beta_1)}{d\alpha}\right\}_{\substack{\xi=\xi_0 \\ \alpha=\alpha_0}} \\ &= \beta_0 + \beta_1 x + (\alpha - 1)\left\{\frac{df(\xi, \beta_0, \beta_1)}{d\alpha}\right\}_{\substack{\xi=\xi_0 \\ \alpha=\alpha_0}} \end{aligned} \tag{5.6}$$

Ahora bien, si se conociera el término entre llaves de esta ecuación, se podría manejar como una variable regresora adicional y sería posible estimar los parámetros β_0, β_1 y α en la ecuación (5.6) mediante mínimos cuadrados. El estimado de α se podría tomar entonces como un estimado mejorado del parámetro de transformación. El término entre llaves de la ecuación (5.6) se puede escribir en la siguiente forma:

$$\left\{\frac{df(\xi, \beta_0, \beta_1)}{d\alpha}\right\}_{\substack{\xi=\xi_0 \\ \alpha=\alpha_0}} = \left\{\frac{df(\xi, \beta_0, \beta_1)}{d\xi}\right\}_{\xi=\xi_0}\left\{\frac{d\xi}{d\alpha}\right\}_{\alpha=\alpha_0}$$

y como se conoce la forma de la transformación, esto es, $\xi = x^{\alpha}$, entonces $d\xi/d\alpha = x \ln x$. Además,

http://carlos2524.jimdo.com/

$$\left\{\frac{df(\xi, \beta_0, \beta_1)}{d\xi}\right\}_{\xi=\xi_0} = \frac{d(\beta_0 + \beta_1 x)}{dx} = \beta_1$$

Este parámetro se puede estimar en forma conveniente ajustando el modelo

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{5.7}$$

por mínimos cuadrados. A continuación se puede calcular un "ajuste" del tanteo inicial $\alpha_{0=1}$ definiendo una segunda variable regresora como $w = x \ln x$ y estimando los parámetros en

$$\begin{aligned} E(y) &= \beta_0^* + \beta_1^* x + (\alpha - 1)\beta_1 w \\ &= \beta_0 + \beta_1^* x + \gamma w \end{aligned} \tag{5.8}$$

por mínimos cuadrados, para obtener

$$\hat{y} = \hat{\beta}_0^* + \hat{\beta}_1^* + \hat{\gamma} w \tag{5.9}$$

y definiendo

$$\alpha_1 = \frac{\hat{\gamma}}{\hat{\beta}_1} + 1 \tag{5.10}$$

como estimado revisado de α. Nótese que $\hat{\beta}_1$ se obtiene con la ecuación (5.7) y $\hat{\gamma}$ con la ecuación (5.9); en general, $\hat{\beta}_1$ y $\hat{\beta}_1^*$ serán distintas. Ahora se puede repetir este procedimiento usando un nuevo regresor, $x' = x^{\alpha_1}$ en los cálculos.

Box y Tidwell [1962] hacen notar que en general este procedimiento converge con mucha rapidez, y que con frecuencia la α_1 que resulta en la primera etapa es un estimado satisfactorio de α. También previenen que el error de redondeo es un problema potencial, y que los valores sucesivos de α pueden oscilar locamente a menos que se conserven los suficientes lugares decimales. Se pueden encontrar problemas de convergencia en casos en los que la desviación estándar σ del error es grande, o cuando el rango del regresor es muy pequeño en comparación con su media. Esta situación implica que los datos no respaldan la necesidad de transformación alguna.

Ejemplo 5.4 *Datos del molino de viento*

Se ilustrará este procedimiento usando los datos del molino de viento del ejemplo 5.2. El diagrama de dispersión de la figura 5.5 parece indicar que la relación entre la CD producida (y) y la velocidad del viento (x) no es una recta, y que sería adecuada alguna transformación de x.

Se comenzará con el tanteo inicial $\alpha_0 = 1$ y se ajustará un modelo de línea recta, que da como resultado $\hat{y} = 0.1309 + 0.2411x$. A continuación, con la definición $w = x \ln x$ se ajusta la ecuación (5.8) y se obtiene

$$\begin{aligned} \hat{y} &= \hat{\beta}_0^* + \hat{\beta}_1^* x + \hat{\gamma} w \\ &= -2.4168 + 1.5344x - 0.4626w \end{aligned}$$

http://carlos2524.jimdo.com/

Con la ecuación (5.10) se calcula

$$\alpha_2 = \frac{\hat{\gamma}}{\hat{\beta}_1} + \alpha_1 = \frac{0.5994}{-6.6784} + (-0.92) = -1.01$$

como estimado mejorado de α. Nótese que este estimado de α se acerca mucho a -1, por lo que la transformación recíproca de x que se usó en realidad en el ejemplo 5.2 queda respaldada por el procedimiento de Box-Tidwell.

Para hacer una segunda iteración se definiría una nueva variable regresora $x' = x^{-0.92}$ y se ajustaría el modelo

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x' = 3.1039 - 6.6784x'$$

A continuación se forma un segundo regresor $w' = x' \ln x'$ y se ajusta

$$\begin{aligned}\hat{y} &= \hat{\beta}_0^* + \hat{\beta}_1^* x' + \hat{\gamma} w' \\ &= 3.2409 - 6.445x' + 0.5994w'\end{aligned}$$

Así, el estimado de segundo paso de α es

$$\alpha_2 = \frac{\hat{\gamma}}{\hat{\beta}_1} + \alpha_1 = \frac{0.5994}{-6.6784} + (-0.92) = -1.01$$

que de nuevo respalda el uso de la transformación recíproca de x.

5.5 MÍNIMOS CUADRADOS GENERALIZADOS Y PONDERADOS

Los modelos de regresión lineal con varianza de error no constante también se pueden ajustar con el método de los **mínimos cuadrados ponderados**. En este método de estivación se multiplica la diferencia entre los valores observados y esperados de y_i por un **peso** w_i, o **factor de ponderación**, que se escoge como inversamente proporcional a la varianza de y_i. Para el caso de la regresión lineal simple, la función de mínimos cuadrados ponderados es

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i)^2 \tag{5.11}$$

Las ecuaciones normales de mínimos cuadrados que resultan son

$$\begin{aligned}\hat{\beta}_0 \sum_{i=1}^{n} w_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i &= \sum_{i=1}^{n} w_i y_i \\ \beta_0 \sum_{i=1}^{n} w_i x_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i^2 &= \sum_{i=1}^{n} w_i x_i y_i\end{aligned} \tag{5.12}$$

Al resolver las ecuaciones (5.12) se obtendrán los estimados ponderados, por mínimos cuadrados, del β_0 y β_1.

http://carlos2524.jimdo.com/

En esta sección se desarrollarán los mínimos cuadrados ponderados para el modelo de regresión múltiple. Se comenzará considerando un caso un poco más general que concierne a la estructura de los errores del modelo.

5.5.1 Mínimos cuadrados generalizados

Las suposiciones que se suelen hacer acerca del modelo de regresión lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ son que $E(\varepsilon) = \mathbf{0}$ y que Var $(\varepsilon) = \sigma^2\mathbf{I}$. Como se ha observado, a veces esas premisas son irrazonables, por lo que ahora se examinará qué modificaciones se necesitan para este procedimiento de mínimos cuadrados ordinarios, cuando Var $(\varepsilon) = \sigma^2\mathbf{V}$, siendo $\mathbf{V}$ una matriz conocida de $n \times n$. Este caso tiene una interpretación fácil: si $\mathbf{V}$ es diagonal pero con elementos diagonales distintos, las observaciones $\mathbf{y}$ son **no correlacionadas**, pero tienen **varianzas desiguales**, mientras que si algunos de los elementos fuera de la diagonal de $\mathbf{V}$ son distintos de cero, las observaciones son **correlacionadas**.

Cuando el modelo es

$$\begin{aligned} y &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ E(\boldsymbol{\varepsilon}) &= \mathbf{0},\ \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V} \end{aligned} \tag{5.13}$$

el estimador de mínimos cuadrados ordinarios $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ ya no es adecuado. Se resolverá este problema transformando el modelo en un nuevo conjunto de observaciones que satisfagan las premisas estándar de mínimos cuadrados. A continuación se usarán mínimos cuadrados ordinarios con los datos transformados. Como $\sigma^2\mathbf{V}$ es la matriz de covarianza de los errores, $\mathbf{V}$ debe ser no singular y positiva definida, y en consecuencia existe una matriz $\mathbf{K}$, no singular y simétrica, de $n \times n$, tal que $\mathbf{K}'\mathbf{K} = \mathbf{K}\mathbf{K} = \mathbf{V}$. A menudo se le llama **raíz cuadrada** de $\mathbf{V}$ a la matriz $\mathbf{K}$. En forma típica, σ^2 se desconoce, y en ese caso $\mathbf{V}$ representa la estructura supuesta de las varianzas y covarianzas entre los errores aleatorios, aparte de una constante.

Se definen las nuevas variables

$$\mathbf{z} = \mathbf{K}^{-1}\mathbf{y}, \quad \mathbf{B} = \mathbf{K}^{-1}\mathbf{X} \quad \text{y} \quad \mathbf{g} = \mathbf{K}^{-1}\varepsilon \tag{5.14}$$

por lo que el modelo de regresión $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ se transforma en $\mathbf{K}^{-1}\mathbf{y} = \mathbf{K}^{-1}\mathbf{X}\beta + \mathbf{K}^{-1}\varepsilon$, es decir

$$\mathbf{z} = \mathbf{B}\beta + \mathbf{g} \tag{5.15}$$

Los errores en este modelo transformado tienen valor esperado cero, esto es, $E(\mathbf{g}) = \mathbf{K}^{-1}E(\varepsilon) = \mathbf{0}$. Además, la matriz de covarianza de $\mathbf{g}$ es

$$\begin{aligned} \mathrm{Var}(\mathbf{g}) &= \{[\mathbf{g} - E(\mathbf{g})][\mathbf{g} - E(\mathbf{g})]'\} \\ &= E(\mathbf{g}\mathbf{g}') \\ &= E(\mathbf{K}^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{K}^{-1}) \\ &= \mathbf{K}^{-1}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{K}^{-1} \\ &= \sigma^2\mathbf{K}^{-1}\mathbf{V}\mathbf{K}^{-1} \\ &= \sigma^2\mathbf{K}^{-1}\mathbf{K}\mathbf{K}\mathbf{K}^{-1} \\ &= \sigma^2\mathbf{I} \end{aligned} \tag{5.16}$$

http://carlos2524.jimdo.com/

Por lo anterior, los elementos de **g** tienen media cero y varianza constante, y no están correlacionados. Como los errores **g** en el modelo (5.15) satisfacen las premisas acostumbradas, se pueden aplicar los mínimos cuadrados ordinarios. La función de mínimos cuadrados es

$$\begin{aligned} S(\boldsymbol{\beta}) &= \mathbf{g}'\mathbf{g} = \boldsymbol{\varepsilon}'\mathbf{V}^{-1}\boldsymbol{\varepsilon} \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned} \tag{5.17}$$

Las ecuaciones de mínimos cuadrados son

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\,\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{5.18}$$

y la solución de esas ecuaciones es

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{5.19}$$

Aquí, la $\hat{\boldsymbol{\beta}}$ se llama **estimador de mínimos cuadrados generalizado** de $\boldsymbol{\beta}$.

No es difícil demostrar que $\hat{\boldsymbol{\beta}}$ es un estimador insesgado de $\boldsymbol{\beta}$. La matriz de covarianza de $\hat{\boldsymbol{\beta}}$ es

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{B}'\mathbf{B})^{-1} = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \tag{5.20}$$

En el apéndice C.10 se demuestra que $\hat{\boldsymbol{\beta}}$ es el mejor estimador lineal insesgado de $\boldsymbol{\beta}$. El análisis de varianza en términos de los mínimos cuadrados generalizados se resume en la tabla 5.8.

5.5.2 Mínimos cuadrados ponderados

Cuando los errores ε no están correlacionados, pero tienen varianzas desiguales de modo que la matriz de covarianza de ε sea, por ejemplo

$$\sigma^2\mathbf{V} = \sigma^2 \begin{bmatrix} \frac{1}{w_1} & & & 0 \\ & \frac{1}{w_2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{w_n} \end{bmatrix}$$

TABLA 5.8 Análisis de varianza para mínimos cuadrados generalizados

Fuente	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0
Regresión	$SS_R = \hat{\boldsymbol{\beta}}'\mathbf{B}'\mathbf{z} = \mathbf{y}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$	p	SS_R/p	MS_R/MS_{Res}
Error	$SS_{Res} = \mathbf{z}'\mathbf{z} - \hat{\boldsymbol{\beta}}'\mathbf{B}'\mathbf{z} = \mathbf{y}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{y}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$	$n - p$	$SS_{Res}/(n-p)$	
Total	$\mathbf{z}'\mathbf{z} = \mathbf{y}'\mathbf{V}^{-1}\mathbf{y}$	n		

http://carlos2524.jimdo.com/

al procedimiento de estimación se le suele llamar **mínimos cuadrados ponderados**. Sea $\mathbf{W} = \mathbf{V}^{-1}$. Como $\mathbf{V}$ es una matriz diagonal, $\mathbf{W}$ también es diagonal y sus elementos diagonales, **pesos**, o **factores de ponderación**, son $w_1, w_2, \ldots, w_n$. De acuerdo con la ecuación (5.18), las ecuaciones normales de mínimos cuadrados ponderados son

$$(\mathbf{X}'\mathbf{W}\mathbf{X})\,\hat{\beta} = \mathbf{X}'\mathbf{W}\mathbf{y}$$

Éste es un análogo, en regresión múltiple, de las ecuaciones normales de mínimos cuadrados ponderados para la regresión lineal simple, de la ecuación (5.12). En consecuencia,

$$\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$$

es el **estimador ponderado de mínimos cuadrados**. Nótese que las observaciones con varianzas grandes tienen menos peso que las de varianzas pequeñas.

Los estimados por mínimos cuadrados ponderados se pueden obtener con facilidad con un programa ordinario de cómputo de mínimos cuadrados. Si se multiplica cada uno de los valores observados de la i-ésima observación (incluyendo el 1 de la ordenada al origen) por la raíz cuadrada del peso de esa observación, se obtendrá entonces un conjunto de datos transformados.

$$\mathbf{B} = \begin{bmatrix} 1\sqrt{w_1} & x_{11}\sqrt{w_1} & \cdots & x_{1k}\sqrt{w_1} \\ 1\sqrt{w_2} & x_{21}\sqrt{w_2} & \cdots & x_{2k}\sqrt{w_2} \\ \vdots & \vdots & & \vdots \\ 1\sqrt{w_n} & x_{n1}\sqrt{w_n} & \cdots & x_{nk}\sqrt{w_n} \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} y_1\sqrt{w_1} \\ y_2\sqrt{w_2} \\ \vdots \\ y_n\sqrt{w_n} \end{bmatrix}$$

Ahora si se aplican los mínimos cuadrados ordinarios a esos datos transformados, se obtendrá

$$\hat{\beta} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{z} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$$

que es el estimado de β por mínimos cuadrados ponderados.

5.5.3 Algunos asuntos prácticos

Para usar mínimos cuadrados ponderados se deben conocer los pesos w_i. A veces se puede recurrir a la experiencia o conocimientos anteriores, o a la información de un modelo teórico, para determinar los pesos. Véase un ejemplo de este método en Weisberg [1985]. También, el análisis de residuales puede indicar que la varianza de los errores puede ser una función de uno de los regresores, por ejemplo, $\text{Var}(\varepsilon_i) = \sigma^2 x_{ij}$, de modo que $w_i = 1/x_{ij}$. En algunos casos, en realidad y_i es un promedio de n_i observaciones en x_i, y si todas las observaciones **originales** tienen varianza constante σ^2, entonces la varianza de y_1 es $\text{Var}(y_i) = \text{Var}(\varepsilon_i) = \sigma^2/n_i$, y los pesos se escogerían como $w_i = n_i$. A veces, la fuente principal de error es la del error de medición, y distintas observaciones se miden con distintos instrumentos de precisión desigual (pero bien estimada). En ese caso los pesos se podrían elegir inversamente proporcionales a las varianzas del error de medición. En muchos casos prácticos se podrán adivinar los pesos, hacer el análisis para después volver a estimar los pesos con base en los resultados. Pueden ser necesarias varias iteraciones.

http://carlos2524.jimdo.com/

Ya que los mínimos cuadrados generalizados o los ponderados requieren suposiciones adicionales acerca de los errores, es interesante preguntar qué sucede cuando no se hace así y se usan los mínimos cuadrados ordinarios en un caso en el que $\text{Var}(\varepsilon) = \sigma^2\mathbf{V}$, siendo $\mathbf{V} \neq \mathbf{I}$. Si los mínimos cuadrados ordinarios se usaran en este caso, el estimador resultante $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ sigue siendo insesgado. Sin embargo, el estimador de mínimos cuadrados ordinarios ya no será estimador de varianza mínima. Esto es, la matriz de covarianza del estimador de mínimos cuadrados ordinarios es

$$\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \tag{5.21}$$

y la matriz de covarianza del estimador de mínimos cuadrados generalizados, ecuación (5.20), produce varianzas menores de los coeficientes de regresión. Así, los mínimos cuadrados generalizados o ponderados son preferibles a los mínimos cuadrados ordinarios siempre que $\mathbf{V} \neq \mathbf{I}$.

Ejemplo 5.5 *Mínimos cuadrados ponderados*

En las columnas a y b de la tabla 5.9 se muestran los ingresos mensuales promedio por ventas de alimentos, y los correspondientes gastos anuales en publicidad para 30 restaurantes. La gerencia se interesa en la relación entre esas variables, por lo que se ajusta un modelo de regresión lineal que relacione a y, las ventas de alimentos, con x, los gastos en publicidad, mediante mínimos cuadrados ordinarios. Se obtiene $\hat{y} = 49\ 443.3838 + 8.0484x$. Los residuales de este ajuste por mínimos cuadrados se grafican en función de $\hat{y}_i$ en la figura 5.11. En esta gráfica se aprecia la violación de la suposición de varianza constante. En consecuencia, el ajuste por mínimos cuadrados ordinarios es inadecuado.

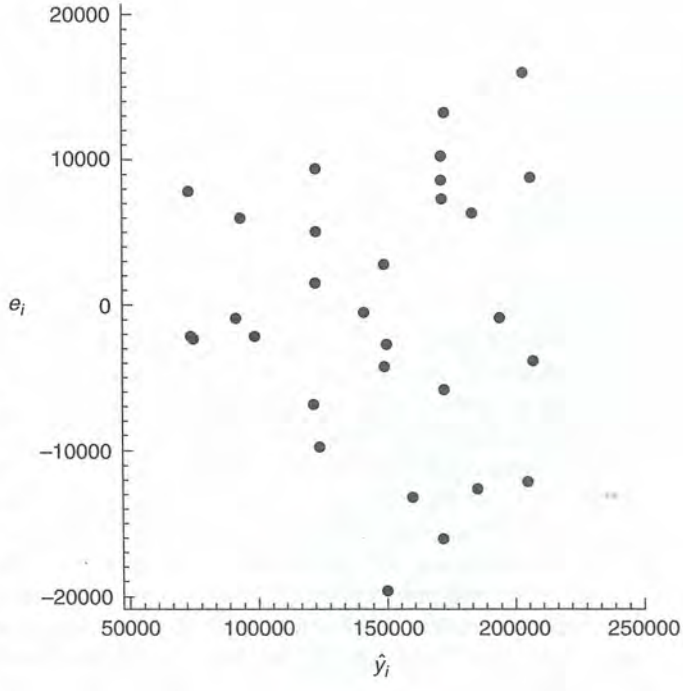


Figura 5.11 Gráfica de los residuales de los mínimos cuadrados ordinarios en función de los valores ajustados, ejemplo 5.5.

http://carlos2524.jimdo.com/

TABLA 5.9 Datos de ventas de alimentos en restaurante

Obs. i	a) Ingreso, y_i	b) Gastos de publicidad, x_i	c) $\bar{x}$	d) s_y^2	e) Pesos, w_i
1	81,464	3,000			6.21771E-08
2	72,661	3,150	3,078.3	26,794,620	5.79507E-08
3	72,344	3,085			5.97094E-08
4	90,743	5,225	5,287.5	30,722,010	2.98667E-08
5	98,588	5,350			2,90195E-08
6	96,507	6,090			2.48471E-08
7	126,574	8,925			1.60217E-08
8	114,133	9,015			1.58431E-08
9	115,814	8,885	8,955.0	52,803,698	1.61024E-08
10	123,181	8,950			1.59717E-08
11	131,434	9,000			1.58726E-08
12	140,564	11,345			1.22942E-08
13	151,352	12,275			1.12852E-08
14	146,926	12,400	12,377.5	77,280,167	1.11621E-08
15	130,963	12,525			1.10416E-08
16	144,630	12,310			1.12505E-08
17	147,041	13,700			1.00246E-08
18	179,021	15,000			9.09750E-09
19	166,200	15,175			8.98563E-09
20	180,732	14,995	15,095.0	120,571,040	9.10074E-09
21	178,187	15,050			9.06525E-09
22	185,304	15,200			8.96988E-09
23	155,931	15,150			9.00144E-09
24	172,579	16,800	16,650.0	132,388,990	8.06478E-09
25	188,851	16,500			8.22031E-09
26	192,424	17,830			7.57287E-09
27	203,112	19,500			6.89136E-09
28	192,482	19,200	19,262.5	138,856,867	7.00460E-09
29	218,715	19,000			7.08218E-09
30	214,317	19,350			6.94752E-09

Para corregir este problema de desigualdad de varianza, se deben conocer los pesos w_i. Al examinar los datos de la tabla 5.9 se nota que hay varios conjuntos de valores de x que son "vecinos cercanos", esto es, que tienen puntos de repetición aproximados en x. Se supondrá que esos vecinos cercanos están lo bastante cerca entre sí para poder considerarse puntos de repetición, y se usará la varianza de las respuestas en dichos puntos para investigar la forma en que Var(y) cambia en función de x. Las columnas c y d de la tabla 5.9 muestran el valor promedio de $x, \bar{x}$, para cada grupo de vecinos cercanos, y la varianza muestral de las y en cada grupo. Graficar s_y^2 en función de la $\bar{x}$ correspondiente, implica que s_y^2 aumenta en forma aproximadamente lineal con $\bar{x}$. Un ajuste por mínimos cuadrados da como resultado

$$\hat{s}_y^2 = -7\,376\,216.04 + 7\,819.77\,\bar{x}$$

http://carlos2524.jimdo.com/

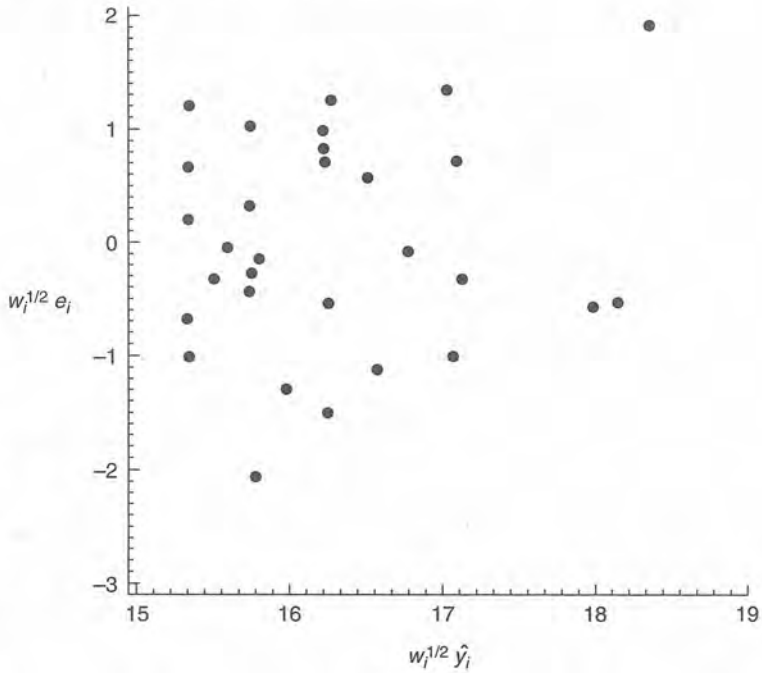


Figura 5.12 Gráfica de los residuales ponderados $w_i^{1/2} e_i$ en función de los valores ajustados ponderados $w_i^{1/2} \hat{y}_i$, ejemplo 5.5.

Al sustituir cada valor de x_i en esta ecuación se obtendrá un estimado de la varianza de la observación y_i correspondiente. Los inversos de esos valores ajustados serán estimados razonables de los pesos w_i. Esos pesos estimados se ven en la columna e de la tabla 5.9.

Al aplicar los mínimos cuadrados a los datos, con los pesos de la tabla 5.9, se obtiene el modelo ajustado

$$\hat{y} = 50\,975.5667 + 7.9222x$$

Ahora se deben examinar los residuales para determinar si el ajuste ha mejorado al usar los mínimos cuadrados ponderados. Para hacerlo, se grafican los residuales **ponderados** $w_i^{1/2} e_i = w_i^{1/2}(y_i - \hat{y}_i)$, donde $\hat{y}_i$ proviene del ajuste por mínimos cuadrados ponderados, en función de $w_i^{1/2}\hat{y}_i$. Esta gráfica se ve en la figura 5.12, que es mucho mejor en comparación con las gráficas previas del ajuste de mínimos cuadrados ordinarios. Se llega a la conclusión que los mínimos cuadrados ponderados han corregido el problema de desigualdad de varianza.

Se deben mencionar otros dos puntos acerca de este ejemplo. El primero es que se tuvo la fortuna de contar con varios vecinos cercanos en el espacio de x. Además, fue fácil identificar esos grupos de puntos, por inspección de la tabla 5.9, porque sólo había implicado un regresor. Cuando hay varios regresores es más difícil la identificación visual de esos grupos. Recuérdese que en el capítulo 4 se presentó un procedimiento analítico para determinar pares de puntos cercanos entre sí en el espacio de x (Sec. 4.5.3). El segundo punto implica el uso de una ecuación de regresión para estimar los pesos. El analista debe comprobar con cuidado los pesos que produce la ecuación, para estar seguro de que son razonables. Por ejemplo, en este problema un valor suficientemente pequeño de x hubiera resultado en un peso negativo, lo cual es irrazonable.

http://carlos2524.jimdo.com/

PROBLEMAS

5.1 Byers y Williams ("Viscosities of Binary and Ternary Mixtures of Polyaromatic Hydrocarbons", *Journal of Chemical and Engineering Data*, **32**, 349–354, 1987) estudiaron el impacto de la temperatura (el regresor) sobre la viscosidad (la respuesta) de las mezclas de tolueno y tetralina. La tabla siguiente muestra los datos para mezclas con fracción molar de tolueno igual a 0.4.

Temperatura (°C)	Viscosidad (mPa · s)
24.9	1.133
35.0	0.9772
44.9	0.8532
55.1	0.7550
65.2	0.6723
75.2	0.6021
85.2	0.5420
95.2	0.5074

a. Trazar un diagrama de dispersión. ¿Parece que será adecuado un modelo de línea recta?
b. Ajustar el modelo rectilíneo. Calcular los estadísticos de resumen y las gráficas de residuales. ¿Cuáles son las conclusiones acerca de la adecuación del modelo?
c. Los principios básicos de la química física indican que la viscosidad es una función exponencial de la temperatura. Repetir la parte b usando la transformación adecuada basada en esta información.

5.2 La siguiente tabla muestra la presión de vapor del agua para diversas temperaturas.

Temperatura (K)	Presión de vapor (mm Hg)
273	4.6
283	9.2
293	17.5
303	31.8
313	55.3
323	92.5
333	149.4
343	233.7
353	355.1
363	525.8
373	760.0

a. Trazar un diagrama de dispersión. ¿Parece que será adecuado un modelo de línea recta?
b. Ajustar el modelo rectilíneo. Calcular los estadísticos de resumen y las gráficas de residuales. ¿Cuáles son las conclusiones acerca de la adecuación del modelo?
c. Según la ecuación de Clausius-Clapeyron, de la química física,

$$\ln(p_v) \ \alpha \ -\frac{1}{T}$$

http://carlos2524.jimdo.com/

Repetir la parte b usando la transformación adecuada basada en esta información.

5.3 Los datos a continuación muestran la cantidad promedio de bacterias que sobreviven en un producto alimenticio enlatado, y los minutos de exposición a la temperatura de 300 °F.

Cantidad de bacterias	Minutos de exposición
175	1
108	2
95	3
82	4
71	5
50	6
49	7
31	8
28	9
17	10
16	11
11	12

a. Trazar un diagrama de dispersión. ¿Es probable que un modelo de línea recta sea adecuado?

b. Ajustar el modelo rectilíneo. Calcular los estadísticos de resumen y trazar las gráficas de residuales. ¿Cuáles son las conclusiones acerca de la adecuación del modelo?

c. Identificar un modelo transformado adecuado para estos datos. Ajustar ese modelo y hacer las pruebas acostumbradas de adecuación del modelo.

5.4 Examinar los datos que se ven a continuación. Trazar un diagrama de dispersión y sugerir una forma adecuada del modelo de regresión. Ajustar este modelo a los datos y hacer las pruebas acostumbradas de adecuación del modelo.

x	10	15	18	12	9	8	11	6
y	0.17	0.13	0.09	0.15	0.20	0.21	0.18	0.24

5.5 Una empresa fabricante de botellas de vidrio ha registrado datos acerca de la cantidad promedio de defectos por 10 000 botellas, debidos a piedras (pequeñas piedras embebidas en la pared de la botella) y la cantidad de semanas a partir de la última reparación general del horno. Los datos son los siguientes:

Defectos por 10 000	Semanas	Defectos por 10 000	Semanas
13.0	4	34.2	11
16.1	5	65.6	12
14.5	6	49.2	13
17.8	7	66.2	14
22.0	8	81.2	15
27.4	9	87.4	16
16.8	10	114.5	17

http://carlos2524.jimdo.com/

a. Ajustar un modelo rectilíneo de regresión a los datos y hacer las pruebas usuales de adecuación del modelo.

b. Sugerir una transformación adecuada que elimine los problemas encontrados en la parte a. Ajustar el modelo transformado y comprobar su adecuación.

5.6 Se tienen los tres modelos siguientes:

a. $y = \beta_0 + \beta_1(1/x) + \varepsilon$

b. $1/y = \beta_0 + \beta_1 x + \varepsilon$

c. $y = x/(\beta_0 - \beta_1 x) + \varepsilon$

Todos estos modelos se pueden linealizar mediante transformaciones recíprocas. Hacer un diagrama del comportamiento de y en función de x. ¿Qué características observadas en el diagrama de dispersión llevarían a la elección de uno de esos modelos?

5.7 Se tienen los datos de formación de catrato de la tabla B.8 del apéndice.

a. Hacer un análisis minucioso de residuales para esos datos.

b. Identificar la transformación más adecuada para esos datos. Ajustar este modelo y repetir el análisis de residuales.

5.8 Para los datos de caída de presión en la tabla B.9 del apéndice:

a. Hacer un análisis minucioso de residuales de esos datos.

b. Identificar la transformación más adecuada para ellos. Ajustar este modelo y repetir el análisis de residuales.

5.9 Para los datos de viscosidad cinemática de la tabla B.10 del apéndice:

a. Hacer un análisis detallado de residuales de esos datos.

b. Identificar la transformación más adecuada para ellos. Ajustar este modelo y repetir el análisis de residuales.

5.10 Vining y Myers ("Combining Taguchi and Response Surface Philosophies: A Dual Response Approach", *Journal of Quality Technology*, **22**, 15-22, 1990) analizan un experimento que habían publicado Box y Draper (*Empirical Model Building and Response Surfaces*, 1988). En ese experimento se estudiaba el efecto de la rapidez (x_1), presión (x_2) y distancia (x_3) sobre la capacidad de una máquina impresora para aplicar tintas de color sobre etiquetas de paquetes. La siguiente tabla resume los resultados experimentales.

http://carlos2524.jimdo.com/

i	x_1	x_2	x_3	y_{i1}	y_{i2}	y_{i3}	$\bar{y}_i$	s_i
1	−1	−1	−1	34	10	28	24.0	12.5
2	0	−1	−1	115	116	130	120.3	8.4
3	1	−1	−1	192	186	263	213.7	42.8
4	−1	0	−1	82	88	88	86.0	3.7
5	0	0	−1	44	178	188	136.7	80.4
6	1	0	−1	322	350	350	340.7	16.2
7	−1	1	−1	141	110	86	112.3	27.6
8	0	1	−1	259	251	259	256.3	4.6
9	1	1	−1	290	280	245	271.7	23.6
10	−1	−1	0	81	81	81	81.0	0.0
11	0	−1	0	90	122	93	101.7	17.7
12	1	−1	0	319	376	376	357.0	32.9
13	−1	0	0	180	180	154	171.3	15.0
14	0	0	0	372	372	372	372.0	0.0
15	1	0	0	541	568	396	501.7	92.5
16	−1	1	0	288	192	312	264.0	63.5
17	0	1	0	432	336	513	427.0	88.6
18	1	1	0	713	725	754	730.7	21.1
19	−1	−1	1	364	99	199	220.7	133.8
20	0	−1	1	232	221	266	239.7	23.5
21	1	−1	1	408	415	443	422.0	18.5
22	−1	0	1	182	233	182	199.0	29.4
23	0	0	1	507	515	434	485.3	44.6
24	1	0	1	846	535	640	673.7	158.2
25	−1	1	1	236	126	168	176.7	55.5
26	0	1	1	660	440	403	501.0	138.9
27	1	1	1	878	991	1161	1010.0	142.5

a. Hacer un análisis detallado de residuales para estos datos.

b. Usar las varianzas muestrales como base para estimar los mínimos cuadrados ponderados para los datos originales (no los promedios muestrales).

c. Vining y Myers sugieren ajustar un modelo lineal a una transformación adecuada de las varianzas muestrales. Usar ese modelo para desarrollar los pesos adecuados y repetir la parte b.

5.11 Schubert, *et al.* ("The Catapult Problem: Enhanced Engineering Modeling Using Experimental Design", *Quality Engineering*, **4**, 463-473, 1992) hicieron un experimento con una catapulta para determinar los efectos del gancho (x_1), longitud de brazo (x_2), ángulo inicial (x_3) y el ángulo de tope (x_4) sobre la distancia a la que una catapulta lanza una bola. Para cada combinación de factores lanzaron la bola tres veces. La siguiente tabla resume los resultados experimentales.

http://carlos2524.jimdo.com/

x_1	x_2	x_3	x_4	y		
− 1	− 1	− 1	− 1	28.0	27.1	26.2
− 1	− 1	1	1	46.3	43.5	46.5
− 1	1	− 1	1	21.9	21.0	20.1
− 1	1	1	− 1	52.9	53.7	52.0
1	− 1	− 1	1	75.0	73.1	74.3
1	− 1	1	− 1	127.7	126.9	128.7
1	1	− 1	− 1	86.2	86.5	87.0
1	1	1	1	195.0	195.9	195.7

a. Hacer un análisis minucioso de residuales para estos datos.

b. Usar las varianzas muestrales como base para la estimación de los mínimos cuadrados ponderados para los datos originales (no la medias muestrales).

c. Ajustar un modelo adecuado a las varianzas muestrales (nota: ¡se requiere una transformación!). Usar este modelo para determinar los pesos adecuados y repetir la parte b.

5.12 Se tiene el modelo de regresión lineal simple $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, en el que la varianza de ε_i es proporcional a x_i^2; esto es, $\text{Var}(\varepsilon_i) = \sigma^2 x_i^2$.

a. Suponer que se usan las transformaciones $y' = y/x$ y que $x' = 1/x$. Esta transformación, ¿es estabilizadora de varianza?

b. ¿Cuáles son las relaciones entre los parámetros de los modelos original y transformados?

c. Supóngase que se usa el método de mínimos cuadrados ponderados con $w_i = 1/x_i^2$. ¿Es equivalente a la transformación introducida en la parte a?

5.13 Suponer que se quiere ajustar el modelo sin ordenada al origen $y = \beta x + \varepsilon$ con mínimos cuadrados ponderados. Suponer que las observaciones son no correlacionadas, pero que tienen varianzas desiguales.

a. Deducir una fórmula general para el estimador de β por mínimos cuadrados ponderados.

b. ¿Cuál es la varianza del estimador de β por mínimos cuadrados ponderados?

c. Suponer que $\text{Var}(y_i) = cx_i$, esto es, que la varianza de y_i es proporcional a la x_i correspondiente. Con los resultados de las partes a y b, determinar el estimador de β por mínimos cuadrados ponderados, y la varianza de este estimador.

d. Suponer que $\text{Var}(y_i) = cx_i^2$, esto es, que la varianza de y_i es proporcional al cuadrado de la x_i correspondiente. Con los resultados de las partes a y b, determinar el estimador de β por mínimos cuadrados y la varianza de ese estimador.

5.14 Se tiene el modelo

$$\mathbf{y} = \mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \varepsilon$$

siendo $E(\varepsilon) = \mathbf{0}$ y $\text{Var}(\varepsilon) = \sigma^2\mathbf{V}$. Suponer que σ^2 y $\mathbf{V}$ son conocidos. Deducir un estadístico adecuado de prueba de las hipótesis

$$H_0: \beta_2 = \mathbf{0}$$
$$H_1: \beta_2 \neq \mathbf{0}$$

Determinar la distribución de acuerdo con las hipótesis nula y alternativa.

http://carlos2524.jimdo.com/

5.15 Se tiene el modelo

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

en donde $E(\varepsilon) = \mathbf{0}$ y $\text{Var}(\varepsilon) = \sigma^2\,\mathbf{V}$. Suponer que se conoce $\mathbf{V}$, pero no σ^2. Demostrar que

$$[\mathbf{y}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{y}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}]/(n-p)$$

es un estimado insesgado de σ^2.

5.16 La tabla B.14 del apéndice contiene datos sobre los puntos transitorios de un inversor electrónico. Omitir la segunda observación, y usar $x_1 - x_4$ como regresores. Ajustar un modelo de regresión múltiple a estos datos.

a. Graficar los residuales ordinarios, los residuales estudentizados y los residuales R de Student en función de la respuesta predicha. Comentar los resultados.

b. Investigar la utilidad de una transformación sobre la variable de respuesta. ¿Mejora eso el modelo?

c. Además de una transformación de la respuesta, considérense transformaciones de los regresores. Usar gráficas de regresión parcial o de residuales parciales como ayuda en esta tarea.

http://carlos2524.jimdo.com/

CAPÍTULO 6

DIAGNÓSTICO PARA BALANCEO E INFLUENCIA

6.1 IMPORTANCIA DE DETECTAR OBSERVACIONES INFLUYENTES

Cuando se calcula el promedio de una muestra, cada observación en la muestra tiene el mismo peso en la determinación del resultado. Éste no es el caso en la regresión, por ejemplo, en la sección 2.7 se hizo notar que la ubicación de las observaciones en el espacio de x puede jugar un papel importante en la determinación de los coeficientes de regresión (véanse las Figs. 2.6 y 2.7). También se ha centrado la atención en los valores atípicos, u observaciones que tienen valores inusuales de y. En la sección 4.4 se dijo que los valores atípicos se identifican con frecuencia por los residuales desacostumbradamente grandes, y que esas observaciones también pueden afectar a los resultados de la regresión. El material en este capítulo es una extensión y una consolidación de algunos de esos asuntos.

Considérese el caso ilustrado en la figura 6.1. El punto identificado con A en esa figura está alejado, en el espacio de x, del resto de la muestra, pero está casi en la recta de regresión que pasa por el resto de los puntos de muestra. Es un ejemplo de un punto de **balanceo**, que es un valor desacostumbrado de x, y que puede controlar ciertas propiedades del modelo. Ahora bien, este punto no afecta las estimaciones de los coeficientes de regresión, pero ciertamente sí tiene un efecto marcado sobre las estadísticas de resumen del modelo, como R^2 y sobre los errores estándar de los coeficientes de regresión. Ahora véase el punto identificado con A en la figura 6.2. Este punto tiene una abscisa x moderadamente inusual, y el valor de y también es inusual. Éste es un punto de **influencia**, o **valor influyente**, esto es, que tiene un impacto notable sobre los coeficientes del modelo, porque "jala" al modelo de regresión en su dirección.

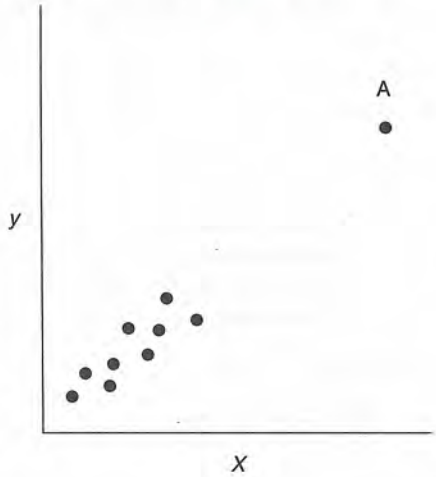


Figura 6.1 Ejemplo de un punto de balanceo.

http://carlos2524.jimdo.com/

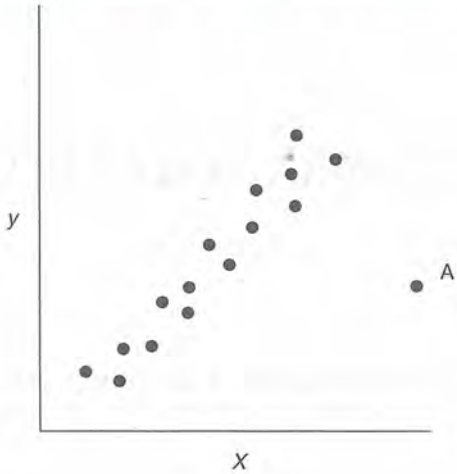


Figura 6.2 Ejemplo de una observación influyente.

A veces se ve que un pequeño subconjunto de los datos ejerce una influencia desproporcionada sobre los coeficientes y las propiedades del modelo. En un caso extremo, los estimados de parámetro pueden depender más del subconjunto influyente de puntos que de la mayor parte de los datos. Es obvio que este caso es indeseable; se prefiere un modelo de regresión que sea representativo de **todas** las observaciones en la muestra, y no sea un artefacto para unas cuantas. En consecuencia, se desea localizar esos puntos influyentes y evaluar su impacto sobre el modelo. Si esos puntos son en realidad valores "malos", se deberían eliminar de la muestra. Por otro lado, puede que no haya nada de malo en ellos, pero si controlan las propiedades clave del modelo, sería bueno conocerlo, porque podrían afectar el uso final del modelo de regresión.

En este capítulo se presentarán varios métodos de diagnóstico para balanceo e influencia. Estos diagnósticos están en la mayor parte de los programas de cómputo para regresión múltiple. Es importante usar esos diagnósticos junto con las técnicas de análisis de residuales que se describieron en el capítulo 4. A veces se encuentra que un coeficiente de regresión puede tener un signo que no tiene sentido desde el punto de vista científico o técnico, o un regresor que se sabe es importante, pero que puede ser insignificante estadísticamente, o un modelo que se ajusta bien a los datos, y que siendo lógico para la aplicación, puede producir malas predicciones. Estos casos pueden deberse a una o quizás a unas pocas observaciones influyentes. Por lo anterior, determinar esas observaciones puede arrojar bastante luz sobre los problemas con el modelo.

6.2 BALANCEO

Como se observó arriba, el lugar de los puntos en el espacio de x tiene importancia potencial en la determinación de las propiedades del modelo de regresión. En particular, los puntos alejados o remotos tienen un impacto desproporcionado sobre los estimados de los parámetros, los errores estándar, valores predichos y estadísticas de resumen del modelo. La matriz de sombrero

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{6.1}$$

http://carlos2524.jimdo.com/

desempeña un papel importante en la identificación de observaciones influyentes. Como se dijo antes, H determina las varianzas y covarianzas de $\hat{\mathbf{y}}$ y de **e**, porque $\text{Var}(\hat{\mathbf{y}}) = \sigma^2\mathbf{H}$ y $\text{Var}(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H})$. Los elementos h_{ij} de la matriz **H** pueden ser vistos como la cantidad de **balanceo** o palanqueo ejercido por la i-ésima observación y_i sobre el i-ésimo valor ajustado $\hat{y}_i$.

Con frecuencia, la atención se dirige hacia los **elementos diagonales** h_{ii} de la matriz **H** sombrero, que se pueden expresar como:

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \tag{6.2}$$

siendo $\mathbf{x}_i'$ el i-ésimo renglón de la matriz **X**. La diagonal de la matriz de sombrero es una medida estandarizada de la distancia de la i-ésima observación al centro (o al centroide) del espacio de x. Así, los elementos grandes en la diagonal indican observaciones que son potencialmente influyentes, por estar lejos, del resto de la muestra en el espacio de x. Sucede que el tamaño promedio de los elementos de la diagonal es $\bar{h} = p/n$ [porque $\Sigma_{i=1}^{n} h_{ii} = \text{rango}(\mathbf{H}) = \text{rango}(\mathbf{X}) = \text{p}$], y por tradición se supone que toda observación para la cual la diagonal del sombrero es más del doble del promedio $2p/n$ está suficientemente alejada del resto de los datos como para considerarse un **punto de balanceo**.

Nótese que todos los puntos de balanceo serán influyentes en los coeficientes de regresión. Por ejemplo, recuérdese el punto A de la figura 6.1. Este punto tendrá un elemento diagonal grande en la matriz de sombrero, y con seguridad será un punto de balanceo, pero casi no tiene efecto sobre los coeficientes de regresión por estar casi en la recta que pasa por las observaciones restantes. Como los elementos diagonales de la matriz de sombrero sólo examinan el lugar de la observación en el espacio de x, algunos analistas prefieren examinar los residuales estudentizados, o los residuales R de Student **junto con** las h_{ii}. Las observaciones con elementos diagonales grandes **y también** con residuales grandes probablemente serán influyentes. Por último, nótese que al usar el valor de corte $2p/n$ también se debe tener cuidado de evaluar las magnitudes tanto de p como de n. Habrá casos en los que $2p/n > 1$, y en esos casos no se aplica el corte.

Ejemplo 6.1 *Datos del tiempo de entrega*

La columna a de la tabla 6.1 muestra los elementos de la diagonal de la matriz de sombrero para el ejemplo 3.1, sobre los tiempos de entrega de gaseosas. Como $p = 3$ y $n = 25$, cualquier punto para el que la diagonal h_{ii} de la matriz de sombrero sea mayor de $2p/n = 2(3)/25 = 0.24$ es un punto de balanceo. Este criterio identifica que las observaciones 9 y 22 son puntos de balanceo. La ubicación lejana de esos puntos (en especial el punto 9), se notó antes ya, al examinar la matriz de diagramas de dispersión en la figura 3.4, y también cuando se ilustraron la interpolación y la extrapolación con este modelo en la figura 3.11.

En el ejemplo 4.1 se calcularon los residuales escalados para los datos de tiempo de entrega. La tabla 4.1 contiene los residuales estudentizados y los R de Student. Esos residuales no son demasiado grandes para la observación 22, e indican que es probable que su influencia sea pequeña sobre el modelo ajustado. Sin embargo, los dos residuales escalados para el punto 9 son moderadamente grandes y parecen indicar que esta observación puede tener una influencia moderada sobre el modelo. Para ilustrar el efecto de esos dos puntos sobre el modelo, se hicieron otros tres análisis: uno donde se eliminó la observación 9, un segundo sin la observación 22, y el tercero, eliminando las observaciones 9 y 22. Los resultados de esas corridas adicionales aparecen en la siguiente tabla:

http://carlos2524.jimdo.com/

Corrida	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	MS_{Res}	R^2
Con 9 y 22	2.341	1.616	0.014	10.624	0.9596
Sin 9	4.447	1.498	0.010	5.905	0.9487
Sin 22	1.916	1.786	0.012	10.066	0.9564
Sin 9 y sin 22	4.643	1.456	0.011	6.163	0.9072

La eliminación de la observación 9 sólo produce un cambio pequeño en $\hat{\beta}_1$, pero en $\hat{\beta}_2$ provoca un cambio aproximado de 28%, y un cambio de 90% en $\hat{\beta}_0$. Esto ilustra que la observación 9 está fuera del plano que pasa por los otros 24 puntos, y que ejerce una influencia moderadamente fuerte sobre el coeficiente de regresión asociado con x_2 (la distancia). No debe sorprender eso, si se considera que el valor de x_2, para esta observación (1 460 pies) es muy distinto de las demás observaciones. De hecho, la observación 9 puede estar originando curvatura en la dirección de x_2. Si se omitiera la observación 9, entonces MS_{Res} se reduciría a 5.905. Nótese que $\sqrt{5.905} = 2.430$, no muy distinto del estimado del error puro $\hat{\sigma} = 1.969$ que se calculó en el análisis de los vecinos cercanos en el ejemplo 4.11. Parece que la mayor parte de la falta de ajuste que se observa en este modelo, en el ejemplo 4.11, se debe al gran residual del punto 9. Si se omite el punto 22 se producen cambios relativamente pequeños en los coeficientes de regresión y en las estadísticas de resumen del modelo. Si se omiten los puntos 9 y 22 se producen cambios parecidos a cuando sólo se elimina el punto 9.

6.3 MEDIDAS DE INFLUENCIA: LA D DE COOK

En la sección anterior se dijo que es preferible tener en cuenta el lugar del punto en el espacio de x y también la variable de respuesta, al medir la influencia. Cook [1977, 1979] ha sugerido una forma de hacerlo, con una medida de la distancia, elevada al cuadrado, entre el estimado por mínimos cuadrados basado en los n puntos $\hat{\beta}$, y el estimado obtenido eliminando el i-ésimo punto, por ejemplo $\hat{\beta}_{(i)}$. Esta medida de la distancia se puede expresar como sigue, en forma general:

$$D_i(\mathbf{M}, c) = \frac{\left(\hat{\beta}_{(i)} - \hat{\beta}\right)'\mathbf{M}\left(\beta_{(i)} - \hat{\beta}\right)}{c}, \quad i = 1, 2, \ldots, n \tag{6.3}$$

Las opciones comunes de $\mathbf{M}$ y c son $\mathbf{M} = \mathbf{X}'\mathbf{X}$ y $c = pMS_{\text{Res}}$, por lo que la ecuación (6.3) se transforma en

$$D_i(\mathbf{X}'\mathbf{X}, pMS_{\text{Res}}) \equiv D_i = \frac{\left(\hat{\beta}_{(i)} - \hat{\beta}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\beta}_{(i)} - \hat{\beta}\right)}{pMS_{\text{Res}}}, \quad i = 1, 2, \ldots, n \tag{6.4}$$

Los puntos con grandes valores de D_i tienen gran influencia sobre el estimado de $\hat{\beta}$ por mínimos cuadrados.

La magnitud de D_i se suele evaluar comparándola con $F_{\alpha, p, n-p}$. Si $D_i = F_{0.5, p, n-p}$, entonces al eliminar el punto i se movería $\hat{\beta}_{(i)}$ hacia la frontera de una región de confianza

http://carlos2524.jimdo.com/

Tabla 6.1 Estadísticas para detectar observaciones influyentes en los datos de tiempo de entrega de bebidas gaseosas

Observación i	(a) h_{ii}	(b) D_i	(c) $DFFITS_i$	(d) Ordenada al origen $DFBETAS_{(0),i}$	(e) Cajas $DFBETAS_{1,i}$	(f) Distancia $DFBETAS_{2,i}$	(e) $COVRATIO_i$
1	0.10180	0.10009	−0.5709	−0.1873	0.4113	−0.4349	0.8711
2	0.07070	0.00338	0.0986	0.0898	−0.0478	0.0144	1.2149
3	0.09874	0.00001	−0.0052	−0.0035	0.0039	−0.0028	1.2757
4	0.08538	0.07766	0.5008	0.4520	0.0883	−0.2734	0.8760
5	0.07501	0.00054	−0.0395	−0.0317	−0.0133	0.0242	1.2396
6	0.04287	0.00012	−0.0188	−0.0147	0.0018	0.0011	1.1999
7	0.08180	0.00217	0.0790	0.0781	−0.0223	−0.0110	1.2398
8	0.06373	0.00305	0.0938	0.0712	0.0334	−0.0538	1.2056
9	0.49829	3.41835	4.2961	−2.5757	0.9287	1.5076	0.3422
10	0.19630	0.05385	0.3987	0.1079	−0.3382	0.3413	1.3054
11	0.08613	0.01620	0.2180	−0.0343	0.0925	−0.0027	1.1717
12	0.11366	0.00160	−0.0677	−0.0303	−0.0487	0.0540	1.2906
13	0.06113	0.00229	0.0813	0.0724	−0.0356	0.0113	1.2070
14	0.07824	0.00329	0.0974	0.0495	−0.0671	0.0618	1.2277
15	0.04111	0.00063	0.0426	0.0223	−0.0048	0.0068	1.1918
16	0.16594	0.00329	−0.0972	−0.0027	0.0644	−0.0842	1.3692
17	0.05943	0.00040	0.0339	0.0289	0.0065	−0.0157	1.2192
18	0.09626	0.04398	0.3653	0.2486	0.1897	−0.2724	1.0692
19	0.09645	0.01192	0.1862	0.1726	0.0236	−0.0990	1.2153
20	0.10169	0.13246	−0.6718	0.1680	−0.2150	−0.0929	0.7598
21	0.16528	0.05086	−0.3885	−0.1619	−0.2972	0.3364	1.2377
22	0.39158	0.45106	−1.1950	0.3986	−1.0254	0.5731	1.3981
23	0.04126	0.02990	−0.3075	−0.1599	0.0373	−0.0527	0.8897
24	0.12061	0.10232	−0.5711	−0.1197	0.4046	−0.4654	0.9476
25	0.06664	0.00011	−0.0176	−0.0168	0.0008	0.0056	1.2311

http://carlos2524.jimdo.com/

aproximada de 50% para β, basándose en el conjunto completo de datos. Es un desplazamiento grande e indica que el estimado por mínimos cuadrados es sensible al i-ésimo punto de datos. Como $F_{0.5,\,p,\,n-p} \simeq 1$, se suelen considerar como influyentes los puntos para los que $D_i > 1$. En el caso ideal sería bueno que cada estimado $\hat{\beta}_{(i)}$ permaneciera dentro de los límites de la región de confianza de 10 o de 20%. Esta recomendación de corte se basa en la semejanza de D_i con la ecuación del elipsoide de confianza de la teoría normal [Ec. (3.50)]. La medida de distancia D_i no es una estadística F. Sin embargo, usar el corte igual a una unidad funciona muy bien en la práctica.

La estadística D_i se puede reexpresar como sigue:

$$D_i = \frac{r_i^2}{p}\frac{\mathrm{Var}(\hat{y}_i)}{\mathrm{Var}(e_i)} = \frac{r_i^2}{p}\frac{h_{ii}}{(1-h_{ii})}, \quad i = 1, 2, \ldots, n \tag{6.5}$$

Así se ve que, además de la constante p, la D_i es el producto del cuadrado del i-ésimo residual estudentizado por $h_{ii}/(1-h_{ii})$. Se puede demostrar que esta relación es la distancia del vector $\mathbf{x}_i$ al centroide de los datos restantes. Así, D_i está formada por un componente que refleja lo bien que se ajusta el modelo a la i-ésima observación y_i, y un componente que mide lo alejado que el punto está del resto de los datos. Cualquiera de los componentes (o ambos), pueden contribuir a un valor grande de D_i. Así, en D_i se combinan la magnitud del residual para la i-ésima observación y la ubicación de ese punto en el espacio de x, para evaluar su influencia.

Ya que $\mathbf{X}\hat{\beta}_{(i)} - \mathbf{X}\hat{\beta} = \hat{\mathbf{y}}_{(i)} - \hat{\mathbf{y}}$, otra forma de expresar la medida de distancia de Cook es

$$D_i = \frac{(\hat{\mathbf{y}}_{(i)} - \hat{\mathbf{y}})'(\hat{\mathbf{y}}_{(i)} - \hat{\mathbf{y}})}{pMS_{\mathrm{Res}}} \tag{6.6}$$

Así, otra forma de interpretar la distancia de Cook es el cuadrado de la distancia euclidiana (sin considerar pMS_{Res}) que se mueve el vector de los valores ajustados cuando se elimina la i-ésima observación.

Ejemplo 6.2 *Datos del tiempo de entrega*

La columna b de la tabla 6.1 contiene los valores de la medida de distancia de Cook para los datos del tiempo de entrega de bebidas gaseosas. Se ilustrarán los cálculos con la primera observación. Los residuales estudentizados para los datos de tiempo de entrega están en la tabla 4.1, y $r_1 = -1.6277$. Así, el valor de D_1 es

$$D_i = \frac{r_i^2}{p}\frac{h_{11}}{(1-h_{11})}$$

$$= \frac{(-1.6277)^2}{3}\frac{0.10180}{(1-0.10180)} = 0.10009$$

El valor máximo de la estadística D_i es $D_9 = 3.41835$, que indica que al eliminar la observación 9 se movería el estimado por mínimos cuadrados hasta, aproximadamente, el límite de la región de confianza de 96% respecto a $\hat{\beta}$. El siguiente valor máximo es $D_{22} = 0.45106$, y la eliminación del punto 22 movería el estimado de $\hat{\beta}$ hasta más o menos la orilla de una región de confianza de 35%. En consecuencia, las conclusiones serían que la observación 9 es influyente en definitiva, al usar el valor de corte igual a la unidad, y

http://carlos2524.jimdo.com/

que la observación 22 no es influyente. Obsérvese que estas conclusiones concuerdan bastante bien con las alcanzadas en el ejemplo 6.1, al examinar por separado las diagonales de la matriz sombrero y los residuales estudentizados.

6.4 MEDIDAS DE INFLUENCIA: *DFFITS* Y *DFBETAS*

La medida de distancia de Cook es un **diagnóstico de eliminación**; es decir, mide la influencia de la i-ésima observación si se eliminara de la muestra. Belsley, Kuh y Welsch [1980] introdujeron otras dos medidas útiles de eliminación de influencia. La primera es una estadística que indica cuánto cambia el coeficiente de regresión $\hat{\beta}_j$, en unidades de desviación estándar, si se omitiera la i-ésima observación. Esta estadística es

$$DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 C_{jj}}} \tag{6.7}$$

siendo C_{jj} el j-ésimo elemento diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$, y $\hat{\beta}_{j(i)}$ el j-ésimo coeficiente de regresión, calculado sin usar la i-ésima observación. Un valor (de magnitud) grande de $DFBETAS_{j,i}$ indica que la observación i tiene gran influencia sobre el j-ésimo coeficiente de regresión. Obsérvese que $DFBETAS_{j,i}$ es una matriz de $n \times p$ que contiene información parecida a la de la medida de distancia de Cook.

Es interesante el cálculo de $DFBETAS_{j,i}$. Se define la matriz de $p \times n$

$$\mathbf{R} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Los n elementos del j-ésimo renglón de $\mathbf{R}$ producen el balanceo que las n observaciones de la muestra tienen sobre $\hat{\beta}_j$. Si se define a $\mathbf{r}_j'$ como el j-ésimo renglón de $\mathbf{R}$, se puede demostrar (véase el Ap. C.12) que

$$\begin{aligned} DFBETAS_{j,i} &= \frac{r_{j,i}}{\sqrt{\mathbf{r}_j\mathbf{r}_j}} \frac{e_i}{S_{(i)}(1 - h_{ii})} \\ &= \frac{r_{j,i}}{\sqrt{\mathbf{r}_j\mathbf{r}_j}} \frac{t_i}{\sqrt{1 - h_{ii}}} \end{aligned} \tag{6.8}$$

en donde t_i es el residual de R de Student. Nótese que $DFBETAS_{j,i}$ mide tanto el balanceo ($r_{j,i}/\sqrt{\mathbf{r}_j\mathbf{r}_j}$ es una medida del impacto de la i-ésima observación sobre $\hat{\beta}_j$) como el efecto de un residual grande. Belsley, Kuh y Welsch [1980] sugieren un valor de corte igual a $2/\sqrt{n}$ para $DFBETAS_{j,i}$; esto es, si $|DFBETAS_{j,i}| > 2/\sqrt{n}$, es necesario examinar la i-ésima observación.

También se puede investigar la influencia de la eliminación de la i-ésima observación sobre el valor predicho o ajustado. Esto conduce al segundo diagnóstico propuesto por Belsley, Kuh y Welsch:

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{S_{(i)}^2 h_{ii}}}, \quad i = 1, 2, \ldots, n \tag{6.9}$$

http://carlos2524.jimdo.com/

en la que $\hat{y}_{(i)}$ es el valor ajustado de y_i, obtenido sin usar la i-ésima observación. El denominador no es más que una estandarización, porque $\text{Var}(\hat{y}_i) = \sigma^2 h_{ii}$. Así, $DFFITS_i$ es la cantidad de desviaciones estándar que **cambia** el valor ajustado $\hat{y}_i$ si se elimina la observación i.

Para los cálculos se puede usar (véanse los detalles en el Ap. C.12):

$$DFFITS_i = \left(\frac{h_{ii}}{1 - h_{ii}}\right)^{1/2} \frac{e_i}{S_{(i)}(1 - h_{ii})^{1/2}}$$

$$= \left(\frac{h_{ii}}{1 - h_{ii}}\right)^{1/2} t_i \tag{6.10}$$

donde t_i es el residual R de Student. Así, $DFFITS_i$ es el valor del R de Student multiplicado por el balanceo de la i-ésima observación $[h_{ii}/(1 - h_{ii})]^{1/2}$. Si el dato es atípico, el residual R de Student tendrá magnitud grande, mientras que si el dato tiene gran balanceo, h_{ii} se aproximará a la unidad. En cualquiera de esos casos, $DFFITS_i$ puede ser grande. Sin embargo, si $h_{ii} \simeq 0$, el efecto del R de Student será moderado. En forma parecida, un residual R de Student cercano a cero, combinado con un punto de gran balanceo podría producir un valor pequeño de $DFFITS_i$. Así se ve que $DFFITS_i$ se puede afectar ya sea por el balanceo como por el error de predicción. Belsley, Kuh y Welsch sugieren que merece investigarse toda observación para $|DFFITS_i| > 2\sqrt{p/n}$.

Observación acerca de los valores de corte

En esta sección hemos mencionado valores de corte recomendados para $DFFITS_i$ y para $DFBETAS_{j,i}$. Recuérdese que esas recomendaciones sólo son lineamientos, porque es muy difícil producir valores de corte que sean correctos en todos los casos. Así, recomendamos al analista que use información sobre qué significa el diagnóstico, y sobre el ambiente de aplicación para seleccionar un valor de corte. Por ejemplo, si $DFFITS_i = 1.0$, se podría traducir en unidades reales de respuesta para determinar cuánto se afecta $\hat{y}_i$ al eliminar la i-ésima observación. Entonces $DFBETAS_{j,i}$ se podría usar para ver si esa observación es responsable del significado (o quizá el no significado) de determinados coeficientes, o de cambios de signo en un coeficiente de regresión. El diagnóstico con $DFBETAS_{j,i}$ también se puede usar para determinar (usando el error estándar del coeficiente) cuánto cambio origina, en **unidades específicas del problema** un dato sobre el coeficiente de regresión. A veces esos cambios serán importantes en un **contexto específico del problema** aun cuando las estadísticas de diagnóstico no rebasen el valor formal de corte.

Obsérvese que el valor de corte recomendado es función de n, el tamaño de la muestra. Tenemos la firme convicción de que cualquier valor corte formal debe ser una función del tamaño de la muestra; sin embargo, de acuerdo con nuestra experiencia, esos valores de corte suelen identificar más datos de los que un analista desease analizar. Esto es válido en especial con muestras pequeñas. Creemos que los valores de corte recomendados por Belsley, Kuh y Welsch tienen sentido para muestras grandes, pero cuando n es pequeña preferimos la forma de diagnóstico que describimos antes.

http://carlos2524.jimdo.com/

Ejemplo 6.3 *Datos del tiempo de entrega*

Las columnas c a f de la tabla 6.1 muestran los valores de $DFFITS_i$ y $DFBETAS_{j,i}$ para los datos de tiempo de entrega de bebidas gaseosas. El valor formal de corte para $DFFITS_i$ es $2\sqrt{p/n} = 2\sqrt{3/25} = 0.69$. Al inspeccionar la tabla 6.1 se aprecia que los puntos 9 y 22 tienen valores $DFFITS_i$ mayores que 0.69, y que además $DFFITS_{20}$ se acerca al valor de corte.

Al examinar los $DFBETAS_{j,i}$, recordando que el valor de corte es $2/\sqrt{25} = 0.40$, se nota de inmediato que los puntos 9 y 22 tienen efectos grandes sobre los tres parámetros. El punto 9 tiene un efecto muy grande sobre la ordenada al origen, y menores efectos sobre $\hat{\beta}_1$ y $\hat{\beta}_2$, mientras que el punto 22 tiene el máximo efecto sobre $\hat{\beta}_1$. Hay algunos otros puntos que producen efectos sobre los coeficientes, que se acercan al corte formal, incluyendo al 1 (sobre $\hat{\beta}_1$ y $\hat{\beta}_2$), el 4 (sobre $\hat{\beta}_0$) y el 24 (sobre $\hat{\beta}_1$ y $\hat{\beta}_2$). Esos puntos producen cambios relativamente pequeños en comparación con el punto 9.

Si se adopta un punto de vista de diagnóstico, es claro que el punto 9 es influyente, porque su eliminación da como resultado un desplazamiento de cada coeficiente de regresión de 0.9 desviaciones estándar, como mínimo. El efecto del punto 22 es mucho menor. Además, al eliminar el punto 9 se desplaza la respuesta predicha en más de cuatro desviaciones estándar. De nuevo se cuenta con una señal clara de que la observación 9 es influyente.

6.5 UNA MEDIDA DEL DESEMPEÑO DEL MODELO

Los diagnósticos D_i, $DFBETAS_{j,i}$, y $DFFITS_i$ permiten conocer el efecto de las observaciones sobre los coeficientes estimados $\hat{\beta}_j$ y los valores ajustados $\hat{y}_i$. No proporcionan información alguna sobre la **precisión general de la estimación**. En vista de que se acostumbra mucho usar el determinante de la matriz de covarianza como una medida escalar cómoda de la precisión, llamada **varianza generalizada**, se podría definir la varianza generalizada (GV, de *generalized variance*) de $\hat{\beta}$ como sigue:

$$\mathrm{GV}(\hat{\beta}) = |\mathrm{Var}(\hat{\beta})| = |\sigma^2(\mathbf{X}'\mathbf{X})^{-1}|$$

Para expresar el papel de la i-ésima observación en la precisión de la estimación se puede definir la relación de covarianzas ($COVRATIO_i$, de *covariance ratio*):

$$COVRATIO_i = \frac{|(\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1}\ S^2_{(i)}|}{|(\mathbf{X}'\mathbf{X})^{-1}\ MS_{\mathrm{Res}}|}, \quad i = 1, 2, \ldots, n \tag{6.11}$$

Es claro que si $COVRATIO_i > 1$, la i-ésima observación mejora la precisión de la estimación, mientras que si $COVRATIO_i < 1$, la inclusión del i-ésimo punto degrada la precisión. Para el cálculo se usa

$$COVRATIO_i = \frac{\left(S^2_{(i)}\right)^p}{MS^p_{\mathrm{Res}}}\left(\frac{1}{1 - h_{ii}}\right) \tag{6.12}$$

Nótese que $[1/(1 - h_{ii})]$ es la relación de $|(\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1}|$ entre $|(\mathbf{X}'\mathbf{X})^{-1}|$, por lo que un punto de alto balanceo hará que $COVRATIO_i$ sea grande. Esto es lógico, porque un punto de alto

http://carlos2524.jimdo.com/

balanceo mejorará la precisión, a menos que ese punto sea un valor atípico en el espacio de y. Si la i-ésima observación es atípica, $S_{(i)}^2/MS_{\text{Res}}$ será mucho menor que la unidad.

No es fácil obtener valores de corte para *COVRATIO*. Belsley, Kuh y Welsch [1980] sugieren que si $COVRATIO_i > 1 + 3p/n$, o si $COVRATIO_i < 1 - 3p/n$, se debería considerar al i-ésimo punto como influyente. La cota inferior no es adecuada cuando $n > 3p$. Estos valores de cortes sólo se recomiendan para muestras grandes.

Ejemplo 6.4 *Datos del tiempo de entrega*

La columna g de la tabla 6.1 contiene los valores de $COVRATIO_i$, para los datos de tiempo de entrega de bebidas gaseosas. El valor de corte formal recomendado para $COVRATIO_i$ es $1 \pm 3p/n = 1 \pm 3(3)/25$, es decir, 0.64 y 1.36. Nótese que los valores de $COVRATIO_9$ y $COVRATIO_{22}$ se salen de estos límites, lo que indica que esos puntos son influyentes. Como $COVRATIO_9 < 1$, esta observación degrada la precisión de estimación, mientras que $COVRATIO_{22} > 1$, por lo que esta observación tiende a mejorar la precisión. Sin embargo, el punto 22 apenas se sale del valor de corte, por lo que desde el punto de vista práctico es bastante pequeña la influencia de esta observación. El punto 9 es influyente con mucho más claridad.

6.6 DETECCIÓN DE GRUPOS DE OBSERVACIONES INFLUYENTES

Se ha descrito aquí principalmente el diagnóstico para eliminación de una observación, por influencia y balanceo. Es obvio que pudieran haber casos en donde un **grupo** de puntos tengan alto balanceo o ejerzan demasiada influencia sobre el modelo de regresión. En Belsley, Kuh y Welsch [1980], Cook y Weisberg [1980] y Rousseeuw y Leroy [1987] se encuentran descripciones muy buenas de este problema.

En principio se puede extender el diagnóstico de un punto al caso de varias observaciones. De hecho, hay varias estrategias para resolver el problema de varias observaciones influyentes o atípicas. Por ejemplo, véase Atkinson[1994] Hadi y Simonoff [1993], Hawkings, Bradu y Kass [1984], Pena y Yohai [1995] y Rousseeuw y van Zomeren [1990]. Para mostrar cómo se podría extender la medida de distancia de Cook para evaluar la influencia simultánea de un grupo de m observaciones, sea $\mathbf{i}$ el vector de $m \times 1$, de los índices que especifican los puntos a eliminar, y se define

$$D_{\mathbf{i}}(\mathbf{X}'\mathbf{X}, pMS_{\text{Res}}) = D_{\mathbf{i}} = \frac{\left(\hat{\boldsymbol{\beta}}_{(\mathbf{i})} - \hat{\boldsymbol{\beta}}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\boldsymbol{\beta}}_{(\mathbf{i})} - \hat{\boldsymbol{\beta}}\right)}{pMS_{\text{Res}}}$$

Es obvio que Di es una versión de la medida de distancia de Cook para observaciones múltiples. La interpretación de $D_{\mathbf{i}}$ se parece a la estadística de una sola observación. Los valores grandes de $D_{\mathbf{i}}$ indican que el conjunto de m puntos es influyente. Sin embargo, no es obvia la selección del subconjunto de puntos que se incluyen en m, porque en algunos conjuntos de datos los subconjuntos de puntos son **influyentes en conjunto**, pero los puntos individuales no lo son. Además no es práctico investigar todas las combinaciones posibles de los n puntos de la muestra al tomar $m = 1, 2, \ldots, n$ puntos cada vez.

http://carlos2524.jimdo.com/

Sebert, Montgomery y Rollier [1998] investigan el uso de **análisis de conglomerados** para determinar el conjunto de observaciones influyentes en regresión. El análisis de conglomerado es una técnica de varias variables para determinar grupos de observaciones semejantes. El procedimiento consiste en definir una medida de similaridad entre las observaciones para después aplicar una serie de reglas y clasificar las observaciones en grupos, con base entre sus similaridades interobservacionales. Usan un procedimiento de conglomerado de un solo eslabón (véase Johnson y Wichern [1992] y Everitt [1993]) aplicado a los residuales y a los valores ajustados por mínimos cuadrados para agrupar $n - m$ observaciones en un grupo "limpio" y un grupo potencialmente influyente de m observaciones. Este grupo se evalúa en subconjuntos de tamaño $1, 2, \ldots, m$, con la versión de la medida de distancia de Cook para observaciones múltiples. Los autores informan que este procedimiento es muy efectivo para determinar el subconjunto de observaciones influyentes. Hay cierto "exceso", esto es identificar demasiadas observaciones como influyentes, pero el empleo de la distancia de Cook elimina con eficiencia las observaciones no influyentes. Al estudiar nueve conjuntos de datos publicados, los autores inform que no hubo incidentes de "encubrimiento", esto es, no poder determinar el subconju correcto de puntos influyentes. También informan buenos resultados con un extenso est dio de desempeño, hecho con simulación Monte Carlo.

6.7 TRATAMIENTO DE LAS OBSERVACIONES INFLUYENTES

Los diagnósticos de balanceo e influencia son parte importante del arsenal de herramientas de quien construye modelos. Pretenden ofrecer al analista mejor comprensión de los datos y señalar cuáles observaciones merecen más escrutinio. ¿Cuánto esfuerzo se debe dedicar al estudio de esos puntos? Es probable que dependa de la cantidad de puntos influyentes identificados, de su impacto real sobre el modelo y de la impo ancia del problema de construcción del modelo. Si el lector dedicó un año a reunir 30 observaciones, es probable que se justifique mucho análisis de seguimiento para los puntos dudosos. Esto es válido en especial si se obtiene un resultado inesperado debido a una sola observación influyente.

¿Se deben desechar las observaciones influyentes? Esta pregunta se parece a la cuestión de descartar los valores atípicos. Por regla general, si hay un error al anotar un valor medido, o si el punto de la muestra realmente es inválido o no es parte de la población que se pretendía muestrear, será adecuado descartar la observación. Sin embargo, si el análisis indica que un punto influyente es una observación válida, no hay justificación para su eliminación.

Un "compromiso" entre eliminar una observación y retenerla es considerar una técnica de estimación que no sea tan sensible a los puntos influyentes como lo son los mínimos cuadrados. Esas técnicas **robustas** de estimación en esencia **aligeran** las observaciones en proporción con la magnitud o influencia residual, de tal modo que una observación muy influyente recibirá menos peso que en el ajuste de mínimos cuadrados. En el capítulo 11 se describirán métodos robustos de regresión.

PROBLEMAS

6.1 Hacer un análisis minucioso de influencia con los datos de pruebas de energía térmica que aparecen en la tabla B.2 del apéndice. Comentar los resultados.

http://carlos2524.jimdo.com/

6.2 Hacer un análisis minucioso de influencia con los datos de valuación de la propiedad que aparecen en la tabla B.4 del apéndice. Comentar los resultados.

6.3 Hacer un análisis minucioso de influencia con las corridas de licuefacción de Belle Ayr que aparecen en la tabla B.5 del apéndice. Comentar los resultados.

6.4 Hacer un análisis minucioso de influencia con los datos del reactor de tubo de flujo que aparecen en la tabla B.6 del apéndice. Comentar los resultados.

6.5 Hacer un análisis detallado de influencia con los datos de desempeño del equipo de la NFL que aparecen en la tabla B.1 del apéndice. Comentar los resultados.

6.6 Hacer un análisis minucioso de influencia con los datos de extracción de aceite que aparecen en la tabla B.7 del apéndice. Comentar los resultados.

6.7 Hacer un análisis minucioso de influencia con los datos de formación de catrato que aparecen en la tabla B.8 del apéndice. Hacer las transformaciones que sean necesarios. Comentar los resultados.

6.8 Hacer un análisis minucioso de influencia con los datos de caída de presión que aparecen en la tabla B.9 del apéndice. Hacer las transformaciones adecuadas. Comentar los resultados.

6.9 Hacer un análisis minucioso de influencia con los datos de viscosidad cinemática que aparecen en la tabla B.10 del apéndice. Hacer las transformaciones necesarias. Comentar los resultados.

6.10 Demostrar formalmente que

$$D_i = \frac{r_i}{p} \frac{h_{ii}}{(1 - h_{ii})}$$

6.11 Demostrar formalmente que

$$COVRATIO_i = \left[\frac{S_{(i)}^2}{MS_{\text{Res}}} \right]^p \left(\frac{1}{(1 - h_{ii})} \right)$$

6.12 La tabla B.11 del apéndice contiene datos de la calidad del vino Pinot Noir. Ajustar un modelo de regresión usando claridad, aroma, cuerpo, sabor y perfume como regresores. Investigar si en este modelo hay observaciones influyentes y comentar lo encontrado.

6.13 La tabla B.12 del apéndice contiene datos reunidos en el tratamiento térmico de engranajes. Ajustar un modelo de regresión a esos datos usando todos los regresores. Investigar si en este modelo hay observaciones influyentes y comentar lo encontrado.

6.14 La tabla B.13 del apéndice contiene datos del empuje de un motor de turbina de reacción. Ajustar un modelo de regresión a esos datos usando todos los regresores. Investigar si en este modelo hay observaciones influyentes y comentar lo encontrado.

6.15 La tabla B.14 del apéndice contiene datos de los puntos transitorios de inversor electrónico. Ajustar un modelo de regresión a las 25 observaciones, pero sólo usar de x_1 a x_4 como regresores. Investigar si en este modelo hay observaciones influyentes y comentar lo encontrado.

http://carlos2524.jimdo.com/

CAPÍTULO 7

MODELOS POLINOMIALES DE REGRESIÓN

7.1 INTRODUCCIÓN

El modelo de regresión lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ es un modelo general de ajuste de toda relación que sea lineal en los parámetros desconocidos β. Entre las relaciones está incluida la clase importante de los **modelos polinomiales de regresión**. Por ejemplo, el polinomio de segundo orden de una variable

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

y el polinomio de segundo orden de dos variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$$

son modelos de regresión lineal.

Los polinomios se usan mucho en casos en los que la respuesta es curvilínea, y aun las relaciones no lineales complejas se pueden modelar en forma adecuada con polinomios dentro de límites razonablemente pequeños de las x. En este capítulo se describirán varios problemas y asuntos relacionados con el ajuste de polinomios.

7.2 MODELOS POLINOMIALES EN UNA VARIABLE

7.2.1 Principios básicos

Como ejemplo de un **modelo polinomial de regresión** se considerará

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \tag{7.1}$$

Este modelo se llama **modelo de segundo orden** en una variable. También a veces se le llama **modelo cuadrático**, porque el valor esperado de y es

$$E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$$

lo cual describe una función cuadrática. Un ejemplo típico se ve en la figura 7.1. Con frecuencia, a β_1 se le llama parámetro de efecto lineal y a β_2 parámetro de efecto cuadrático. El parámetro β_0 es el promedio de y cuando $x = 0$, si el rango de los datos incluye a $x = 0$. En caso contrario, β_0 no tiene interpretación física.

http://carlos2524.jimdo.com/

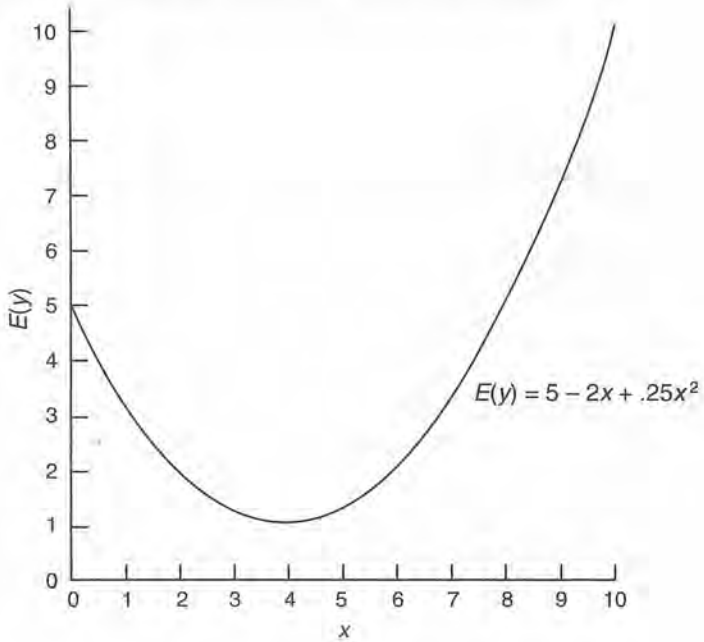


Figura 7.1 Un ejemplo de polinomio cuadrático.

En general, el modelo polinomial de k-ésimo orden en una variable es

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \varepsilon \tag{7.2}$$

Si se define $x_j = x^j$, $j = 1, 2, \ldots, k$, la ecuación (7.2) se transforma en un modelo de regresión lineal múltiple con los k regresores $x_1, x_2, \ldots, x_k$. Así, un modelo polinomial de orden k se puede ajustar con las técnicas que ya se estudiaron.

Los modelos polinomiales son útiles en casos cuando el analista sabe que hay efectos curvilíneos presentes en la función verdadera de respuesta. También son útiles como funciones de aproximación a relaciones no lineales, desconocidas y posiblemente muy complejas. En este sentido, el modelo polinomial es simplemente el desarrollo en serie de Taylor de la función desconocida. Esta clase de aplicaciones parece presentarse con la mayor frecuencia en la práctica.

Hay varias consideraciones importantes que se presentan cuando se ajusta un polinomio de una variable. Algunas de ellas se describirán a continuación.

1. Orden del modelo Es importante mantener **tan bajo como sea posible** el orden del modelo. Cuando la función de respuesta parezca ser curvilínea, se deben intentar **transformaciones** para mantener al modelo como de primer orden. Los métodos que se describieron en el capítulo 5 son útiles en este aspecto, si falla lo anterior, se debe intentar un polinomio de segundo orden. Como regla general, se debe evitar el uso de polinomios de orden superior (con $k > 2$), a menos que se puedan justificar por razones ajenas a los datos. Un modelo de orden menor en una variable transformada casi siempre es preferible a un modelo de orden superior en la métrica original. El ajuste arbitrario de polinomios de orden superior es un grave abuso del análisis de regresión. Siempre se debe mantener un sentido de **parsimonia**, esto es, se debe usar el modelo más simple posible que sea consistente con los datos y el conocimiento del ambiente del problema. Recuérdese que en un caso extremo siempre es posible hacer pasar un polinomio de orden $n - 1$ por

http://carlos2524.jimdo.com/

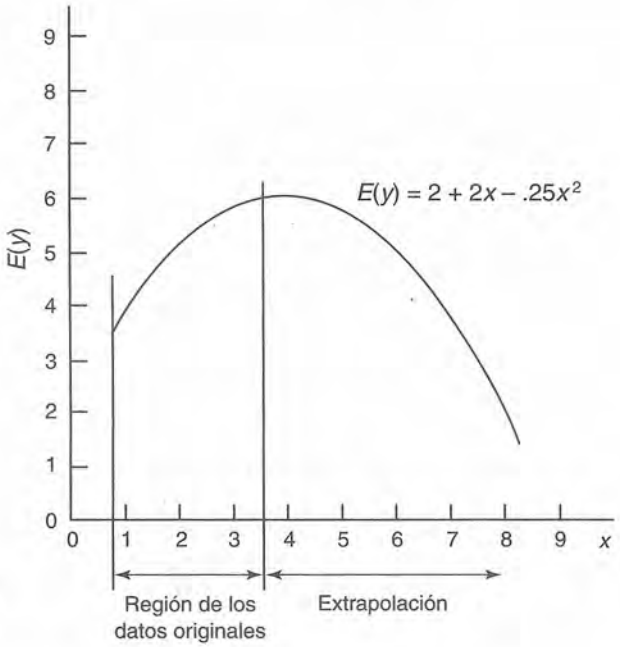


Figura 7.2 Peligro de extrapolar.

n puntos, por lo que siempre se puede encontrar un polinomio con grado suficientemente alto que produzca un ajuste "bueno" con los datos. Ese modelo no contribuiría a mejorar el conocimiento de la función desconocida, ni es probable que sea buen predictor.

2. Estrategia para la construcción del modelo Se han sugerido diversas estrategias para elegir el orden de un polinomio de aproximación. Un método es ajustar en forma sucesiva modelos de orden creciente hasta que la prueba t para el término de orden máximo sea no significativa. Un procedimiento alterno es ajustar el modelo de orden máximo adecuado, y a continuación eliminar términos, uno por uno, comenzando con el de orden máximo hasta que el término que quede de orden máximo tenga una estadística t significativa. Esos dos procedimientos se llaman **selección en avance** y **eliminación en reversa**, respectivamente, no necesariamente conducen al mismo modelo. En vista del comentario del punto 1, se deben usar con cuidado esos procedimientos. En la mayor parte de los casos se debería restringir la atención a polinomios de primero y segundo orden.

3. Extrapolación La extrapolación con modelos polinomiales puede ser peligrosa en extremo, por ejemplo, véase el modelo de segundo orden en la figura 7.2, si se extrapola más allá del rango de los datos originales, la respuesta predicha se va hacia abajo; esto puede ser contrario al comportamiento real del sistema, En general, los modelos polinomiales pueden dirigirse hacia direcciones imprevistas e inadecuadas, tanto en la interpolación como en la extrapolación.

4. Mal acondicionamiento I A medida que aumenta el orden del polinomio, la matriz $\mathbf{X}'\mathbf{X}$ se vuelve **mal acondicionada**. Esto quiere decir que los cálculos de inversión de matriz serán inexactos y se puede introducir error considerable en los estimados de los parámetros, véase, por ejemplo, Forsythe [1957]. El **mal acondicionamiento no esencial** causado por la elección arbitraria del origen se puede eliminar, **centrando** primero las **variables regresoras**, es decir, corregir x por su promedio $\bar{x}$; pero como hacen notar Bradley y Srivastava [1979], centrando los datos todavía se pueden obtener grandes correlaciones

http://carlos2524.jimdo.com/

muestrales entre ciertos coeficientes de regresión. En la sección 7.4 se describirá un método para tratar este problema.

5. Mal acondicionamiento II Si los valores de x se limitan a un rango estrecho, puede haber mal acondicionamiento o multicolinealidad apreciables en las columnas de la matriz $\mathbf{X}$. Por ejemplo, si x varía entre 1 y 2, entonces x^2 varía entre 1 y 4, lo cual podría crear una fuerte multicolinealidad entre x y x^2.

6. Jerarquía El modelo de regresión

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$$

se llama **jerárquico** porque contiene todos los términos de orden tres y menores. En contraste, el modelo

$$y = \beta_0 + \beta_1 x + \beta_3 x^3 + \varepsilon$$

no es jerárquico. Peixoto [1987, 1990] hace notar que sólo los modelos jerárquicos son invariantes a la transformación lineal, y sugiere que todos los modelos polinomiales deben tener esta propiedad (con frecuencia se usa la frase "un modelo jerárquicamente bien formulado"). Nosotros tenemos ideas encontradas acerca de esto. Es cierto que es atractivo que la forma del modelo se preserve después de una transformación lineal (como ajustar el modelo con variables codificadas y convertirlo después a un modelo con las variables naturales), pero eso no es más que belleza matemática. Hay muchos modelos mecanísticos que no son jerárquicos; por ejemplo, la ley de Newton de la gravedad es de cuadrado inverso, y la ley del dipolo magnético es de cubo inverso. Además, hay muchos casos en que se usa un modelo polinomial de regresión para representar los resultados de un experimento diseñado, en el que los datos respaldarían un modelo como:

$$y = \beta_0 + \beta_1 x_1 + \beta_{12} x_1 x_2 + \varepsilon$$

en donde el término de producto cruzado representa una interacción de dos factores. Ahora bien, un modelo jerárquico necesitaría la inclusión del otro efecto principal, x_2. Sin embargo, ese otro término podría ser totalmente innecesario desde una perspectiva de significado estadístico. Podrá ser perfectamente lógico, desde el punto de vista científico o técnico, tener una interacción en el modelo sin uno de los efectos principales individuales (o hasta en algunos casos, los dos). Esto sucede con frecuencia cuando algunas de las variables implicadas en la interacción son categóricas. El mejor consejo es ajustar un modelo que contenga todos los términos significativos y usar el conocimiento de la disciplina, más que una regla arbitraria, como guía adicional para formular el modelo.

Ahora se ilustrarán algunos de los análisis que se relacionan en forma característica con el ajuste de un modelo polinomial en una variable.

Ejemplo 7.1 *Datos de madera dura*

La tabla 7.1 presenta datos acerca de la resistencia del papel kraft y el porcentaje de madera dura en el lote de pulpa con el que se fabricó. En la figura 7.3 se ve el diagrama de dispersión para esos datos. Esta presentación y el conocimiento del proceso de producción parecen indicar que un modelo cuadrático puede describir en forma adecuada la relación entre la resistencia a la tensión y la concentración de fibra corta (es decir, de madera

http://carlos2524.jimdo.com/

TABLA 7.1 Concentración de madera dura en la pulpa, y resistencia del papel kraft a la tensión, ejemplo 7.1

x_i, concentración de madera dura (%)	y_i, resistencia a la tensión (psi)
1	6.3
1.5	11.1
2	20.0
3	24.0
4	26.1
4.5	30.0
5	33.8
5.5	34.0
6	38.1
6.5	39.9
7	42.0
8	46.1
9	53.1
10	52.0
11	52.5
12	48.0
13	42.8
14	27.8
15	21.9

dura). Si se adopta la recomendación de que al centrar los datos se puede eliminar el mal acondicionamiento no esencial, se ajustará el modelo

$$y = \beta_0 + \beta_1(x - \bar{x}) + \beta_2(x - \bar{x})^2 + \varepsilon$$

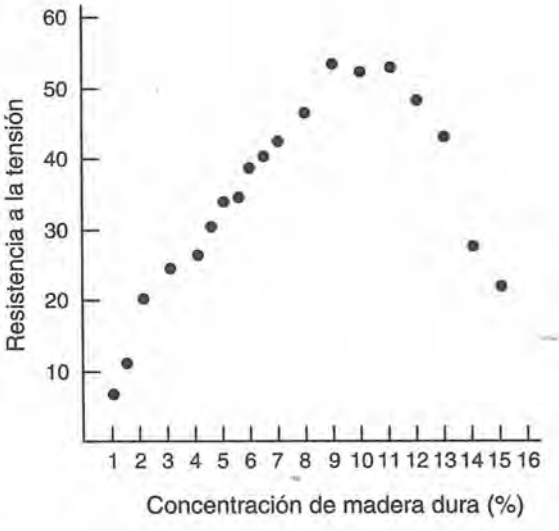


Figura 7.3 Diagrama de dispersión, ejemplo 7.1.

http://carlos2524.jimdo.com/

TABLA 7.2 Análisis de varianza para el modelo cuadrático del ejemplo 7.1

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Regresión	3104.247	2	1552.123	79.434	4.91×10^{-9}
Residuales	312.638	16	19.540		
Total	3416.885	18			

En vista de que ajustar este modelo equivale a ajustar un modelo de regresión con dos variables, se puede adoptar el método general del capítulo 3. El modelo ajustado es

$$\hat{y} = 45.295 + 2.546(x - 7.2632) - 0.635(x - 7.2632)^2$$

En la tabla 7.2 se presenta el análisis de varianza para este modelo. El valor observado de $F_0 = 79.434$ y el valor de P es muy pequeño, por lo que se rechaza la hipótesis H_0: $\beta_1 = \beta_2 = 0$. Se llega a la conclusión que el término lineal o el cuadrático (o ambos) contribuyen al modelo en forma significativa. Las demás estadísticas de resumen para este modelo son $R^2 = 0.9085$, el error estándar $\text{se}(\hat{\beta}_1) = 0.254$ y $\text{se}(\hat{\beta}_2) = 0.062$.

En la figura 7.4 se ve la gráfica de residuales en función de $\hat{y}_i$. En ella no se ve inadecuación grave del modelo. En la figura 7.5 se muestra la gráfica de probabilidad normal de los residuales, que es algo perturbadora, porque indica que la distribución de errores tiene colas más gruesas que la distribución normal. Sin embargo, aún no se cuestiona seriamente la suposición de normalidad.

Ahora supóngase que se desea investigar la contribución del término cuadrático al modelo, esto es, se quiere probar

$$H_0: \beta_2 = 0$$
$$H_1: \beta_2 \neq 0$$

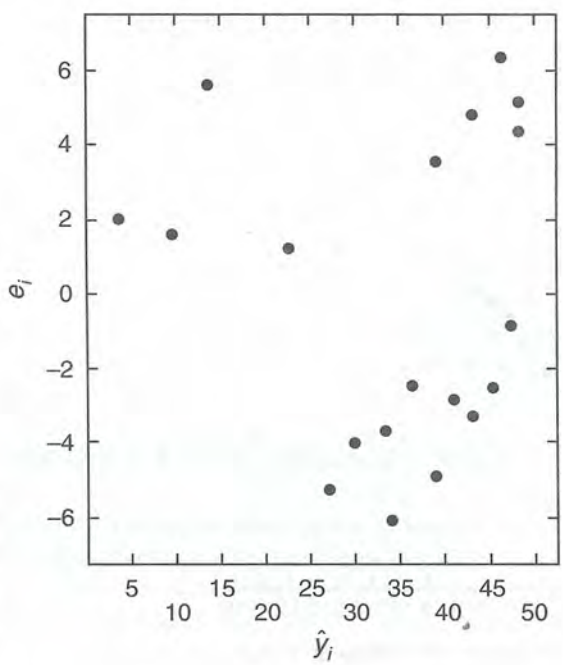


Figura 7.4 Gráfica de residuales e_i en función de valores ajustados $\hat{y}_i$, ejemplo 7.1.

http://carlos2524.jimdo.com/

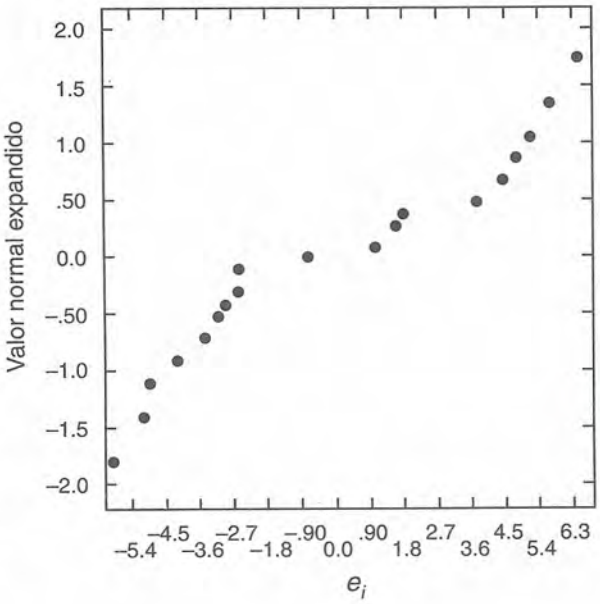


Figura 7.5 Gráfica de probabilidad normal de los residuales, ejemplo 7.1.

Se probará esta hipótesis con el método de la suma extra de cuadrados. Si $\beta_2 = 0$, el modelo reducido es la recta $y = \beta_0 + \beta_1(x - \bar{x}) + \varepsilon$. El ajuste por mínimos cuadrados es

$$\hat{y} = 34.184 + 1.771(x - 7.2632)$$

Las estadísticas de resumen para este modelo son $MS_{\text{Res}} = 139.615$, $R^2 = 0.3054$, $\text{se}(\hat{\beta}_1) = 0.648$ y $SS_{\text{R}}(\beta_1|\beta_0) = 1\,043.427$. Se ve que al eliminar el término cuadrático se afectó R^2 en forma sustancial, así como MS_{Res} y $\text{se}(\beta_1)$. Estas estadísticas de resumen son mucho peores que las del modelo cuadrático. La suma extra de cuadrados para probar H_0: $\beta_2 = 0$ es

$$\begin{aligned} SS_{\text{R}}(\beta_2|\beta_1, \beta_0) &= SS_{\text{R}}(\beta_1, \beta_2|\beta_0) - SS_{\text{R}}(\beta_1|\beta_0) \\ &= 3\,104.247 - 1\,043.427 \\ &= 2\,060.820 \end{aligned}$$

con un grado de libertad. La estadística F es

$$F_0 = \frac{SS_{\text{R}}(\beta_2|\beta_1, \beta_0)/1}{MS_{\text{Res}}} = \frac{2\,060.820/1}{19.540} = 105.47$$

y como $F_{0.01,1,16} = 8.53$, se llega a la conclusión que $\beta_2 \neq 0$. Por lo anterior, el término cuadrático contribuye al modelo en forma significativa.

7.2.2 Ajuste polinomial por segmentos (curvas SPLINE)

A veces se ve que un polinomio de bajo orden proporciona mal ajuste a los datos, y que al aumentar en forma modesta, el orden del polinomio no mejora mucho la situación. Los indicios de esto son que no se estabiliza la suma de cuadrados de residuales, o que las gráficas

http://carlos2524.jimdo.com/

de residuales muestran una estructura remanente inexplicable. Este problema se puede presentar cuando la función se comporta en forma distinta en diferentes partes del rango de x. A veces, las transformaciones de x y/o de y eliminan este problema; sin embargo, el método acostumbrado es dividir el rango de x en segmentos y ajustar la curva adecuada en cada segmento. Las **funciones SPLINE** ofrecen una forma útil de hacer este tipo de **ajuste polinomial por segmentos**.

Las splines son polinomios de orden k por segmentos. Los puntos de unión de los segmentos se suelen llamar **nudos**. Por lo general, se requiere que los valores de la función y de las primeras $k - 1$ derivadas concuerden en los nudos, para que la spline sea una función continua con $k - 1$ derivadas continuas. La **spline cúbica** ($k = 3$) suele ser adecuada para la mayor parte de los problemas prácticos.

Una función spline cúbica con h nudos, $t_1 < t_2 < \cdots < t_h$, con primera y segunda derivadas continuas, se puede escribir como sigue:

$$E(y) = S(x) = \sum_{j=0}^{3} \beta_{0j} x^j + \sum_{i=1}^{h} \beta_i (x - t_i)^3_+ \tag{7.3}$$

en la que

$$(x - t_i)_+ = \begin{cases} (x - t_i) & \text{si } x - t_i > 0 \\ 0 & \text{si } x - t_i \leq 0 \end{cases}$$

Se supondrá que se conocen las posiciones de los nudos, si esas son parámetros que se deben estimar, el problema que resulta es de regresión no lineal; sin embargo, cuando se conocen las posiciones de los nudos, se puede ajustar la ecuación (7.3) aplicando mínimos cuadrados lineales en forma directa.

No es fácil decidir la cantidad y la posición de los nudos, ni el orden del polinomio en cada segmento. Wold [1974] sugiere que debe tenerse un mínimo posible de nudos, y en cada segmento al menos cuatro o cinco puntos de datos. Se debe tener mucho cuidado aquí, porque la gran flexibilidad de las funciones spline provoca fácilmente "sobreajustar" los datos. Wold también sugiere que no debe haber más de un punto extremo (máximo o mínimo) y uno de inflexión por segmento. Hasta donde sea posible, los puntos extremos deberían estar centrados en el segmento, y los puntos de inflexión deberían estar cerca de los nudos. Disponer de información previa acerca del proceso que genera los datos puede ser útil para ubicar los nudos.

El modelo básico de función spline cúbica (7.3) se puede modificar con facilidad para ajustar polinomios de orden distinto en cada segmento, y para imponer distintas restricciones de continuidad en los nudos. Si todos los $h + 1$ tramos polinomiales son de orden 3, un modelo de función spline cúbica sin restricciones de continuidad es

$$E(y) = S(x) = \sum_{j=0}^{3} \beta_{0j} x^j + \sum_{i=1}^{h} \sum_{j=0}^{3} \beta_{ij} (x - t_i)^j_+ \tag{7.4}$$

en el que $(x - t)_+^0$ es igual a 1 si $x > t$, e igual a 0 si $x \leq t$. Así, si en el modelo hay un término $\beta_{ij}(x - t_i)^j_+$, éste causa una discontinuidad de la j-ésima derivada de $S(x)$ en t_i. Si ese término no existe, la j-ésima derivada de $S(x)$ es continua en t_i. Mientras menores restricciones de continuidad se requieran, el ajuste es mejor, porque en el modelo habrá más parámetros; mientras que cuanto más restricciones de continuidad se requieran el ajuste será peor, pero la curva final será más uniforme. Se puede determinar tanto el orden de los

http://carlos2524.jimdo.com/

segmentos del polinomio como las restricciones de continuidad que no degraden en forma sustancial el ajuste, usando los métodos estándar de prueba de hipótesis para regresión múltiple.

Como ilustración, se tiene una función local cúbica con un solo nudo en t, y sin restricciones de continuidad, como la siguiente:

$$E(y) = S(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3 + \beta_{10}(x - t)_+^0 + \beta_{11}(x - t)_+^1 + \beta_{12}(x - t)_+^2 + \beta_{13}(x - t)_+^3$$

Nótese que ni $S(x)$, $S'(x)$ ni $S''(x)$ son necesariamente continuas en t, por la presencia de los términos con β_{10}, β_{11} y β_{12} en el modelo. Para determinar si al imponer restricciones de continuidad se reduce la calidad del ajuste, se prueban las hipótesis H_0: $\beta_{10} = 0$ [continuidad de $S(x)$], H_0: $\beta_{10} = \beta_{11} = 0$ [continuidad de $S(x)$ y $S'(x)$] y H_0: $\beta_{10} = \beta_{11} = \beta_{12} = 0$ [continuidad de $S(x)$, $S'(x)$ y $S''(x)$]. Para determinar si la spline cúbica se ajusta a los datos mejor que un solo polinomio cúbico dentro del rango de x, sólo se prueba H_0: $\beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = 0$.

En Smith [1979] se presenta una excelente descripción de este método para ajustar funciones spline. Una desventaja potencial de este método es que la matriz $\mathbf{X}'\mathbf{X}$ se vuelve mal acondicionada si hay una gran cantidad de nudos. Este problema se puede resolver usando una representación distinta de la función spline, llamada **spline *B* cúbica**. Las spline B cúbicas se definen en función de las diferencias divididas

$$B_i(x) = \sum_{j=i-4}^{i} \left[\frac{(x - t_j)_+^3}{\prod_{\substack{m=i-4 \\ m \neq j}}^{i} (t_j - t_m)} \right], \quad i = 1, 2, \ldots, h + 4 \tag{7.5}$$

y

$$E(y) = S(x) = \sum_{i=1}^{h+4} \gamma_i B_i(x) \tag{7.6}$$

en donde γ_i, $i = 1, 2, \ldots, h + 4$ son parámetros por estimar. En la ecuación (7.5) hay ocho nudos adicionales, $t_{-3} < t_{-2} < t_{-1} < t_0$, y $t_{h+1} < t_{h+2} < t_{h+3} < t_{h+4}$. Se suelen igualar $t_0 = x_{\text{mín}}$ y $t_{h+1} = x_{\text{máx}}$; los demás nudos son arbitrarios. Para mayor documentación acerca de funciones spline véanse Buse y Lim [1977], Curry y Schoenberg [1966], Eubank [1988], Gallant y Fuller [1973], Hayes [1970, 1974], Poirer [1973, 1975] y Wold [1974].

Ejemplo 7.2 *Datos de caída de voltaje*

La caída de voltaje en la batería del motor de un misil guiado, que se observa durante el tiempo de vuelo del misil, se muestra en la tabla 7.3. El diagrama de dispersión de la figura 7.6 parece indicar que la caída de voltaje se comporta en forma distinta en diferentes intervalos de tiempo, por lo que se modelarán los datos con una spline cúbica usando dos nudos en $t_1 = 6.5$ y $t_2 = 13$ segundos después del lanzamiento, respectivamente. La colocación de los nudos concuerda en forma aproximada con los cambios de curso del proyectil

http://carlos2524.jimdo.com/

TABLA 7.3 Datos de caída de voltaje

Observación i	Tiempo (segundos) x_i	Caída de voltaje y_i	Observación i	Tiempo (segundos) x_i	Caída de voltaje y_i
1	0.0	8.33	21	10.0	14.48
2	0.5	8.23	22	10.5	14.92
3	1.0	7.17	23	11.0	14.37
4	1.5	7.14	24	11.5	14.63
5	2.0	7.31	25	12.0	15.18
6	2.5	7.60	26	12.5	14.51
7	3.0	7.94	27	13.0	14.34
8	3.5	8.30	28	13.5	13.81
9	4.0	8.76	29	14.0	13.79
10	4.5	8.71	30	14.5	13.05
11	5.0	9.71	31	15.0	13.04
12	5.5	10.26	32	15.5	12.60
13	6.0	10.91	33	16.0	12.05
14	6.5	11.67	34	16.5	11.15
15	7.0	11.76	35	17.0	11.15
16	7.5	12.81	36	17.5	10.14
17	8.0	13.30	37	18.0	10.08
18	8.5	13.88	38	18.5	9.78
19	9.0	14.59	39	19.0	9.80
20	9.5	14.05	40	19.5	9.95
			41	20.0	9.51

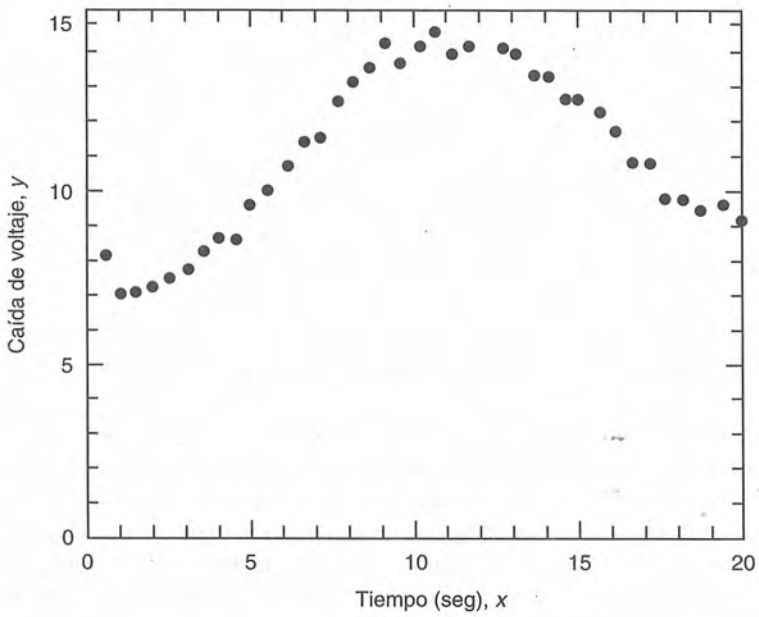


Figura 7.6 Gráfica de dispersión para los datos de caída de voltaje.

http://carlos2524.jimdo.com/

(con los cambios asociados en necesidades de energía), que se conocen por los datos de la trayectoria. El modelo de caída de voltaje se pretende usar en un modelo digital-analógico de simulación del misil.

El modelo de función local cúbica es

$$y = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3 + \beta_1(x-6.5)_+^3 + \beta_2(x-13)_+^3 + \varepsilon$$

y el ajuste por mínimos cuadrados es

$$\hat{y} = 8.4657 - 1.4531x + 0.4899x^2 - 0.0295x^3 + 0.0247(x-6.5)_+^3 + 0.0271(x-13)_+^3$$

Las estadísticas de resumen para el modelo se ven en la tabla 7.4. En las figuras 7.7 y 7.8 se ven gráficas de los residuales en función de $\hat{y}$ y de probabilidad normal de los residuales, respectivamente. Esos datos no indican que haya desviaciones graves respecto a las premisas, por lo que la conclusión es que el modelo de spline cúbica es un ajuste adecuado a los datos de caída de voltaje.

Se pueden comparar con facilidad el ajuste de modelo con spline cúbica y el obtenido con un solo polinomio cúbico en todo el tiempo de vuelo del misil; por ejemplo

$$\hat{y} = 6.4910 + 0.7032x + 0.0340x^2 - 0.0033x^3$$

Éste es un modelo más sencillo que contiene menos parámetros, y que sería preferible al de la spline cúbica si proporcionara un ajuste satisfactorio. En la figura 7.9 se grafican los residuales de este polinomio cúbico en función de $\hat{y}$. En esta gráfica se ve una fuerte indicación de curvatura, y con base en la estructura remanente inexplicada, se llega a la conclusión de que el polinomio cúbico simple es un modelo inadecuado para los datos de caída de voltaje.

TABLA 7.4 Estadísticas de resumen para el modelo de Spline cúbica, con datos de caída de voltaje

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Regresión	260.1784	5	52.0357	725.52	< 0.0001
Residual	2.5102	35	0.0717		
Total	262.6886	40			

Parámetros	Estimado	Error Estándar	Valor t para $H_0: \beta = 0$	Valor P
β_{00}	8.4657	0.2005	42.22	< 0.0001
β_{01}	− 1.4531	0.1816	− 8.00	< 0.0001
β_{02}	0.4899	0.0430	11.39	< 0.0001
β_{03}	− 0.0295	0.0028	− 10.54	< 0.0001
β_1	0.0247	0.0040	6.18	< 0.0001
β_2	0.0271	0.0036	7.53	< 0.0001
		$R^2 = 0.9904$		

http://carlos2524.jimdo.com/

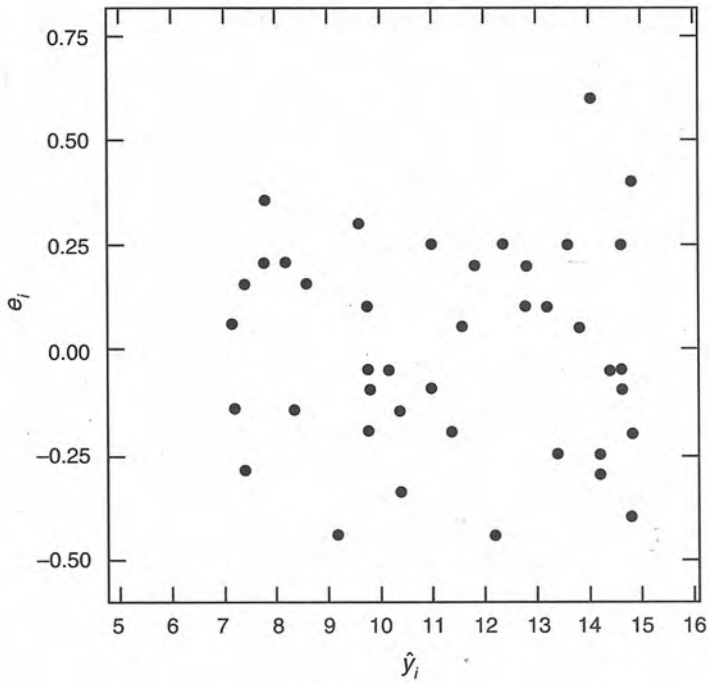


Figura 7.7 Gráfica de residuales e_i en función de valores ajustados $\hat{y}_i$, para el modelo con spline cúbica.

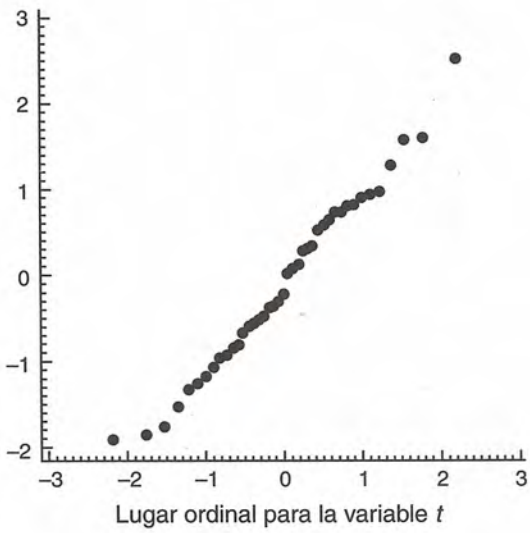


Figura 7.8 Gráfica de probabilidad normal de los residuales R de Student para el modelo con spline cúbica.

También se puede investigar si el modelo de spline cúbica mejora el ajuste, probando la hipótesis H_0: $\beta_1 = \beta_2 = 0$, con el método de la suma extra de cuadrados. La suma de cuadrados de la regresión para el polinomio cúbico es

$$SS_R(\beta_{01}, \beta_{02}, \beta_{03} | \beta_{00}) = 230.4444$$

http://carlos2524.jimdo.com/

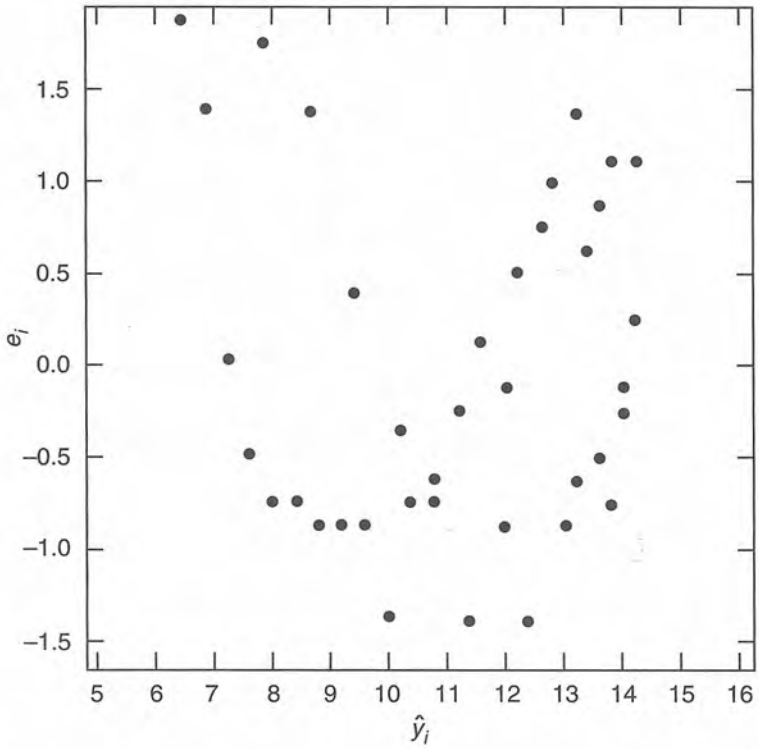


Figura 7.9 Gráfica de residuales e_i en función de valores ajustados $\hat{y}_i$, para el modelo de polinomio cúbico.

con tres grados de libertad. La suma extra de cuadrados para probar H_0: $\beta_1 = \beta_2 = 0$ es

$$\begin{aligned} SS_R(\beta_1, \beta_2 | \beta_{00}, \beta_{01}, \beta_{02}, \beta_{03}) &= SS_R(\beta_{01}, \beta_{02}, \beta_{03}, \beta_2, \beta_2 | \beta_{00}) \\ &\quad - SS_R(\beta_{01}, \beta_{02}, \beta_{03} | \beta_{00}) \\ &= 260.1784 - 230.4444 \\ &= 29.7340 \end{aligned}$$

con dos grados de libertad. Ya que

$$F_0 = \frac{SS_R(\beta_1, \beta_2 | \beta_{00}, \beta_{01}, \beta_{02}, \beta_{03})/2}{MS_{Res}} = \frac{29.7340/2}{0.0717} = 207.35$$

que se compararía con la distribución $F_{2,35}$, se rechaza la hipótesis que H_0: $\beta_1 = \beta_2 = 0$. La conclusión es que el modelo de la spline cúbica proporciona mejor ajuste.

Ejemplo 7.3 *Regresión lineal por segmentos*

Un caso especial, de interés práctico, implica ajustar modelos de regresión lineal por segmentos. Esto se puede manejar con facilidad usando **spline lineales**, por ejemplo, supóngase que hay un solo nudo en t, y que en él podría haber un cambio de pendiente y una discontinuidad. El modelo de spline lineal resultante es

$$E(y) = S(x) = \beta_{00} + \beta_{01}x + \beta_{10}(x - t)_+^0 + \beta_{11}(x - t)_+^1$$

http://carlos2524.jimdo.com/

Ahora bien, si $x \leq t$, el modelo de línea recta es

$$E(y) = \beta_{00} + \beta_{01}x$$

y si $x > t$, el modelo es

$$\begin{aligned} E(y) &= \beta_{00} + \beta_{01}x + \beta_{10}(1) + \beta_{11}(x - t) \\ &= (\beta_{00} + \beta_{10} - \beta_{11}t) + (\beta_{01} + \beta_{11})x \end{aligned}$$

Esto es, si $x \leq t$, el modelo tiene ordenada al origen β_{00} y pendiente β_{01}, mientras que si $x > t$, la ordenada al origen es $\beta_{00} + \beta_{10} - \beta_{11}t$ y la pendiente es $\beta_{01} + \beta_{11}$. La función de regresión se muestra en la figura 7.10a. Nótese que el parámetro β_{10} representa la diferencia en la respuesta promedio en el nudo t.

Si se pidiera que la función de regresión sea continua en el nudo se obtendría una función más lisa. Eso se hace con facilidad eliminando el término $\beta_{10}(x - t)_+^0$ del modelo original, y se obtiene

$$E(y) = S(x) = \beta_{00} + \beta_{01}x + \beta_{11}(x - t)_+^1$$

Ahora bien, si $x \leq t$, el modelo es

$$E(y) = \beta_{00} + \beta_{01}x$$

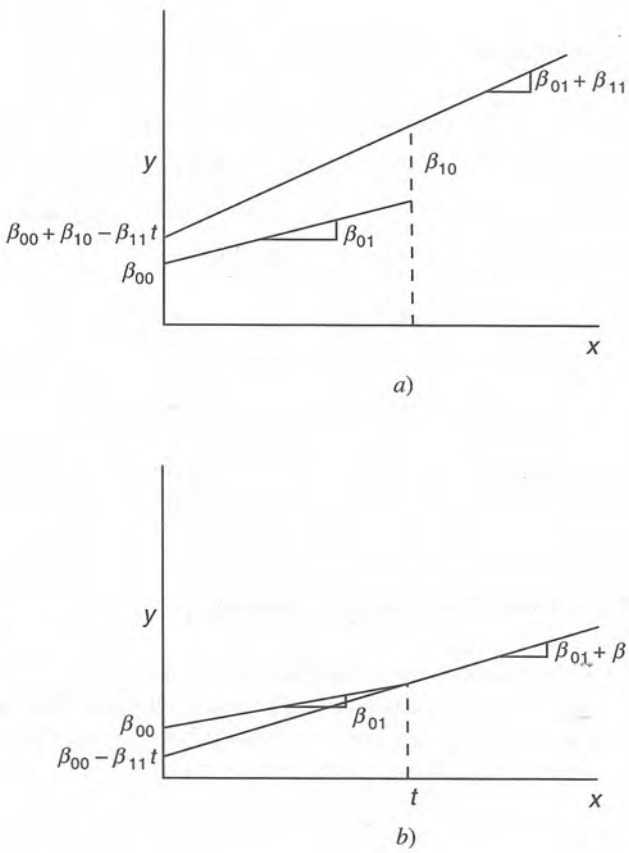


Figura 7.10 Regresión lineal por intervalos: a) discontinuidad en el nudo; b) modelo de regresión lineal continua por intervalos.

http://carlos2524.jimdo.com/

y si $x > t$, el modelo es

$$\begin{aligned} E(y) &= \beta_{00} + \beta_{01}x + \beta_{11}(x - t) \\ &= (\beta_{00} - \beta_{11}t) + (\beta_{01} + \beta_{11})x \end{aligned}$$

En la figura 7.10b se ven las dos funciones de regresión.

7.2.3 Términos polinomiales y trigonométricos

A veces conviene considerar modelos donde se combinan términos tanto polinomiales como trigonométricos, como modelos alternativos a los que sólo contienen términos polinomiales. En particular, si el diagrama de dispersión indica que puede haber cierta periodicidad o comportamiento cíclico de los datos, podrá ser muy benéfico agregar términos trigonométricos, porque puede resultar un modelo con menos términos que si sólo se emplean términos trigonométricos. Esta ventaja la hacen notar Graybill [1976] y Eubank y Speckman [1990].

El modelo para un solo regresor x es

$$y = \beta_0 + \sum_{i=1}^{d} \beta_i x^i + \sum_{j=1}^{r} \left[\delta_j \operatorname{sen}(jx) + \gamma_j \cos(jx)\right] + \varepsilon$$

Si el regresor x está igualmente espaciado, los pares de términos sen(jx) y cos(jx) son ortogonales. Aun sin que el espaciamiento sea exactamente igual, la correlación entre esos términos será, en general, muy pequeña.

Eubank y Speckman [1990] usaron los datos de caída de voltaje del ejemplo 7.2 para ilustrar el ajuste de un modelo de regresión polinomial-trigonométrico. Primero escalan el regresor x (tiempo) de tal modo que todas las observaciones queden dentro del intervalo $(0, 2\pi)$, y ajustan el modelo anterior con $d = 2$ y $r = 1$, de tal modo que el modelo es cuadrático en el tiempo, y tiene un par de términos en seno y coseno. Así, su modelo sólo tiene cuatro términos, mientras que el modelo de regresión spline tuvo cinco. Eubank y Speckman obtuvieron $R^2 = 0.9895$ y $MS_{\text{Res}} = 0.0767$, resultados que se parecen mucho a los encontrados con el modelo con spline (véase la Tabla 7.4). Como los datos de caída de voltaje muestran cierto indicio de periodicidad en el diagrama de dispersión (Fig. 7.6), el modelo de regresión polinomial-trigonométrico es definitivamente una buena alternativa al modelo con spline. Tiene un término menos (siempre es una propiedad deseable), aun cuando un cuadrado medio residual sea un poco mayor. Algunos usuarios también consideran que es una desventaja potencial el trabajar con una versión reescalada de la variable regresora.

7.3 REGRESIÓN NO PARAMÉTRICA

Muy relacionada con la regresión polinomial por segmentos está la **regresión no paramétrica**. La idea básica de esta regresión es desarrollar una base **de modelo libre** para predecir la respuesta sobre el rango de los datos. Los primeros métodos de regresión no paramétrica se basaron mucho en la estimación no paramétrica de función de densidad. La mayor parte de las publicaciones sobre regresión no paramétrica se enfocan hacia un solo regresor, sin embargo, muchas de las ideas básicas se aplican también a más de uno solo.

http://carlos2524.jimdo.com/

Una perspectiva fundamental de la regresión no paramétrica es la naturaleza del valor predicho. Considérense los mínimos cuadrados ordinarios estándar. Se recordará que

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{H}\mathbf{y} \\ &= \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}\end{aligned}$$

por lo que

$$\hat{y}_i = \sum_{j=1}^{n} h_{ij}\, y_j$$

En otras palabras, el valor predicho para la i-ésima respuesta es tan sólo una combinación lineal de los datos originales.

7.3.1 Regresión Kernel

Uno de los primeros métodos no paramétricos alternativos es el alisador de **Kernel**, donde se usa un promedio ponderado de los datos. Sea $\tilde{y}_i$ el estimado de alisador de núcleo para la i-ésima respuesta. Para un alisador Kernel,

$$\tilde{y}_i = \sum_{j=1}^{n} w_{ij} y_j$$

siendo $\Sigma_{j=1}^{n} w_{ij} = 1$. Como resultado,

$$\mathbf{y} = \mathbf{S}\mathbf{y}$$

donde $\mathbf{S} = [w_{ij}]$ es la matriz de "alisamiento". Comúnmente, los pesos o factores de ponderación se escogen de tal modo que $w_{ij} \cong 0$ para todas las y_i fuera de una "proximidad" del lugar de interés específico. Estos alisadores Kernel usan un ancho de banda b para definir la proximidad de interés. Un valor grande de b da como resultado que un mayor número de datos se usen para predecir la respuesta en el lugar específico. En consecuencia, la gráfica resultante de valores predichos se vuelve mucho más lisa a medida que b aumenta. Al revés, cuando b disminuye se usan menos de los datos para generar la predicción y la gráfica que resulta se ve más "vacilante" o con irregularidades.

Este método se llama alisador Kernel porque usa una función Kernel K para especificar los pesos. Las propiedades de las funciones Kernel son las siguientes:

- $K(t) \geq 0$ para toda t
- $\int_{-\infty}^{\infty} K(t)\, dt = 1$
- $K(-t) = K(t)$ (simetría)

http://carlos2524.jimdo.com/

TABLA 7.5 Resumen de las funciones Kernel que se usan en S-PLUS

Box	$K(t) = \begin{cases} 1 & \lvert t\rvert \le 0.5 \\ 0 & \lvert t\rvert > 0.5 \end{cases}$
Triángulo	$K(t) = \begin{cases} 1 - \dfrac{\lvert t\rvert}{c} & \lvert t\rvert \le \dfrac{1}{c} \\ 0 & \lvert t\rvert > \dfrac{1}{c} \end{cases}$
Parzen	$K(t) = \begin{cases} \dfrac{k_1 - t^2}{k^2} & \lvert t\rvert \le C_1 \\ \dfrac{t^2}{k_3} - k_4\lvert t\rvert + k_5 & C_1 < \lvert t\rvert \le C_2 \\ 0 & \lvert t\rvert > C_2 \end{cases}$
Normal	$K(t) = \dfrac{1}{\sqrt{2\pi}k_6}\exp\left\{\dfrac{-t^2}{2k_6^2}\right\}$

Estas propiedades también son las de una función de densidad simétrica de probabilidades, que enfatiza la relación con la estimación no paramétrica de densidad.

Los pesos específicos del alisador Kernel se calculan con

$$w_{ij} = \frac{K\left(\dfrac{x_i - x_j}{b}\right)}{\displaystyle\sum_{k=1}^{n} K\left(\dfrac{x_i - x_k}{b}\right)}$$

La tabla 7.5 es un resumen de los Kernel que usa el programa de computadora S-PLUS. Hastie y Tibshirani (1990) hacen notar que las propiedades del alisador Kernel dependen mucho más de la elección del ancho de banda que de la función Kernel real.

7.3.2 Regresión ponderada localmente (Loess)

Otra alternativa no paramétrica es la **regresión ponderada localmente**, a la que con frecuencia se le llama **loess**. Al igual que la regresión Kernel, la loess usa los datos próximos al lugar específico. Normalmente, la proximidad se define como el **tramo**, que es la fracción de los puntos totales que se usan para formar las proximidades. Un tramo de 0.5 indica que la mitad más próxima del total de puntos de datos se usa como proximidad. El procedimiento loess usa entonces los puntos en la proximidad para generar un estimado por mínimos cuadrados ponderados, de la respuesta específica. El procedimiento de mínimos cuadrados ponderados usa un polinomio de bajo orden, que suele ser la regresión lineal simple, o un modelo cuadrático de regresión. Los pesos o factores de ponderación para la parte de mínimos cuadrados ponderados de la estimación se basan en la distancia de los puntos que se usaron en la estimación, al lugar específico de interés. En S-PLUS se usa la función de ponderación **tricubo** como la de default. Sea x_0 el lugar específico de interés,

http://carlos2524.jimdo.com/

y sea $\Delta(x_0)$ la distancia del punto más alejado de la proximidad al punto específico de interés. La función de peso tricubo es

$$W\left[\frac{|x_0 - x_j|}{\Delta(x_0)}\right]$$

en donde

$$W(t) = \begin{cases} (1 - t^3)^3 & \text{para } 0 \leq t < 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

El proceso loess de estimación se resume como sigue:

$$\mathbf{y} = \mathbf{S}\mathbf{y}$$

en donde $\mathbf{S}$ es la matriz de alisamiento creada por la regresión localmente ponderada.

El concepto de suma de residuales al cuadrado lleva en forma directa a la regresión no paramétrica. En particular

$$\begin{aligned} SS_{\text{Res}} &= \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \\ &= (\mathbf{y} - \mathbf{S}\mathbf{y})'(\mathbf{y} - \mathbf{S}\mathbf{y}) \\ &= \mathbf{y}'[\mathbf{I} - \mathbf{S}'][\mathbf{I} - \mathbf{S}]\mathbf{y} \\ &= \mathbf{y}'[\mathbf{I} - \mathbf{S}' - \mathbf{S} + \mathbf{S}'\mathbf{S}]\mathbf{y} \end{aligned}$$

En forma asintótica, estos procedimientos de alisamiento son insesgados, por consiguiente, el valor asintótico esperado de SS_{Res} es

$$\begin{aligned} &\text{traza}\left[(\mathbf{I} - \mathbf{S}' - \mathbf{S} + \mathbf{S}'\mathbf{S})\sigma^2\mathbf{I}\right] \\ &\quad = \sigma^2/\text{traza}\left[\mathbf{I} - \mathbf{S}' - \mathbf{S} + \mathbf{S}'\mathbf{S}\right] \\ &\quad = \sigma^2\left[\text{traza}(\mathbf{I}) - \text{traza}(\mathbf{S}') - \text{traza}(\mathbf{S}) + \text{traza}(\mathbf{S}'\mathbf{S})\right] \end{aligned}$$

Es importante observar que $\mathbf{S}$ es una matriz cuadrada de $n \times n$. Entonces, la traza[$\mathbf{S}'$] = traza[$\mathbf{S}$], y así

$$E(SS_{\text{Res}}) = \sigma^2[n - 2\,\text{traza}(\mathbf{S}) + \text{traza}(\mathbf{S}'\mathbf{S})]$$

En cierto sentido [2 traza($\mathbf{S}$) – traza($\mathbf{S}'\mathbf{S}$)] representa los grados de libertad asociados con el modelo total. En algunos programas de cómputo [2 traza($\mathbf{S}$) – traza($\mathbf{S}'\mathbf{S}$)] es llamada la cantidad equivalente de parámetros, y representa una medida de la complejidad del procedimiento de estimación. Un estimado común de σ^2 es:

$$\tilde{\sigma}^2 = \frac{\sum_{i=1} (y_i - \tilde{y}_i)^2}{n - 2\,\text{traza}(\mathbf{S}) + \text{traza}(\mathbf{S}'\mathbf{S})}$$

http://carlos2524.jimdo.com/

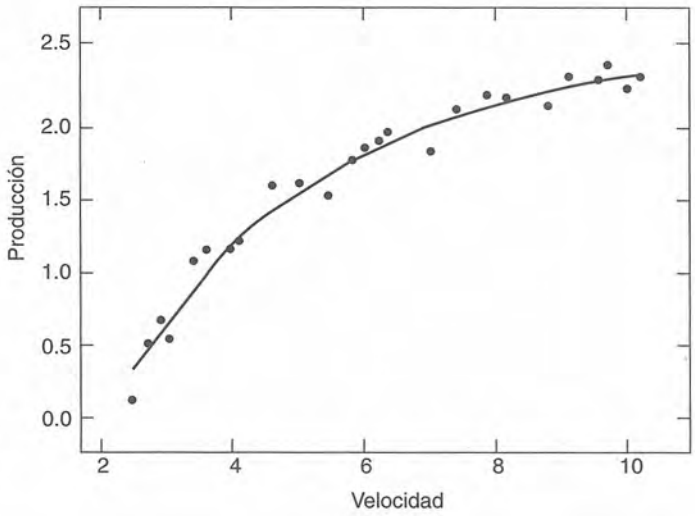


Figura 7.11 Ajuste loess para los datos del molino de viento.

Por último, se puede definir una versión de R^2 como sigue:

$$R^2 = \frac{SS_T - SS_{Res}}{SS_T}$$

cuya interpretación es la misma que antes, en los mínimos cuadrados ordinarios. Todo esto se amplía en forma natural al caso de la regresión múltiple, y el programa S–PLUS tiene estas posibilidades.

Ejemplo 7.4 *Aplicación de la regresión loess a los datos del molino de viento*

En el ejemplo 5.2 se describieron los datos reunidos por un ingeniero que investigaba la relación entre la velocidad del viento y la CD producida por un molino de viento, la tabla 5.5 resume esos datos. El final en este ejemplo se desarrolló en un modelo de regresión lineal simple, donde interviene la inversa de la velocidad del viento. Este modelo proporcionó una

TABLA 7.6 Resultados del ajuste loess para los datos del molino de viento, obtenidos con S-PLUS

```
***Local Regression Model***
Call:
loess (formula=output ~ velocity, data=windmill,
       na.action=na.exclude, span=0.75, degree=2,
       normalize=T, family= ``gaussian´´)
Number of Observations:            25
Equivalent Number of Parameters:  4.4
Residual Standard Error:          0.1017
Multiple R-squared:               0.98
Residuals:
min              1st Q        median          3rd Q          max
-0.1756       -0.07051       0.01575        0.06744       0.1543
```

http://carlos2524.jimdo.com/

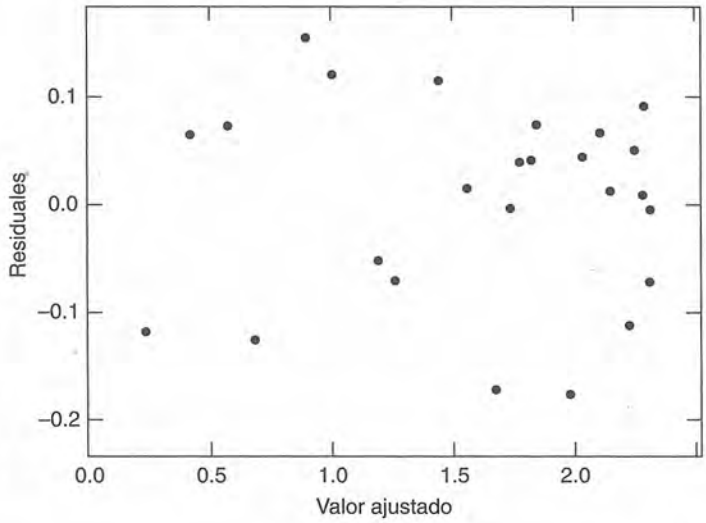


Figura 7.12 Residuales en función de valores ajustados, para ajuste loess con los datos del molino de viento.

buena base para modelar el hecho que hay una cota superior verdadera para la salida real de CD que puede generar el molino de viento.

Un método alternativo para este ejemplo usa la regresión loess. La figura 7.11 muestra este ajuste a los datos, obtenido con los ajustes predeterminados de S-PLUS, y la tabla 7.6 resume el informe de S-PLUS. La figura 7.12, que muestra los residuales en función de los valores ajustados, parece no indicar problemas. La figura 7.13 muestra la gráfica de probabilidad normal que, aunque no es perfecta, no indica problema serio alguno. De hecho, se ve muy parecida a la gráfica que se generó con la regresión lineal simple tomando la inversa de la velocidad del viento.

El ajuste a los datos con loess es bastante bueno, y se compara favorablemente con el que se había generado con anterioridad con los mínimos cuadrados ordinarios y la inversa de la velocidad del viento. Los resultados indican una R^2 de 0.98, igual que la del modelo con regresión lineal simple. Aunque los dos valores de R^2 no se pueden comparar en forma

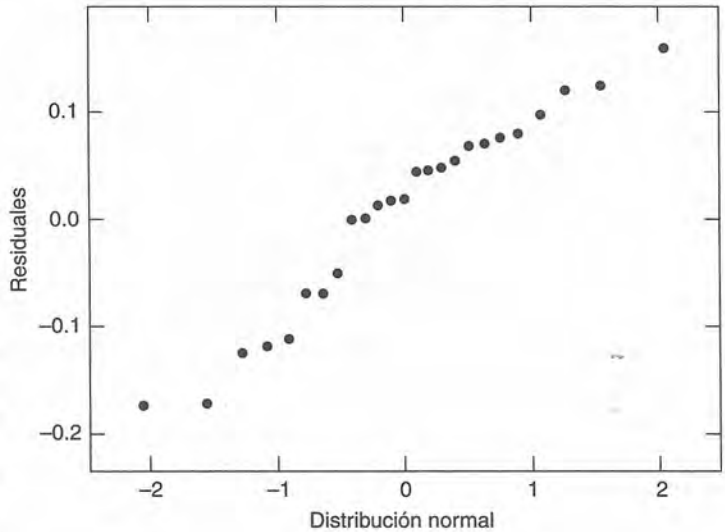


Figura 7.13 Gráfica de probabilidad normal de los residuales, para el ajuste loess con los datos del molino de viento.

http://carlos2524.jimdo.com/

directa, ambos indican un ajuste muy bueno. El MS_{Res} es 0.1017, en comparación con el valor de 0.0089 con el modelo de regresión lineal simple. Es claro que ambos modelos son competitivos entre sí. Es interesante que para el ajuste loess se requiere una cantidad equivalente de parámetros de 4.4, algo intermedia entre un modelo cúbico y uno cuártico. Por otro lado, el modelo de regresión lineal simple usando la inversa de la velocidad del viento sólo requiere dos parámetros, por lo que es mucho más simple. En último término se prefiere el modelo de regresión lineal simple, por ser más simple y porque corresponde a la teoría conocida; por otra parte, el modelo loess es más complicado, y se parece algo a una "caja negra".

7.3.3 Precauciones finales

Los análisis de regresión paramétricos y no paramétricos tienen cada uno sus ventajas y desventajas. Con frecuencia, los modelos paramétricos se guían por la teoría vigente en el área. Los modelos no paramétricos reflejan, casi siempre, empirismo puro.

Siempre se debe preferir un modelo paramétrico sencillo cuando produzca un ajuste razonable y satisfactorio a los datos; el asunto de la complejidad no es trivial. Los modelos simples proporcionan una base fácil y cómoda para la predicción, además, con frecuencia los términos del modelo tienen interpretaciones importantes. Hay casos, como el de los datos del molino de viento, en los que se requieren transformaciones de la respuesta o del regresor para proporcionar un ajuste adecuado a los datos. Aquí también se debe preferir el modelo paramétrico, en especial cuando la teoría en ese campo respalda el modelo de transformación.

Por otra parte, hay muchos casos en los que ningún modelo paramétrico simple proporciona un ajuste adecuado o satisfactorio a los datos, cuando en el área hay poca o ninguna teoría que guíe al analista, y donde ninguna transformación sencilla parece adecuada. En esos casos, la regresión no paramétrica tiene mucho sentido. Se prefiere aceptar la complejidad relativa y la naturaleza de "caja negra" de la estimación para proporcionar un ajuste adecuado a los datos.

7.4 MODELOS POLINOMIALES CON DOS O MÁS VARIABLES

El ajuste de un modelo polinomial con dos o más variables regresoras es una extensión directa del método de la sección 7.2.1. Por ejemplo, un modelo polinomial en dos variables sería

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \tag{7.7}$$

Nótese que este modelo contiene dos parámetros de efecto lineal, β_1 y β_2, dos parámetros de efecto cuadrático, β_{11} y β_{22}, y un parámetro de efecto de interacción, β_{12}.

El ajuste de un modelo de segundo orden, como la ecuación (7.7), ha recibido considerable atención, tanto de los investigadores como de los profesionistas. A la función de regresión

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x x_2^2 + \beta_{12} x_1 x_2$$

http://carlos2524.jimdo.com/

se le llama **superficie de respuesta**. Se puede representar en forma gráfica la superficie de respuesta bidimensional trazando los ejes x_1 y x_2 en el plano del papel, y visualizando al eje $E(y)$ como perpendicular a ese plano. Al trazar las curvas de nivel constante de respuesta esperada $E(y)$ se produce la superficie de respuesta, por ejemplo, véase la figura 3.3, que muestra la superficie de respuesta

$$E(y) = 800 + 10x_1 + 7x_2 - 8.5\,x_1^2 - 5\,x_2^2 + 4x_1x_2$$

Nótese que esta superficie de respuesta es una colina que contiene un punto de respuesta máximo. Hay otras posibilidades, como un valle que contiene un punto de respuesta mínimo, y como un sistema de silla de montar. La **metodología de superficie de respuesta** (RSM, de *response surface methodology*) se aplica mucho en la industria para modelar la o las superficies resultantes de un proceso en función de las variables controlables importantes, para a continuación determinar las condiciones de operación que optimicen la respuesta. Para conocer una descripción detallada de los métodos de superficie de respuesta, véase Box y Draper [1987], Box, Hunter y Hunter [1978], Khuri y Cornell [1996], Montgomery [2001] y Myers y Montgomery [1995].

Ahora se ilustrará el ajuste de una superficie de respuesta de segundo orden con dos variables. La parte A de la tabla 7.7 presenta datos de un experimento para estudiar el efecto de dos variables: la temperatura de reacción T y la concentración de reactivo C, sobre el porcentaje de conversión y en un proceso químico. Los ingenieros de proceso habían usado un método para mejorar este proceso basado en **experimentos diseñados**. El primer experimento fue uno de cribado, donde intervinieron algunos factores que aislaban a la temperatura y la concentración con carácter de las dos variables más importantes. Como los experimentadores creían que este proceso estaba trabajando cerca del óptimo, optaron por ajustar un modelo cuadrático que relacionara el rendimiento con la temperatura y la concentración.

La parte A de la tabla 7.7 muestra los niveles que se usaron de T y C en las unidades naturales de medida. La parte B muestra los niveles en función de las variables codificadas x_1 y x_2. La figura 7.14 muestra el diseño experimental, en forma gráfica, de la tabla 7.5.

TABLA 7.7 Diseño compuesto central para el ejemplo del proceso químico

	A			B		
Observación	**Orden de la corrida**	**Temperatura (°C)** T	**Conc. (%)** C	x_1	x_2	y
1	4	200	15	−1	−1	43
2	12	250	15	1	−1	78
3	11	200	25	−1	1	69
4	5	250	25	1	1	73
5	6	189.65	20	−1.414	0	48
6	7	260.35	20	1.414	0	76
7	1	225	12.93	0	−1.414	65
8	3	225	27.07	0	1.414	74
9	8	225	20	0	0	76
10	10	225	20	0	0	79
11	9	225	20	0	0	83
12	2	225	20	0	0	81

http://carlos2524.jimdo.com/

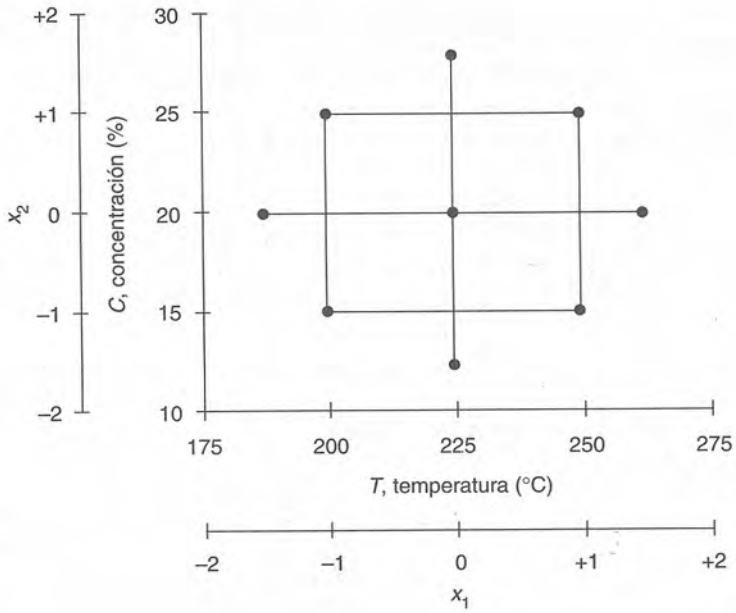


Figura 7.14 Diseño compuesto central para el ejemplo del proceso químico.

A este diseño se le llama **diseño compuesto central** y se usa mucho para ajustar una superficie de respuesta de segundo orden. Nótese que el diseño consiste en cuatro corridas en las esquinas de un cuadrado, más cuatro corridas en el centro de ese cuadrado, más cuatro corridas en los ejes. En términos de las variables codificadas, las esquinas del cuadrado son $(x_1, x_2) = (-1, -1)$, $(1, -1)$, $(-1, 1)$ y $(1, 1)$; los puntos centrales están en $(x_1, x_2) = (0, 0)$, y las corridas axiales están en $(x_1, x_2) = (-1.414, 0)$, $(1.414, 0)$, $(0, -1.414)$ y $(0, 1.414)$.

Se ajustará el modelo de segundo orden

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$$

usando las variables codificadas, ya que esa es la práctica común en la metodología de la superficie de respuesta. La matriz **X** y el vector **y** son:

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccccc} & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{array} \\ \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1.414 & 0 & 2 & 0 & 0 \\ 1 & 1.414 & 0 & 2 & 0 & 0 \\ 1 & 0 & -1.414 & 0 & 2 & 0 \\ 1 & 0 & 1.414 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}, \quad \mathbf{y} = \begin{bmatrix} 43 \\ 78 \\ 69 \\ 73 \\ 48 \\ 76 \\ 65 \\ 74 \\ 76 \\ 79 \\ 83 \\ 81 \end{bmatrix}$$

http://carlos2524.jimdo.com/

Nótese que se han mostrado las variables asociadas con cada columna arriba de esa columna, en la matriz **X**. Los elementos de las columnas asociados con x_1^2 y x_2^2 se calculan elevando al cuadrado los elementos de las columnas x_1 y x_2, respectivamente, y los elementos de la columna x_1x_2 se calculan multiplicando cada elemento de x_1 por el elemento correspondiente en x_2. La matriz $\mathbf{X}'\mathbf{X}$ y el vector $\mathbf{X}'\mathbf{y}$ son

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 12 & 0 & 0 & 8 & 8 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 \\ 8 & 0 & 0 & 12 & 4 & 0 \\ 8 & 0 & 0 & 4 & 12 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}, \qquad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 845.000 \\ 78.592 \\ 33.726 \\ 511.000 \\ 541.000 \\ -31.000 \end{bmatrix}$$

y a partir de $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbf{y}$ se obtiene

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 79.75 \\ 9.83 \\ 4.22 \\ -8.88 \\ -5.13 \\ -7.75 \end{bmatrix}$$

En consecuencia, el modelo ajustado del porcentaje de conversión es

$$\hat{y} = 79.75 + 9.83\,x_1 + 4.22\,x_2 - 8.88\,x_1^2 - 5.13\,x_2^2 - 7.75x_1\,x_2$$

En términos de las variables naturales, el modelo es

$$\hat{y} = -1105.56 + 8.0242T + 22.94C + 0.0142T^2 + 0.20502C^2 + 0.062TC$$

La tabla 7.8 muestra el análisis de varianza para este modelo. Como el diseño experimental tiene cuatro corridas de réplica, se puede agrupar la suma de cuadrados de residuales en componentes de error puro y de falta de ajuste; la prueba de falta de ajuste de la tabla 7.8 es la de la falta de ajuste para el modelo cuadrático. El valor P en esta prueba es grande ($P = 0.8120$), y eso implica que el modelo cuadrático es adecuado. Por lo anterior, el cuadrado medio residual con seis grados de libertad se usará en el análisis restante. La prueba F para el significado de la regresión es $F_0 = 58.86$, y como el valor P es muy pequeño, se rechaza la hipótesis H_0: $\beta_1 = \beta_2 = \beta_{11} = \beta_{22} = \beta_{12} = 0$, y la conclusión es que al menos algunos de esos parámetros son distintos de cero. Esta tabla también muestra la suma de cuadrados para probar la contribución sólo de los términos lineales al modelo [$SS_R(\beta_1, \beta_2 \mid \beta_0) = 918.4$, con dos grados de libertad], y la suma de cuadrados para probar la contribución de los términos cuadráticos, dado que el modelo ya contiene a los términos lineales [$SS_R(\beta_{11}, \beta_{22}, \beta_{12} \mid \beta_0, \beta_1, \beta_2) = 819.2$, con tres grados de libertad]. Al comparar los dos cuadrados medios correspondientes con el cuadrado medio de residuales se obtiene la siguiente estadística F:

$$F_0 = \frac{SS_R(\beta_1, \beta_2 \mid \beta_0)/2}{MS_{Res}} = \frac{914.4/2}{5.89} = \frac{457.2}{5.89} = 77.62$$

http://carlos2524.jimdo.com/

TABLA 7.8 Análisis de varianza para el ejemplo del proceso químico

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Regresión	1733.6	5	346.71	58.86	< 0.0001
$SS_R(\beta_1, \beta_2 \mid \beta_0)$	(914.4)	(2)	(457.20)		
$SS_R(\beta_{11}, \beta_{22}, \beta_{12} \mid \beta_0, \beta_1, \beta_2)$	(819.2)	(3)	(273.10)		
Residuales	35.3	6	5.89		
Falta de ajuste	(8.5)	(3)	(2.83)	0.3176	0.8120
Error puro	(26.8)	(3)	(8.92)		
Total	1768.9	11			
$R^2 = 0.9800$	$R^2_{adj} = 0.9634$			PRESS = 108.7	

para la cual el valor $P = 5.2 \times 10^{-5}$, y

$$F_0 = \frac{SS_R(\beta_{11}, \beta_{22}, \beta_{12} \mid \beta_0, \beta_1, \beta_2)/3}{MS_{Res}} = \frac{819.2/3}{5.89} = \frac{273.1}{5.89} = 46.37$$

para la cual el valor $P = 0.0002$. En consecuencia, tanto los términos lineales como los cuadráticos contribuyen al modelo en forma significativa.

La tabla 7.9 muestra las pruebas t de cada variable individual. Todos los valores t son lo bastante grandes como para que la conclusión sea que no hay términos no significativos en el modelo. Si algunas de esas estadísticas t hubieran sido pequeñas, algunos analistas eliminarían del modelo las variables no significativas y obtendrían un modelo cuadrático reducido para el proceso. En general, nosotros preferimos ajustar el modelo cuadrático completo siempre que sea posible, a menos que haya grandes diferencias entre los modelos completo y reducido en función de PRESS y de R^2 ajustada. La tabla 7.8 indica que los valores de R^2 y R^2 ajustada para este modelo son satisfactorios. La $R^2_{\text{predicción}}$ basada en PRESS es

$$\begin{aligned} R^2_{\text{predicción}} &= 1 - \frac{\text{PRESS}}{SS_T} \\ &= 1 - \frac{108.7}{1768.9} \\ &= 0.9385 \end{aligned}$$

TABLA 7.9 Pruebas con las variables individuales, modelo cuadrático del proceso químico

Variable	Estimado del coeficiente	Error estándar	t para H_0 Coeficiente = 0	Valor P
Ordenada al origen	79.75	1.21	65.72	
x_1	9.83	0.86	11.45	0.0001
x_2	4.22	0.86	4.913	0.0027
x_1^2	− 8.88	0.96	− 9.250	0.0001
x_2^2	− 5.13	0.96	− 5.341	0.0018
x_1x_2	− 7.75	1.21	− 6.386	0.0007

http://carlos2524.jimdo.com/

TABLA 7.10 Valores observados, valores predichos, residuales y otros diagnósticos para el ejemplo del proceso químico

Obs.	Valor real	Valor predicho	Residual	h_{ij}	Residual estudentizado	D de Cook	R de Student
1	43.00	43.96	−0.96	0.625	−0.643	0.115	−0.609
2	78.00	79.11	−1.11	0.625	−0.745	0.154	−0.714
3	69.00	67.89	1.11	0.625	0.748	0.155	0.717
4	73.00	72.04	0.96	0.625	0.646	0.116	0.612
5	48.00	48.11	−0.11	0.625	−0.073	0.001	−0.067
6	76.00	75.90	0.10	0.625	−0.073	0.001	−0.067
7	65.00	63.54	1.46	0.625	0.982	0.268	0.979
8	74.00	75.46	−1.46	0.625	−0.985	0.269	−0.982
9	76.00	79.75	−3.75	0.250	−1.784	0.177	−2.377
10	79.00	79.75	−0.75	0.250	−0.357	0.007	−0.329
11	83.00	79.75	3.25	0.250	1.546	0.133	1.820
12	81.00	79.75	1.25	0.250	0.595	0.020	0.560

que indican que es probable que el modelo explique un gran porcentaje (el 94% más o menos) de la variabilidad con nuevos datos.

La tabla 7.10 contiene los valores observados y predichos de la conversión porcentual, los residuales y otras estadísticas de diagnóstico para este modelo. Ninguno de los residuales estudentizados ni de los valores de R de Student son lo bastante grandes como para indicar que hay problema potencial con valores atípicos. Nótese que las diagonales h_{ii} de la matriz de sombrero sólo asumen dos valores: 0.625 o 0.250. Los valores de $h_{ii} = 0.625$ se asocian con las cuatro corridas en las esquinas del cuadrado, en el diseño, y con las cuatro corridas axiales. Estos ocho puntos están equidistantes al centro del diseño, es la razón por la que todos los valores de h_{ii} son idénticos. Los cuatro puntos centrales tienen $h_{ii} = 0.250$.

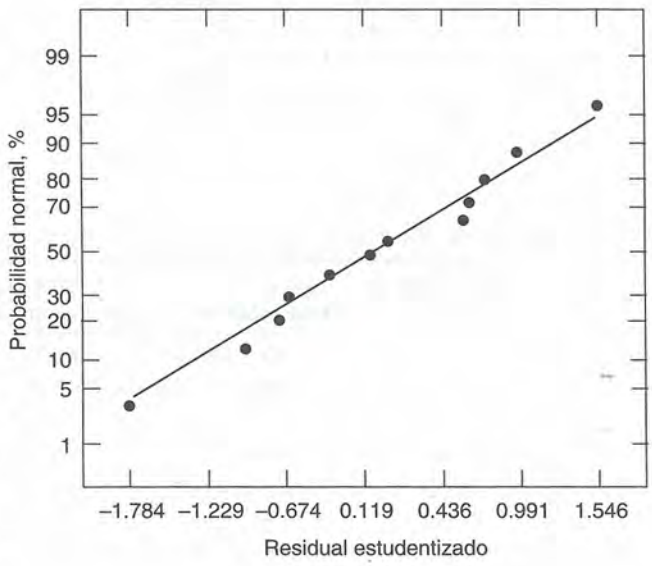


Figura 7.15 Gráfica de probabilidad normal para los residuales estudentizados del ejemplo del proceso químico.

http://carlos2524.jimdo.com/

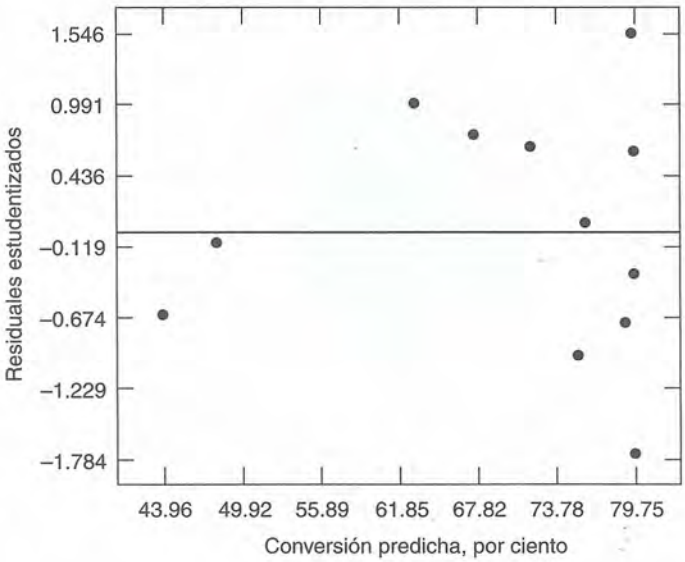


Figura 7.16 Gráfica de residuales estudentizados en función de la conversión predicha, para el ejemplo del proceso químico.

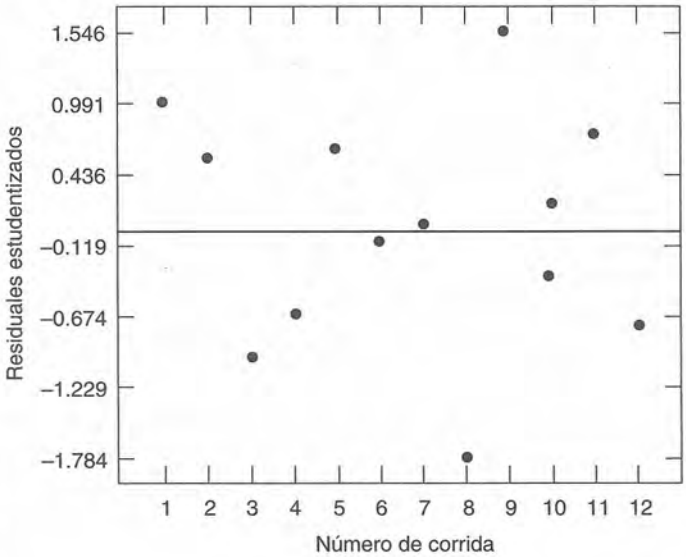


Figura 7.17 Gráfica de los residuales estudentizados en función del número de la corrida.

Las figuras 7.15, 7.16 y 7.17 muestran una gráfica de probabilidad normal de los residuales estudentizados, una gráfica de los residuales estudentizados en función de los valores predichos $\hat{y}_i$, y una gráfica de los residuales estudentizados en función del número progresivo de la corrida. Ninguna de esas gráficas indica inadecuación alguna del modelo.

Las gráficas de la superficie de respuesta de conversión y sus curvas de nivel se ven en las partes *a* y *b* de la figura 7.18, respectivamente; son para el modelo ajustado. Las gráficas de superficie de respuesta indican que la conversión porcentual máxima está a unos 245°C y 20% de concentración.

En muchos problemas de superficie de respuesta a quien hace el experimento le interesa predecir la respuesta *y*, o estimar la respuesta media en determinado punto del espacio de la variable del proceso. Las gráficas de superficie de respuesta de la figura 7.19 son una

http://carlos2524.jimdo.com/

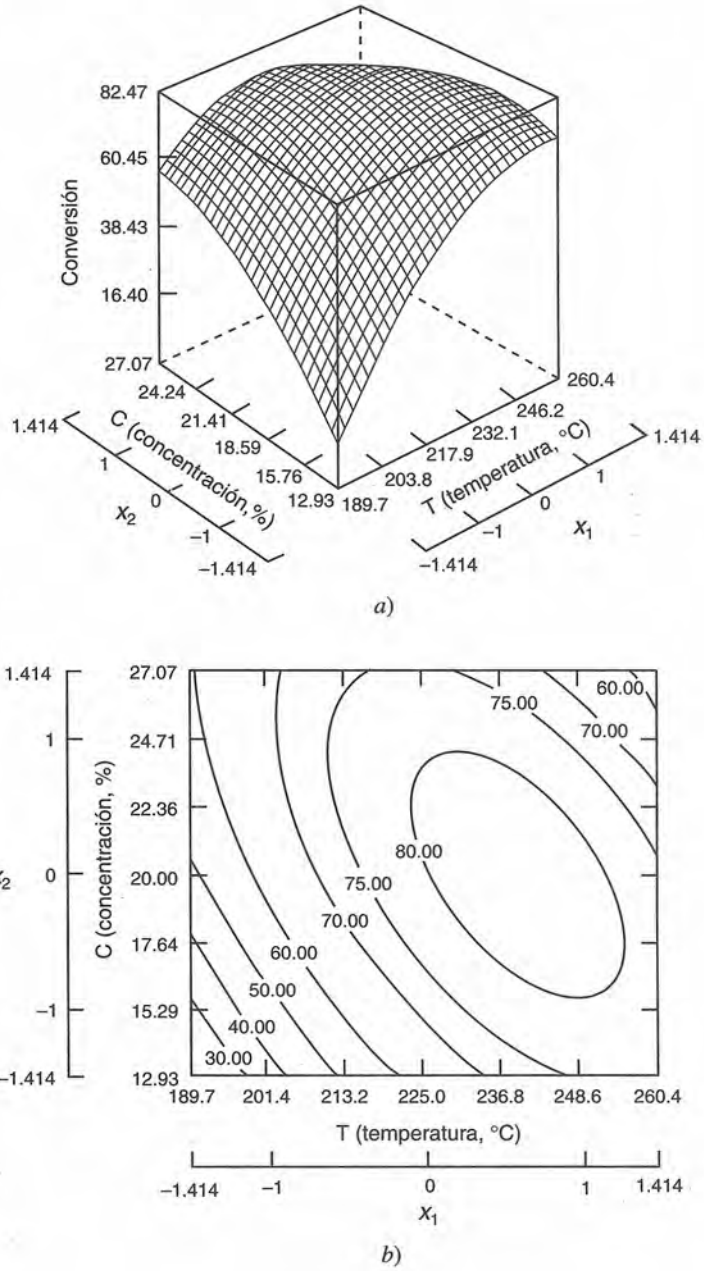
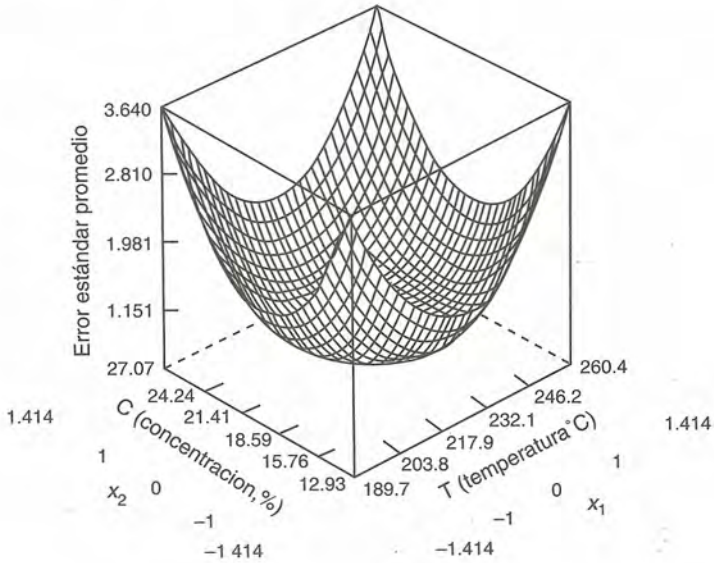


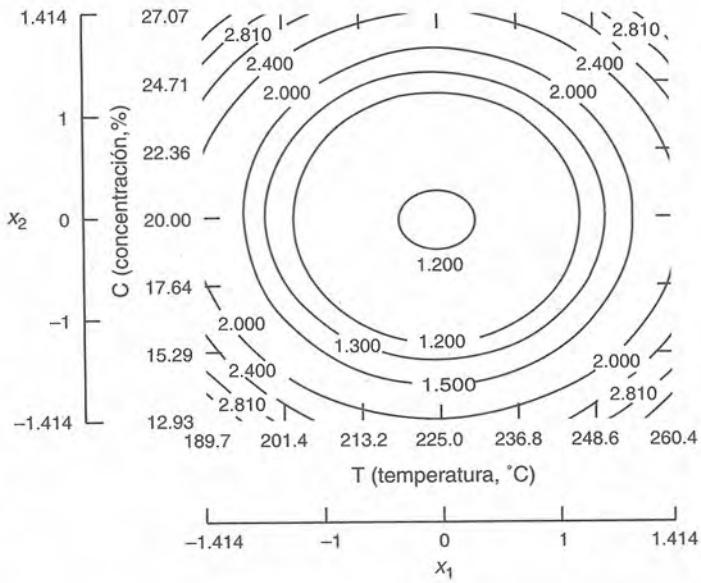
Figura 7.18
a) Superficie de respuesta de la conversión predicha.
b) Gráfica de curvas de nivel de la conversión predicha.

presentación gráfica de esas cantidades. Comúnmente también interesa la varianza de la predicción, porque es una medida directa del error probable asociado con el estimado de punto que produce el modelo. Recuérdese que la varianza del estimado de la respuesta promedio en el punto $\mathbf{x}_0$ se determina con $\text{Var}[\hat{y}(\mathbf{x}_0)] = \sigma^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$. Las gráficas de $\sqrt{\text{Var}[\hat{y}(\mathbf{x}_0)]}$, habiendo estimado σ^2 con el cuadrado medio de residuales $MS_{\text{Res}} = 5.89$ para este modelo, para todos los valores de $\mathbf{x}_0$ en la región de experimentación, se presentan en las partes *a* y *b* de la figura 7.19. Tanto la superficie de respuesta de la figura 7.19a como las curvas de nivel de $\sqrt{\text{Var}[\hat{y}(\mathbf{x}_0)]}$ constante en la figura 7.19b demuestran que la $\sqrt{\text{Var}[\hat{y}(\mathbf{x}_0)]}$ es igual

http://carlos2524.jimdo.com/



a)



b)

Figura 7.19
a) Gráfica de la superficie de respuesta de $\sqrt{\text{Var}[\hat{y}(\mathbf{x}_0)]}$.
b) Gráfica de curvas de nivel de $\sqrt{\text{Var}[\hat{y}(\mathbf{x}_0)]}$.

para todos los puntos $\mathbf{x}_0$ que están a la misma distancia del centro del diseño, lo cual es resultado del espaciado de las corridas axiales en el diseño central compuesto, a 1.414 unidades del origen (en las variables codificadas), y es una propiedad del diseño llamada **rotabilidad**. Es una propiedad muy importante para un diseño de superficie de respuesta de segundo orden, y se describe con detalle en las referencias citadas para RSM (metodología de superficie de respuesta).

http://carlos2524.jimdo.com/

7.5 POLINOMIOS ORTOGONALES

Se ha hecho notar que en el ajuste de modelos polinomiales en una variable, aun cuando se elimine el mal acondicionamiento no esencial mediante el centrado, pueden haber todavía altos niveles de multicolinealidad. Algunas de esas dificultades se pueden eliminar usando **polinomios ortogonales** para ajustar el modelo; supóngase que el modelo es

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{7.8}$$

En general, las columnas de la matriz $\mathbf{X}$ no serán ortogonales, además, si se aumenta el orden del polinomio agregándole un término $\beta_{k+1}x^{k+1}$ se debe calcular $(\mathbf{X}'\mathbf{X})^{-1}$ y cambiarán los estimados de los parámetros de órdenes inferiores $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$.

Ahora, supóngase que se ajusta el modelo

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \alpha_2 P_2(x_i) + \cdots + \alpha_k P_k(x_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{7.9}$$

en donde $P_u(x_i)$ es un polinomio ortogonal de u-ésimo orden, definido de tal modo que

$$\sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0, \quad r \neq s, \quad r, s = 0, 1, 2, \ldots, k$$

$$P_0(x_i) = 1$$

Entonces el modelo se transforma en $\mathbf{y} = \mathbf{X}\alpha + \varepsilon$, siendo la matriz $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} P_0(x_1) & P_1(x_1) & \cdots & P_k(x_1) \\ P_0(x_2) & P_1(x_2) & \cdots & P_k(x_2) \\ \vdots & \vdots & & \vdots \\ P_0(x_n) & P_1(x_n) & \cdots & P_k(x_n) \end{bmatrix}$$

Como esta matriz tiene columnas ortogonales, la matriz $\mathbf{X}'\mathbf{X}$ es

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \sum_{i=1}^{n} P_0^2(x_i) & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^{n} P_1^2(x_i) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sum_{i=1}^{n} P_K^2(x_i) \end{bmatrix}$$

http://carlos2524.jimdo.com/

Los estimadores de α por mínimos cuadrados se calculan a partir de $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ como sigue:

$$\hat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(x_i)y_i}{\sum_{i=1}^{n} P_j^2(x_i)}, \qquad j = 0, 1, \ldots, k \tag{7.10}$$

Como $P_0(x_i)$ es un polinomio de grado cero, se puede igualar $P_0(x_i) = 1$ y en consecuencia

$$\hat{\alpha}_0 = \hat{y}$$

La suma de cuadrados de residuales es

$$SS_{\text{Res}}(k) = SS_{\text{T}} - \sum_{i=1}^{k} \hat{\alpha}_j \left[\sum_{i=1}^{n} P_j(x_i)y_i \right] \tag{7.11}$$

La suma de cuadrados de la regresión, para cualquier parámetro del modelo, no depende de los demás parámetros de éste. Esta suma de cuadrados de regresión es

$$SS_{\text{R}}(\alpha_j) = \hat{\alpha}_j \sum_{i=1}^{n} P_j(x_i)y_i \tag{7.12}$$

Si se desea evaluar el significado del término de orden máximo, se debe probar H_0: $\alpha_k = 0$, esto equivale a probar H_0: $\beta_k = 0$ en la ecuación (7.4); en este caso se usaría

$$F_0 = \frac{SS_{\text{R}}(\alpha_k)}{SS_{\text{Res}}(k)/(n-k-1)} = \frac{\hat{\alpha}_k \sum_{i=1}^{n} P_k(x_i)y_i}{SS_{\text{Res}}(k)/(n-k-1)} \tag{7.13}$$

como el estadístico F. Además, nótese que si cambia el orden del modelo a $k + r$, sólo se deben calcular los r nuevos coeficientes. Los coeficientes $\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_k$ no cambian, por la propiedad de ortogonalidad de los polinomios. Así se facilita el cómputo del ajuste secuencial del modelo.

Los polinomios ortogonales $P_j(x_i)$ se forman con facilidad para el caso en el que los niveles de x tienen espaciamiento igual. Los cinco primeros polinomios ortogonales son

$$P_0(x_i) = 1$$

$$P_1(x_i) = \lambda_1 \left[\frac{x_i - \bar{x}}{d} \right]$$

$$P_2(x_i) = \lambda_2 \left[\left(\frac{(x_i - \bar{x})}{d} \right)^2 - \left(\frac{n^2 - 1}{12} \right) \right]$$

http://carlos2524.jimdo.com/

TABLA 7.11 Resultados de la simulación del inventario, ejemplo 7.5

Cantidad del pedido, x_i	Costo anual promedio, y_i
50	\$335
75	326
100	316
125	313
150	311
175	314
200	318
225	328
250	337
275	345

$$P_3(x_i) = \lambda_3\left[\left(\frac{x_i - \bar{x}}{d}\right)^3 - \left(\frac{x_i - \bar{x}}{d}\right)\left(\frac{3n^2 - 7}{20}\right)\right]$$

$$P_4(x_i) = \lambda_4\left[\left(\frac{x_i - \bar{x}}{d}\right)^4 - \left(\frac{x_i - \bar{x}}{d}\right)^2\left(\frac{3n^2 - 13}{14}\right) + \frac{3(n^2 - 1)(n^2 - 9)}{560}\right]$$

en las que d es el espacio entre los niveles de x, y las $\{\lambda_j\}$ son constantes que se eligen de tal modo que los polinomios tengan valores enteros. En la tabla A.5 del apéndice se presentan algunos valores numéricos de esos polinomios ortogonales. Se pueden encontrar tablas más extensas en DeLury [1960] y en Pearson y Hartley [1966]; también se pueden formar polinomios ortogonales, para usarlos en casos en los que las x no están espaciadas por igual. En Seber [1977, Cap. 8] se presenta una descripción general de métodos para generar polinomios ortogonales.

Ejemplo 7.5 *Polinomios ortogonales*

Un analista de investigación de operaciones ha desarrollado un modelo de cómputo para simulación de un sistema de inventarios con un solo artículo. Ha probado ese modelo para investigar el efecto de diversas cantidades de pedidos sobre el costo promedio anual del inventario. Los datos obtenidos se ven en la tabla 7.11.

Como se sabe que el costo anual promedio del inventario es una función compleja de las cantidades de pedidos, se cree que un polinomio de segundo orden tiene el orden máximo que se debe tener en cuenta, en consecuencia, se ajustará

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \alpha_2 P_2(x_i) + \varepsilon_i, \quad i = 1, 2, \ldots, 10$$

Los coeficientes de los polinomios ortogonales $P_0(x_i)$, $P_1(x_i)$ y $P_2(x_i)$, obtenidos de la tabla A.5 del apéndice, se ven en la tabla 7.12.

http://carlos2524.jimdo.com/

TABLA 7.12 Coeficientes de los polinomios ortogonales para el ejemplo 7.5

i	$P_0(x_i)$	$P_1(x_i)$	$P_2(x_i)$
1	1	−9	6
2	1	−7	2
3	1	−5	−1
4	1	−3	−3
5	1	−1	−4
6	1	1	−4
7	1	3	−3
8	1	5	−1
9	1	7	2
10	1	9	6
	$\sum_{i=1}^{10} P_0^2(x_i) = 10$	$\sum_{i=1}^{10} P_1^2(x_i) = 330$	$\sum_{i=1}^{10} P_2^2(x_i) = 132$
		$\lambda_1 = 2$	$\lambda_2 = 1/2$

Entonces,

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \sum_{i=1}^{10} P_0^2(x_i) & 0 & 0 \\ 0 & \sum_{i=1}^{10} P_1^2(x_i) & 0 \\ 0 & 0 & \sum_{i=1}^{10} P_2^2(x_i) \end{bmatrix} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 330 & 0 \\ 0 & 0 & 132 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} \sum_{i=1}^{10} P_0(x_i)y_i \\ \sum_{i=1}^{10} P_1(x_i)y_i \\ \sum_{i=1}^{10} P_2(x_i)y_i \end{bmatrix} = \begin{bmatrix} 3243 \\ 245 \\ 369 \end{bmatrix}$$

y

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{330} & 0 \\ 0 & 0 & \frac{1}{132} \end{bmatrix} \begin{bmatrix} 3243 \\ 245 \\ 369 \end{bmatrix} = \begin{bmatrix} 324.3000 \\ 0.7424 \\ 2.7955 \end{bmatrix}$$

El modelo ajustado es

$$\hat{y} = 324.30 + 0.7424P_1(x) + 2.7955P_2(x)$$

http://carlos2524.jimdo.com/

TABLA 7.13 Análisis de varianza para el modelo cuadrático del ejemplo 7.5

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Regresión	1213.43	2	606.72	159.24	< 0.0001
Lineal, α_1	(181.89)	1	181.89	47.74	< 0.0002
Cuadrática, α_2	(1031.54)	1	1031.54	270.75	< 0.0001
Residual	26.67	7	3.81		
Total	1240.10	9			

La suma de cuadrados de regresión es

$$\begin{aligned} SS_R(\alpha_1, \alpha_2) &= \sum_{j=1}^{2} \hat{\alpha}_j \left[\sum_{i=1}^{10} P_j(x_i) y_i \right] \\ &= 0.7424(245) + 2.7955(369) \\ &= 181.89 + 1\,031.54 = 1\,213.43 \end{aligned}$$

El análisis de varianza se ve en la tabla 7.13. Tanto los términos lineales como los cuadráticos contribuyen en forma significativa al modelo. Como esos términos explican la mayor parte de la variación en los datos, se adoptará el modelo cuadrático en forma tentativa, a reserva de hacerle un análisis satisfactorio de residuales.

Se puede obtener una ecuación ajustada en términos del regresor original, sustituyendo $P_j(x_j)$ como sigue:

$$\begin{aligned} \hat{y} &= 324.30 + 0.7424P_1(x) + 2.7955P_2(x) \\ &= 324.30 + 0.7424(2)\left(\frac{x - 162.5}{25}\right) \\ &\quad + 2.7955\frac{1}{2}\left[\left(\frac{x - 162.5}{25}\right)^2 - \frac{(10)^2 - 1}{12}\right] \\ &= 312.7686 + 0.0594(x - 162.5) + 0.0022(x - 162.5)^2 \end{aligned}$$

Ésta es la forma de la ecuación que se debe proporcionar al usuario.

PROBLEMAS

7.1 Para los valores de x mostrados a continuación:

$$x = 1.00,\ 1.70,\ 1.25,\ 1.20,\ 1.45,\ 1.85,\ 1.60,\ 1.50,\ 1.95,\ 2.00$$

Supóngase que se desea ajustar un modelo de segundo orden, utilizando estos valores de la variable regresora x. Calcular la correlación entre x y x^2. ¿Se ven algunas dificultades potenciales para ajustar el modelo?

http://carlos2524.jimdo.com/

7.2 Un combustible sólido para cohetes pierde peso después de haber sido producido. Se dispone de los siguientes datos:

Meses después de producido, x	Pérdida de peso, y (kg)
0.25	1.42
0.50	1.39
0.75	1.55
1.00	1.89
1.25	2.43
1.50	3.15
1.75	4.05
2.00	5.15
2.25	6.43
2.50	7.89

a. Ajustar un polinomio de segundo orden que exprese la pérdida de peso en función de la cantidad de meses después de haber sido producido.
b. Pruebe la significancia de la regresión.
c. Pruebe la hipótesis H_0: $\beta_2 = 0$. Comente la necesidad del término cuadrático en este modelo.
d. ¿Hay riesgos potenciales al extrapolar con este modelo?

7.3 Acerca del problema 7.2, calcular los residuales del modelo de segundo orden. Analizar los residuales y comentar la adecuación del modelo.

7.4 Se tienen los datos siguientes:

x	y
4.00	24.60
4.00	24.71
4.00	23.90
5.00	39.50
5.00	39.60
6.00	57.12
6.50	67.11
6.50	67.24
6.75	67.15
7.00	77.87
7.10	80.11
7.30	84.67

a. Ajustar un modelo de polinomio de segundo orden a esos datos.
b. Probar la significancia de la regresión.
c. Probar la falta de ajuste y comentar sobre la adecuación del modelo de segundo orden.
d. Probar la hipótesis H_0: $\beta_2 = 0$. ¿Se puede eliminar el término cuadrático de esta ecuación?

7.5 Para el problema 7.4 calcular los residuales del modelo de segundo orden. Analizar los residuales y llegar a conclusiones acerca de la adecuación del modelo.

http://carlos2524.jimdo.com/

7.6 El grado de carbonatación de una bebida gaseosa se afecta por la temperatura del producto y por la presión de funcionamiento de la llenadora. Se obtuvieron 12 observaciones, y los datos resultantes se presentan a continuación.

Carbonatación, y	Temperatura, x_1	Presión, x_2
2.60	31.0	21.0
2.40	31.0	21.0
17.32	31.5	24.0
15.60	31.5	24.0
16.12	31.5	24.0
5.36	30.5	22.0
6.19	31.5	22.0
10.17	30.5	23.0
2.62	31.0	21.5
2.98	30.5	21.5
6.92	31.0	22.5
7.06	30.5	22.5

a. Ajustar un polinomio de segundo orden.
b. Probar la significancia de la regresión.
c. Probar la falta de ajuste y llegar a conclusiones.
d. ¿Contribuye al modelo el término de interacción, en forma significativa?
e. ¿Contribuyen al modelo los términos de segundo orden, en forma significativa?

7.7 Para el problema 7.6, calcular los residuales del modelo de segundo orden. Analizar esos residuales y comentar la adecuación del modelo.

7.8 Examínense los datos del problema 7.2.
a. Ajustar un modelo de segundo orden a esos datos, usando polinomios ortogonales.
b. Suponer que se desea investigar la adición de un término de tercer orden a este modelo. Comentar la necesidad de este término adicional. Respaldar las conclusiones con un análisis estadístico adecuado.

7.9 Suponer que se trata de ajustar el polinomio cuadrático por tramos, con un nudo en $x = t$:

$$E(y) = S(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{10}(x-t)_+^0 + \beta_{11}(x-t)_+^1 + \beta_{12}(x-t)_+^2$$

a. Demostrar cómo se debe probar la hipótesis de que este modelo de spline cuadrática se ajusta a los datos, en forma significativamente mejor que un polinomio cuadrático ordinario.
b. El modelo de spline cuadrática no es continuo en el nudo t. ¿Se puede modificar el modelo para obtener continuidad en $x = t$?
c. Indicar cómo se puede modificar el modelo para que tanto $E(y)$ como $dE(y)/dx$ sean continuas en $x = t$.
d Comentar el significado de las restricciones de continuidad para el modelo, en las partes b y c. En la práctica, ¿cómo se seleccionaría el tipo de restricciones de continuidad que se van a imponer?

http://carlos2524.jimdo.com/

7.10 Se trata de ajustar un modelo polinomial por segmentos con tres segmentos: si $x < t_1$, el polinomio es lineal; si $t_1 \leq x < t_2$, el polinomio es cuadrático, y si $x > t_2$, el polinomio es lineal. Considérese el modelo

$$\begin{aligned} E(y) = S(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{10}(x - t_1)_+^0 + \beta_{11}(x - t_1)_+^1 \\ + \beta_{12}(x - t_1)_+^2 + \beta_{20}(x - t_2)_+^0 + \beta_{21}(x - t_2)_+^1 \\ + \beta_{22}(x - t_2)_+^2 \end{aligned}$$

a. ¿Satisface los requisitos este polinomio segmentado? Si no es así, indicar cómo se puede modificar para que los satisfaga.

b. Indicar cómo se modificaría el modelo segmentado para asegurar que $E(y)$ sea continua en los nudos t_1 y t_2.

c. Indicar cómo se modificaría el modelo segmentado para asegurar que tanto $E(y)$ como $dE(y)/dx$ sean continuas en los nudos t_1 y t_2.

7.11 Un analista de investigación de operaciones estudia la relación entre el tamaño del lote de producción, x, y el costo unitario promedio de producción, y. Un estudio de las operaciones recientes produjo los siguientes datos:

x	100	120	140	160	180	200	220	240	260	280	300
y	\$9.73	9.61	8.15	6.98	5.87	4.98	5.09	4.79	4.02	4.46	3.82

El analista cree que un modelo de regresión lineal por segmentos debe ajustar a esos datos. Estimar los parámetros de ese modelo, suponiendo que la pendiente de la recta cambia en $x = 200$ unidades. ¿Respaldan los datos el uso de ese modelo?

7.12 Modificar el modelo del problema 7.11 para investigar la posibilidad de que exista una discontinuidad en la función de regresión en $x = 200$ unidades. Estimar los parámetros en este modelo. Probar las hipótesis adecuadas para determinar si la función de regresión cambia, tanto de pendiente como de ordenada, en $x = 200$ unidades.

7.13 Para el modelo polinomial del problema 7.11, calcular los factores de inflación de varianza, y comentar sobre la multicolinealidad en este modelo.

7.14 Se tienen los datos del problema 7.2.

a. Ajustar un modelo de segundo orden $y = \beta_0 + \beta_1 x + \beta_{11}x^2 + \varepsilon$ a esos datos. Evaluar los factores de inflación de la varianza.

b. Ajustar un modelo de segundo orden $y = \beta_0 + \beta_1(x - \bar{x}) + \beta_{11}(x - \bar{x})^2 + \varepsilon$ a los datos. Evaluar los factores de inflación de la varianza.

c. ¿Cuál sería la conclusión acerca del impacto, sobre la multicolinealidad, de centrar las x en un modelo polinomial?

7.15 Con frecuencia, en ingeniería química y mecánica se necesita conocer la presión de vapor de agua a diversas temperaturas; para ello se pueden usar las "infames" tablas de vapor. A continuación se presentan datos de la presión de vapor y del agua a diversas temperaturas.

http://carlos2524.jimdo.com/

y = presión de vapor (mm Hg)	x = Temperatura (°C)
9.2	10
17.5	20
31.8	30
55.3	40
92.5	50
149.4	60

a. Ajustar un modelo de primer orden a los datos. Sobreponer el modelo ajustado al diagrama de dispersión de y en función de x. Comentar el ajuste aparente del modelo.

b. Preparar un diagrama de dispersión de valores ajustados de y en función de los valores de y observada. ¿Qué parece indicar ese diagrama acerca del ajuste del modelo?

c. Graficar los residuales en función de la y ajustada o predicha. Comentar la adecuación del modelo.

d. Ajustar un modelo de segundo orden a los datos. ¿Hay evidencia de que el término cuadrático sea estadísticamente importante?

e. Repetir las partes de a a c, usando el modelo de segundo orden. ¿Hay pruebas de que el modelo de segundo orden produce un mejor ajuste para los datos de presión de vapor?

7.16 Un artículo en la revista *Journal of Pharmaceutical Sciences* (**80**, 971-977, 1991) presenta datos sobre la solubilidad observada, en fracción molar, de un soluto a temperatura

Observación número	y	x_1	x_2	x_3
1	0.22200	7.3	0.0	0.0
2	0.39500	8.7	0.0	0.3
3	0.42200	8.8	0.7	1.0
4	0.43700	8.1	4.0	0.2
5	0.42800	9.0	0.5	1.0
6	0.46700	8.7	1.5	2.8
7	0.44400	9.3	2.1	1.0
8	0.37800	7.6	5.1	3.4
9	0.49400	10.0	0.0	0.3
10	0.45600	8.4	3.7	4.1
11	0.45200	9.3	3.6	2.0
12	0.11200	7.7	2.8	7.1
13	0.43200	9.8	4.2	2.0
14	0.10100	7.3	2.5	6.8
15	0.23200	8.5	2.0	6.6
16	0.30600	9.5	2.5	5.0
17	0.09230	7.4	2.8	7.8
18	0.11600	7.8	2.8	7.7
19	0.07640	7.7	3.0	8.0
20	0.43900	10.3	1.7	4.2
21	0.09440	7.8	3.3	8.5
22	0.11700	7.1	3.9	6.6
23	0.07260	7.7	4.3	9.5
24	0.04120	7.4	6.0	10.9
25	0.25100	7.3	2.0	5.2
26	0.00002	7.6	7.8	20.7

http://carlos2524.jimdo.com/

constante, junto con x_1 = solubilidad parcial de la dispersión, x_2 = solubilidad parcial dipolar y x_3 = solubilidad parcial de Hansen por puentes de hidrógeno. La respuesta y es el logaritmo negativo de la solubilidad en fracción mol.

a. Ajustar un modelo cuadrático completo a los datos.
b. Probar el significado de la regresión y formar las estadísticas t para cada parámetro del modelo. Interpretar esos resultados.
c. Graficar los residuales y comentar la adecuación del modelo.
d. Usar el método de suma extra de cuadrados para probar la contribución de todos los términos de segundo orden al modelo.

7.17 Véase el modelo cuadrático de regresión del problema 7.16. Determinar los factores de inflación de varianza, y comentar la multicolinealidad en ese modelo.

7.18 Para los datos de solubilidad del problema 7.16, suponer que un punto de interés es x_1 = 8.0, x_2 = 3.0 y x_3 = 5.0.

a. Para el modelo cuadrático del problema 7.16, pronosticar la respuesta en el punto de interés y determinar un intervalo de confianza de 95% para la respuesta media en ese punto.
b. Ajustar un modelo que sólo incluya los efectos principales, e interacciones de dos factores, de los datos de solubilidad. Usar este modelo para predecir la respuesta en el punto de interés. Determinar un intervalo de confianza de 95% para la respuesta media en ese punto.
c. Comparar las longitudes de los intervalos de confianza en las partes a y b. ¿Puede el lector llegar a conclusiones acerca del óptimo modelo de acuerdo con esta comparación?

7.19 A continuación se presentan datos de y = concentración de licor verde (g/l) y x = velocidad de la máquina de papel (pies/min) en una fábrica de papel kraft. Los datos se tomaron de una gráfica, de un artículo en la revista *Tappi Journal*, marzo de 1986.

y	16.0	15.8	15.6	15.5	14.8
x	1700	1720	1730	1740	1750
y	14.0	13.5	13.0	12.0	11.0
x	1760	1770	1780	1790	1795

a. Ajustar el modelo $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ a los datos.
b. Probar el significado de la regresión con $\alpha = 0.05$. ¿Cuáles son sus conclusiones?
c. Probar la contribución del término cuadrático al modelo, y la del término lineal, usando una dócima F. Si $\alpha = 0.05$, ¿qué conclusión se puede sacar?
d. Traficar los residuales del modelo. ¿Parece satisfactorio el ajuste del modelo?

7.20 Reconsiderar los datos del problema 7.19. Suponer que es importante predecir la respuesta en los puntos x = 1 750 y x = 1 775.

a. Calcular la respuesta predicha en esos puntos, y los intervalos de predicción de 95% para la respuesta observada en el futuro, en esos puntos.
b. Suponer que también se considera un modelo de primer orden. Ajustar ese modelo y determinar la respuesta predicha en esos puntos. Calcular los intervalos de predicción de 95% para la respuesta futura observada en esos puntos. ¿Permite lo anterior decidir cuál modelo se debe preferir?

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

CAPÍTULO 8

VARIABLES INDICADORAS

8.1 EL CONCEPTO GENERAL DE LAS VARIABLES INDICADORAS

Las variables que se emplean en el análisis de regresión se suelen llamar **variables cuantitativas**, lo que significa que las variables tienen una escala bien definida de medición. Las variables como temperatura, distancia, presión e ingreso son variables cuantitativas, en ocasiones es necesario usar variables **cualitativas** o **categóricas** como variables predictoras en la regresión. Como ejemplos de variables cualitativas o categóricas están los operadores, estado de empleo (empleado o desempleado), los turnos (diurno, mixto o nocturno) y el sexo (masculino o femenino). En general, una variable cualitativa no tiene escala natural de medida. Se debe asignar un conjunto de niveles a una variable cualitativa para tener en cuenta el efecto que pueda tener la variable sobre la respuesta, esto se hace usando **variables indicadoras;** a veces, a las variables indicadoras se les llama **variables ficticias**.

Supóngase que un ingeniero mecánico desea relacionar la vida útil y de una herramienta (un *buril*) en un torno, con la velocidad del torno, en revoluciones por minuto (x_1), y con la clase de buril que se usa. La segunda variable regresora, la clase de buril, es cualitativa y puede tener dos niveles (por ejemplo, los tipos A y B de buril). Se usará una variable indicadora que asuma los valores 0 y 1 para identificar las clases de la variable regresora "tipo de buril", sea

$$x_2 = \begin{cases} 0 \text{ si la observación procede de la herramienta tipo A} \\ 1 \text{ si la observación procede de la herramienta tipo B} \end{cases}$$

La elección de 0 y 1 para identificar los niveles de una variable cualitativa es arbitraria. Dos valores cualesquiera distintos para x_2 serían satisfactorios, aunque por lo general los mejores son 0 y 1.

Suponiendo que es adecuado un modelo de primer orden, se tiene

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{8.1}$$

Para interpretar los parámetros en este modelo, se examinará el primer tipo de buril, el A, para el cual $x_2 = 0$. El modelo de regresión se transforma en

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2(0) + \varepsilon \\ &= \beta_0 + \beta_1 x_1 + \varepsilon \end{aligned} \tag{8.2}$$

Así, la relación entre la vida útil y la velocidad del torno para el buril tipo A es una recta con ordenada al origen β_0 y pendiente β_1. Para el tipo de herramienta B, $x_2 = 1$ y

http://carlos2524.jimdo.com/

$$y = \beta_0 + \beta_1 x_1 + \beta_2(1) + \varepsilon$$
$$= (\beta_0 + \beta_2) + \beta_1 x_1 + \varepsilon \tag{8.3}$$

Esto es, para el buril de tipo B, la relación entre la vida de la herramienta y la velocidad del torno también es una recta β_1, pero con ordenada al origen $\beta_0 + \beta_2$.

Las dos funciones de respuesta se ven en la figura 8.1. Los modelos (8.2) y (8.3) describen dos líneas de regresión **paralelas**, esto es, dos rectas con una pendiente común β_1 y distintas ordenadas al origen. También, se supone que la varianza de los errores ε es igual para ambos tipos de buril, A y B. El parámetro β_2 expresa la diferencia de alturas entre las dos líneas de regresión, ya que, β_2 es una medida de la diferencia de vida media de herramienta que resulta de cambiar del tipo A al tipo B.

Se puede generalizar este método a factores cualitativos con cualquier cantidad de niveles, por ejemplo, supóngase que interesan tres tipos de herramienta: A, B y C, se requieren dos variables indicadoras, como x_2 y x_3, para manejar los tres niveles de tipo de herramienta en el modelo. Los niveles de las variables indicadoras son

x_2	x_3	
0	0	si la observación procede del buril tipo A
1	0	si la observación procede del buril tipo B
0	1	si la observación procede del buril tipo C

y el modelo de regresión es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

En general, una variable cualitativa con a niveles se representa con $a - 1$ variables indicadoras, y cada una asume los valores 0 y 1.

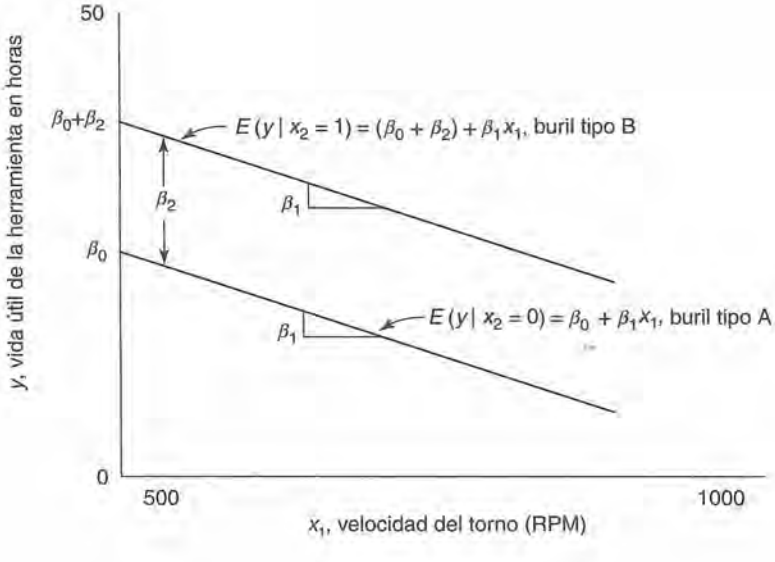


Figura 8.1 Funciones de respuesta para el ejemplo de la vida útil del buril.

http://carlos2524.jimdo.com/

Ejemplo 8.1 *Datos de vida de herramienta*

En la tabla 8.1 se presentan 20 observaciones de duración de herramienta y velocidad del torno, y el diagrama de dispersión se ve en la figura 8.2. Al revisar este diagrama de dispersión se ve que se requieren dos líneas de regresión para modelar bien estos datos, y que la ordenada al origen depende del tipo de herramienta que se usa; en consecuencia, se ajustará el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

TABLA 8.1 Datos, valores ajustados y residuales para el ejemplo 8.1

i	y_i(Horas)	x_{i1} (rpm)	Tipo de herramienta	$\hat{y}_i$	e_i
1	18.73	610	A	20.7552	−2.0252
2	14.52	950	A	11.7087	2.8113
3	17.43	720	A	17.8284	−0.3984
4	14.54	840	A	14.6355	−0.0955
5	13.44	980	A	10.9105	2.5295
6	24.39	530	A	22.8838	1.5062
7	13.34	680	A	18.8927	−5.5527
8	22.71	540	A	22.6177	0.0923
9	12.68	890	A	13.3052	−0.6252
10	19.32	730	A	17.5623	1.7577
11	30.16	670	B	34.1630	−4.0030
12	27.09	770	B	31.5023	−4.4123
13	25.40	880	B	28.5755	−3.1755
14	26.05	1000	B	25.3826	0.6674
15	33.49	760	B	31.7684	1.7216
16	35.62	590	B	36.2916	−0.6716
17	26.07	910	B	27.7773	−1.7073
18	36.78	650	B	34.6952	2.0848
19	34.95	810	B	30.4380	4.5120
20	43.67	500	B	38.6862	4.9838

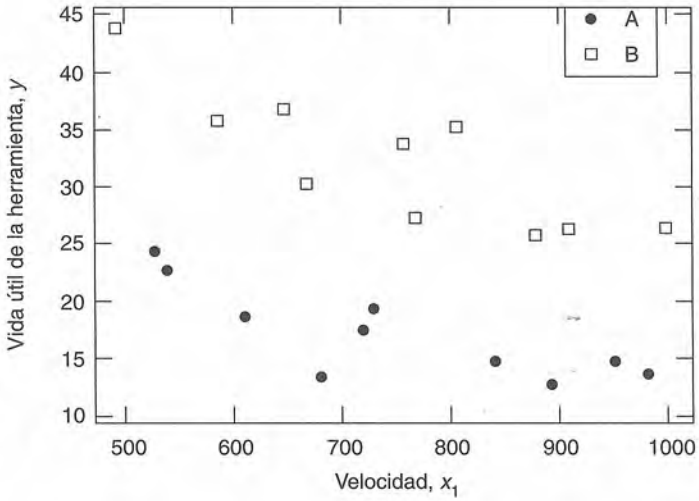


Figura 8.2 Gráfica de la vida útil de la herramienta y en función de la velocidad del torno x_1, para los tipos de buril A y B.

http://carlos2524.jimdo.com/

en donde la variable indicadora $x_2 = 0$ si la observación procede de la herramienta tipo A, y $x_2 = 1$ si procede de la herramienta tipo B. La matriz **X** y el vector **y** para ajustar este modelo son

$$\mathbf{X} = \begin{bmatrix} 1 & 610 & 0 \\ 1 & 950 & 0 \\ 1 & 720 & 0 \\ 1 & 840 & 0 \\ 1 & 980 & 0 \\ 1 & 530 & 0 \\ 1 & 680 & 0 \\ 1 & 540 & 0 \\ 1 & 890 & 0 \\ 1 & 730 & 0 \\ 1 & 670 & 1 \\ 1 & 770 & 1 \\ 1 & 880 & 1 \\ 1 & 1000 & 1 \\ 1 & 760 & 1 \\ 1 & 590 & 1 \\ 1 & 910 & 1 \\ 1 & 650 & 1 \\ 1 & 810 & 1 \\ 1 & 500 & 1 \end{bmatrix} \quad \text{y} \quad \mathbf{y} = \begin{bmatrix} 18.73 \\ 14.52 \\ 17.43 \\ 14.54 \\ 13.44 \\ 24.39 \\ 13.34 \\ 22.71 \\ 12.68 \\ 19.32 \\ 30.16 \\ 27.09 \\ 25.40 \\ 26.05 \\ 33.49 \\ 35.62 \\ 26.07 \\ 36.78 \\ 34.95 \\ 43.67 \end{bmatrix}$$

El ajuste por mínimos cuadrados es

$$\hat{y} = 36.986 - 0.027x_1 + 15.004x_2$$

En la tabla 8.2 se presenta el análisis de varianza y otras estadísticas de resumen para este modelo. Como el valor observado de F_0 tiene un valor P muy pequeño, se rechaza la hipótesis de significancia de la regresión, y en vista de que las estadísticas t para β_1 y β_2 tienen valores P pequeños, se llega a la conclusión de que los dos regresores, x_1 (rpm) y x_2 (tipo de herramienta) contribuyen al modelo. El parámetro β_2 es el cambio en la duración

TABLA 8.2 Estadísticas de resumen para el modelo de regresión del ejemplo 8.1

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Regresión	1418.034	2	709.017	76.75	3.12×10^{-9}
Residual	157.055	17	9.239		
Total	1575.089	19			

Coeficiente	Estimado	Error estándar	t_0	Valor P
β_0	36.986			
β_1	−0.027	0.005	−5.887	8.97×10^{-6}
β_2	15.004	1.360	11.035	1.79×10^{-9}
	$R^2 = 0.9003$			

http://carlos2524.jimdo.com/

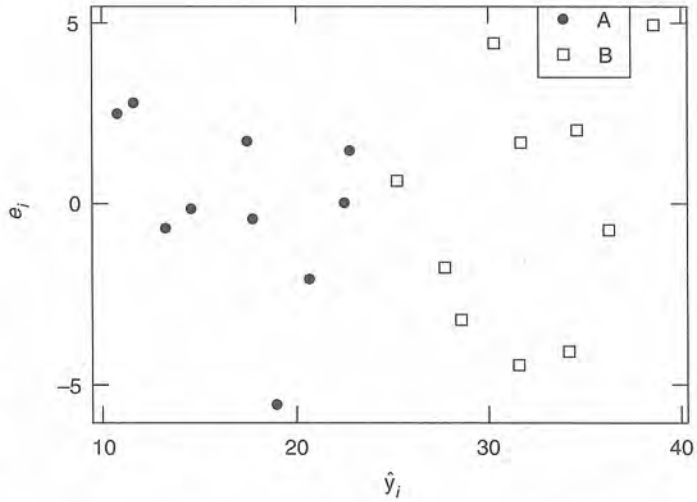


Figura 8.3 Gráfica de residuales e_i en función de valores ajustados $\hat{y}_i$, ejemplo 8.1.

promedio de la herramienta, debido a un cambio del tipo A al tipo B de herramienta. El intervalo de confianza de 95% para β_2 es

$$\hat{\beta}_2 - t_{0.025,\,17}\text{se}\left(\hat{\beta}_2\right) \le \beta_2 \le \hat{\beta}_2 + t_{0.025,\,17}\text{se}\left(\hat{\beta}_2\right)$$
$$= 15.004 - 2.110(1.360) \le \beta_2 \le 15.004 + 2.110(1.360)$$

o sea que

$$12.135 \le \beta_2 \le 17.873$$

En vista de lo anterior, se tiene el 95% de confianza que al cambiar del tipo de buril A al B aumenta la duración de la herramienta entre 12.135 horas y 17.873 horas.

Los valores ajustados $\hat{y}_i$ y los residuales e_i de este modelo se ven en las dos últimas columnas de la tabla 8.1. En la figura 8.3 se muestra la gráfica de los residuales en función de $\hat{y}_i$; en esta gráfica, se identifican por tipo de herramienta (A o B); si la varianza de los errores no fuera la misma para los dos tipos, se vería en esta gráfica. Nótese que los residuales "B" en la figura 8.3 muestran un poco más dispersión que los "A", y eso implica que puede haber un problema moderado de desigualdad de varianza. La figura 8.4 es la gráfica de probabilidad normal de los residuales. No hay indicios de inadecuación grave del modelo.

En vista de que se emplean dos líneas de regresión distintas para modelar la relación entre la vida útil del buril y la velocidad del torno, desde el principio se podrían haber ajustado dos modelos separados rectilíneos, en lugar de uno solo con una variable indicadora. Sin embargo, se prefiere el método con un solo modelo, porque el analista sólo tiene una ecuación final con la que trabajar, y no dos; es un resultado práctico mucho más simple, además, como se supone que las dos rectas tienen la misma pendiente, tiene sentido combinar los datos de ambos tipos para producir un solo estimado de este parámetro común; este método también proporciona una estimación de la varianza común del error σ^2, y se tienen más grados de libertad que los que resultarían de ajustar dos líneas separadas de regresión.

Ahora supóngase que se espera que las rectas de regresión que relacionan la vida útil del buril con la velocidad del torno difieren tanto en la ordenada al origen como en

http://carlos2524.jimdo.com/

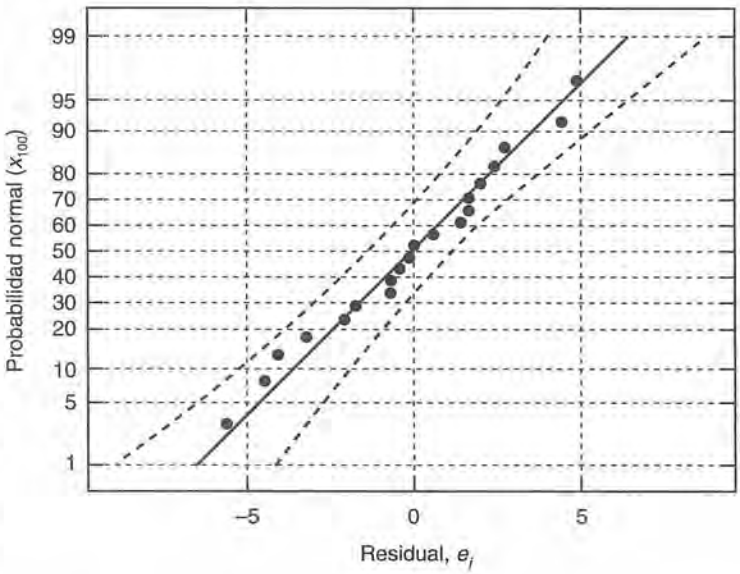


Figura 8.4 Gráfica de probabilidad normal de los residuales, ejemplo 8.1.

la pendiente. Es posible modelar este caso con una sola ecuación de regresión, usando variables indicadoras. El modelo es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon \tag{8.4}$$

Al comparar la ecuación (8.4) con la (8.1) se observa que se agregó al modelo un producto cruzado entre x_1, la velocidad del torno y la variable indicadora que representa el tipo de buril, x_2. Para interpretar los parámetros en este modelo, se examinará primero la herramienta tipo A, para la cual $x_2 = 0$. El modelo (8.4) se transforma en

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2(0) + \beta_3 x_1(0) + \varepsilon \\ &= \beta_0 + \beta_1 x_1 + \varepsilon \end{aligned} \tag{8.5}$$

que es una recta con ordenada al origen β_0 y pendiente β_1. Para el buril tipo B, $x_2 = 1$ y

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2(1) + \beta_3 x_1(1) + \varepsilon \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_1 + \varepsilon \end{aligned} \tag{8.6}$$

Es un modelo rectilíneo con ordenada al origen $\beta_0 + \beta_2$ y pendiente $\beta_1 + \beta_3$. Las dos funciones de regresión se grafican en la figura 8.5. Nótese que la ecuación (8.4) define dos rectas de regresión con distintas pendientes y ordenadas al origen. En consecuencia, el parámetro β_2 refleja el cambio de ordenada al origen asociado con el cambio de buril tipo A a buril tipo B (las clases 0 y 1 de la variable indicadora x_2), y β_3 indica el cambio de pendiente asociado con el cambio de tipos de herramienta, de A a B.

El ajuste del modelo (8.4) equivale a ajustar dos ecuaciones de regresión separadas. Una ventaja del uso de variables indicadoras es que las pruebas de hipótesis se pueden hacer en forma directa, con el método de la suma extra de cuadrados. Por ejemplo, para probar si los dos modelos de regresión son idénticos, las hipótesis serían

$$H_0\colon \beta_2 = \beta_3 = 0$$
$$H_1\colon \beta_2 \neq 0 \text{ y/o } \beta_3 \neq 0.$$

http://carlos2524.jimdo.com/

Si no se rechaza H_0: $\beta_2 = \beta_3 = 0$, entonces un solo modelo de regresión puede explicar la relación entre la vida útil de la herramienta y la velocidad del torno. Para probar si las dos rectas de regresión tienen la misma pendiente pero quizá distintas ordenadas al origen, las hipótesis son

$$H_0: \beta_3 = 0$$
$$H_1: \beta_3 \neq 0$$

Si se usa el modelo (8.4), las dos rectas de regresión se pueden ajustar, y se pueden hacer esas pruebas en una sola corrida de computadora, siempre que el programa produzca las sumas de cuadrados $SS_R(\beta_1|\beta_0)$, $SS_R(\beta_2|\beta_0, \beta_1)$ y $SS_R(\beta_3|\beta_0, \beta_1, \beta_2)$.

Ejemplo 8.2 *Datos de duración de herramienta*

Se ajustará el modelo de regresión

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

a los datos de vida útil de herramienta de la tabla 8.1. La matriz **X** y el vector **y** para e modelo son:

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccc} & x_1 & x_2 & x_1x_2 \end{array} \\ \begin{bmatrix} 1 & 610 & 0 & 0 \\ 1 & 950 & 0 & 0 \\ 1 & 720 & 0 & 0 \\ 1 & 840 & 0 & 0 \\ 1 & 980 & 0 & 0 \\ 1 & 530 & 0 & 0 \\ 1 & 680 & 0 & 0 \\ 1 & 540 & 0 & 0 \\ 1 & 890 & 0 & 0 \\ 1 & 730 & 0 & 0 \\ 1 & 670 & 1 & 670 \\ 1 & 770 & 1 & 770 \\ 1 & 880 & 1 & 880 \\ 1 & 1000 & 1 & 1000 \\ 1 & 760 & 1 & 760 \\ 1 & 590 & 1 & 590 \\ 1 & 910 & 1 & 910 \\ 1 & 650 & 1 & 650 \\ 1 & 810 & 1 & 810 \\ 1 & 500 & 1 & 500 \end{bmatrix} \end{array} \quad \text{y} \quad \mathbf{y} = \begin{bmatrix} 18.73 \\ 14.52 \\ 17.43 \\ 14.54 \\ 13.44 \\ 24.39 \\ 12.54 \\ 22.71 \\ 12.68 \\ 19.32 \\ 30.16 \\ 27.09 \\ 25.40 \\ 26.05 \\ 33.49 \\ 35.62 \\ 26.07 \\ 36.78 \\ 34.95 \\ 43.67 \end{bmatrix}$$

El modelo de regresión ajustado es

$$\hat{y} = 32.775 - 0.021\, x_1 + 23.971 x_2 - 0.012 x_1 x_2$$

En la tabla 8.3 se ve el resumen de análisis para este modelo. Para probar la hipótesis que las dos rectas de regresión son idénticas (H_0: $\beta_2 = \beta_3 = 0$), se usa la estadística

$$F_0 = \frac{SS_R(\beta_2, \beta_3 | \beta_1, \beta_0)/2}{MS_{Res}}$$

http://carlos2524.jimdo.com/

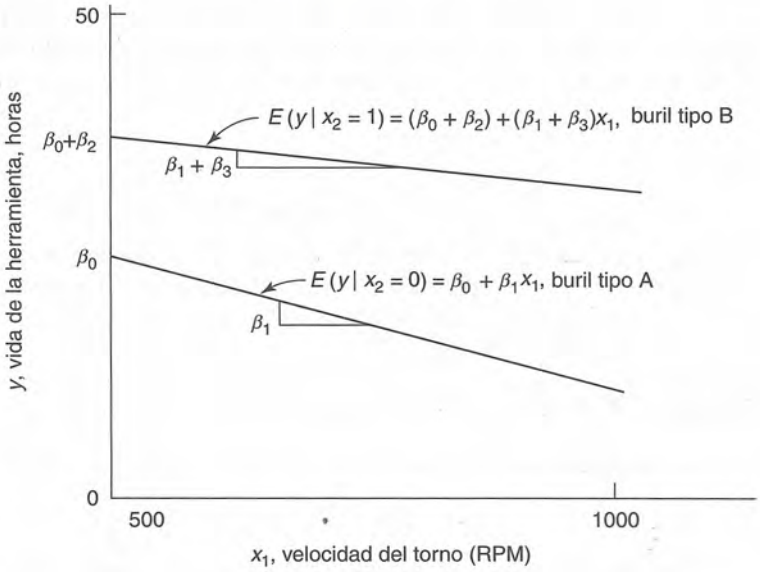


Figura 8.5 Funciones de respuesta para la ecuación (8.4).

En vista de que

$$\begin{aligned} SS_R(\beta_2, \beta_3 | \beta_1, \beta_0) &= SS_R(\beta_1, \beta_2, \beta_3 | \beta_0) - SS_R(\beta_1 | \beta_0) \\ &= 1434.112 - 293.005 \\ &= 1141.107 \end{aligned}$$

la estadística de prueba es

$$F_0 = \frac{SS_R(\beta_2, \beta_3 | \beta_1, \beta_0)/2}{MS_{Res}} = \frac{1141.107/2}{8.811} = 64.75$$

TABLA 8.3 Estadísticas de resumen para el modelo de regresión de la vida útil de herramientas, en el ejemplo 8.2

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor F
Tratamientos	1434.112	3	478.037	54.25	1.32×10^{-9}
Error	140.976	16	8.811		
Total	1575.088	19			

Coeficiente	Estimado	Error estándar	t_0	Suma de cuadrados
β_0	32.775			
β_1	–0.021	0.0061	–3.45	$SS_R(\beta_1 \mid \beta_0) = 293.005$
β_2	23.971	6.7690	3.54	$SS_R(\beta_2 \mid \beta_1, \beta_0) = 1125.029$
β_3	–0.012	0.0088	–1.35	$SS_R(\beta_3 \mid \beta_2, \beta_1, \beta_0) = 16.078$
		$R^2 = 0.9105$		

http://carlos2524.jimdo.com/

y como el valor P para esta estadística es $P = 2.14 \times 10^{-8}$, se concluye que los dos modelos de regresión no son idénticos. Para probar la hipótesis que las dos rectas tienen distintas ordenadas al origen y una pendiente común (H_0: $\beta_3 = 0$) se usa la estadística

$$F_0 = \frac{SS_R(\beta_3|\beta_2, \beta_1, \beta_0)/1}{MS_{Res}} = \frac{16.078}{8.811} = 1.82$$

y como el valor P para esta estadística es $P = 0.20$, se llega a la conclusión de que las pendientes de las dos rectas son iguales. Esto también se puede determinar con la estadística t para β_2 y β_3 en la tabla 8.3.

Las variables indicadoras son útiles en diversos casos de regresión. Ahora se presentarán más aplicaciones características de ellas.

Ejemplo 8.3 Una variable indicadora con más de dos niveles

Una empresa eléctrica está investigando el efecto que tiene el tamaño de la vivienda familiar y el tipo del acondicionamiento de aire que se usa en ella, sobre el consumo total de electricidad durante los meses calurosos. Sea y el consumo eléctrico total (en kilowatt-horas), durante el periodo de junio a septiembre, y x_1 el tamaño de la casa (pies cuadrados de construcción). Hay cuatro tipos de sistemas de acondicionamiento de aire: 1) sin acondicionamiento, 2) unidades de ventana, 3) bomba térmica, y 4) acondicionamiento central. Los cuatro niveles de ese factor se pueden modelar con tres variables indicadoras, x_2, x_3 y x_4, que se definen como sigue:

Tipo de acondicionamiento de aire	x_2	x_3	x_4
Sin acondicionamiento de aire	0	0	0
Unidades de ventana	1	0	0
Bomba térmica	0	1	0
Acondicionamiento central de aire	0	0	1

El modelo de regresión es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \tag{8.7}$$

Si la casa no tiene acondicionamiento de aire, la ecuación (8.7) se transforma en

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Si la casa tiene unidades de ventanas, entonces

$$y = (\beta_0 + \beta_2) + \beta_1 x_1 + \varepsilon$$

Si la casa tiene bomba térmica, el modelo de regresión es

$$y = (\beta_0 + \beta_3) + \beta_1 x_1 + \varepsilon$$

http://carlos2524.jimdo.com/

y si la casa tiene acondicionamiento central, entonces

$$y = (\beta_0 + \beta_4) + \beta_1 x_1 + \varepsilon$$

Así, en el modelo (8.7) se supone que la relación entre el consumo eléctrico en tiempo caluroso, y el tamaño de la casa es lineal, y que la pendiente no depende del tipo de sistema de acondicionamiento de aire que se emplee. Los parámetros β_2, β_3 y β_4 modifican la altura (u ordenada al origen) del modelo de regresión para los distintos sistemas de acondicionamiento de aire. Esto es, β_2, β_3 y β_4 miden el efecto de las unidades de ventana, de bomba térmica y de acondicionamiento central, respectivamente, en comparación con la falta de acondicionamiento de aire; además, se pueden determinar otros efectos comparando en forma directa los coeficientes adecuados de regresión. Por ejemplo, $\beta_3 - \beta_4$ refleja la eficiencia relativa de una bomba térmica respecto al acondicionamiento central de aire, también nótese la hipótesis que la varianza del consumo de energía no depende del tipo de sistema de acondicionamiento usado; esta hipótesis puede ser inadecuada.

En este problema parece irreal suponer que la pendiente de la función de regresión que relaciona el consumo eléctrico medio con el tamaño de la vivienda no depende del tipo de sistema de acondicionamiento de aire. Por ejemplo, cabría esperar que el consumo eléctrico medio aumentará al aumentar el tamaño de la casa, pero la tasa de aumento debería ser distinta para un sistema central de acondicionamiento de aire que para las unidades de ventana, porque el primero debería ser más eficiente que las unidades de ventana para las casas más grandes. Esto es, debería haber una **interacción** entre el tamaño de la casa y la clase de sistema de acondicionamiento. Esto se puede incorporar al modelo ampliando la ecuación (8.7) para incluir términos de interacción. El modelo resultante es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 x_2 + \beta_6 x_1 x_3 + \beta_7 x_1 x_4 + \varepsilon \qquad (8.8)$$

Los cuatro modelos de regresión, que corresponden a las cuatro clases de sistema de acondicionamiento de aire son:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon_1 \qquad \text{(sin acondicionamiento de aire)}$$
$$y = (\beta_0 + \beta_2) + (\beta_1 + \beta_5)x_1 + \varepsilon \qquad \text{(unidades de ventana)}$$
$$y = (\beta_0 + \beta_3) + (\beta_1 + \beta_6)x_1 + \varepsilon \qquad \text{(bomba térmica)}$$
$$y = (\beta_0 + \beta_4) + (\beta_1 + \beta_7)x_1 + \varepsilon \qquad \text{(acondicionamiento central de aire)}$$

Nótese que el modelo (8.8) implica que cada clase de sistema de acondicionamiento de aire puede tener una recta separada de regresión, con su pendiente y ordenada al origen correspondientes.

Ejemplo 8.4 Más de una variable indicadora

Con frecuencia hay varias variables cualitativas diferentes que se deben incorporar al modelo. Para ilustrarlo, supóngase que en el ejemplo 8.1 se debe considerar un segundo factor cualitativo, el tipo de lubricante de corte que se usa, suponiendo que este factor tiene dos niveles, se puede definir una segunda variable indicadora, x_3, como sigue:

$$x_3 = \begin{cases} 0 \text{ si se usa aceite de baja viscosidad} \\ 1 \text{ si se usa aceite de viscosidad intermedia} \end{cases}$$

http://carlos2524.jimdo.com/

Un modelo de regresión que relacione la vida útil de la herramienta (y) con la velocidad de corte (x_1), el tipo de herramienta (x_2) y el tipo de lubricante de corte (x_3) es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \tag{8.9}$$

Es claro que la pendiente β_1 del modelo de regresión que relaciona la vida de la herramienta con la velocidad de corte no depende ni del tipo de herramienta ni del tipo de lubricante de corte. La ordenada al origen de la recta de regresión sí depende de esos factores, en una forma aditiva.

Se pueden agregar diversos tipos de efecto de interacción al modelo. Por ejemplo, supóngase que se consideran interacciones entre la velocidad de corte y los dos factores cualitativos, por lo que el modelo (8.9) se transforma en

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \varepsilon \tag{8.10}$$

Lo anterior conlleva la siguiente situación:

Tipo de buril	Lubricante de corte	Modelo de regresión
A	Baja viscosidad	$y = \beta_0 + \beta_1 x_1 + \varepsilon$
B	Baja viscosidad	$y = (\beta_0 + \beta_2) + (\beta_1 + \beta_4)x_1 + \varepsilon$
A	Viscosidad intermedia	$y = (\beta_0 + \beta_3) + (\beta_1 + \beta_5)x_1 + \varepsilon$
B	Viscosidad intermedia	$y = (\beta_0 + \beta_2 + \beta_3) + (\beta_1 + \beta_4 + \beta_5)x_1 + \varepsilon$

Nótese que cada combinación de tipo de buril y de lubricante de corte da como resultado una línea de regresión aparte, con distintas pendientes y ordenadas al origen, sin embargo, el modelo sigue siendo aditivo con respecto a los niveles de las variables de regresión. Esto es, al cambiar el lubricante de corte de viscosidad baja a intermedia, la ordenada al origen cambia en β_3, y la pendiente en β_5, independientemente de la clase de buril que se use.

Supóngase que se agrega un término de producto cruzado, que implica las dos variables indicadoras x_2 y x_3 al modelo, y se obtiene

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \varepsilon \tag{8.11}$$

En este caso se obtiene lo siguiente:

Tipo de buril	Lubricante de corte	Modelo de regresión
A	Baja viscosidad	$y = \beta_0 + \beta_1 x_1 + \varepsilon$
B	Baja viscosidad	$y = (\beta_0 + \beta_2) + (\beta_1 + \beta_4)x_1 + \varepsilon$
A	Viscosidad intermedia	$y = (\beta_0 + \beta_3) + (\beta_1 + \beta_5)x_1 + \varepsilon$
B	Viscosidad intermedia	$y = (\beta_0 + \beta_2 + \beta_3 + \beta_6) + (\beta_1 + \beta_4 + \beta_5)x_1 + \varepsilon$

La adición del término de producto cruzado $\beta_6 x_2 x_3$ en la ecuación (8.11) da como resultado que el efecto de una variable indicadora sobre la ordenada al origen dependa del nivel de la otra variable indicadora, lo que significa que al cambiar el lubricante de corte de viscosidad baja a intermedia, cambia la ordenada al origen en β_3, si se usa el buril del tipo A, pero ese mismo cambio de lubricante de corte cambia la ordenada al origen en $\beta_3 + \beta_6$

http://carlos2524.jimdo.com/

si se usa el buril del tipo B. Si se agregara un término de interacción $\beta_7 x_1 x_2 x_3$ al modelo (8.11), entonces al cambiar el lubricante de corte de viscosidad baja a viscosidad intermedia tendría un efecto **tanto** en la ordenada al origen **como** en la pendiente, y ese efecto dependería de la clase de buril empleado.

A menos que se disponga de información anterior sobre el efecto esperado del tipo de buril y de la viscosidad del lubricante de corte sobre la duración de la herramienta, habrá que dejar que los datos guíen a uno para seleccionar la forma correcta del modelo. Esto se puede hacer, en general, probando hipótesis sobre los coeficientes individuales de regresión, aplicando la prueba F parcial. Por ejemplo, la prueba de H_0: $\beta_6 = 0$ para el modelo (8.11) permitiría discriminar entre los dos modelos candidato, el (8.11) y el (8.10).

Ejemplo 8.5 *Comparación de modelos de regresión*

Se examinará el caso de la regresión lineal simple, en el que las n observaciones se pueden dividir en M grupos, y el m-ésimo grupo tiene n_m observaciones. El modelo más general consiste en M ecuaciones separadas, como por ejemplo

$$y = \beta_{0m} + \beta_{1mx} + \varepsilon, \qquad m = 1, 2, \ldots, M \tag{8.12}$$

Con frecuencia interesa comparar este modelo general con uno más restrictivo; las variables indicadoras son útiles en este respecto. Se considerarán los siguientes casos:

a. Líneas paralelas En este caso todas las M pendientes son idénticas, $\beta_{11} = \beta_{12} = \cdots = \beta_{1M}$, pero las ordenadas al origen pueden ser distintas, nótese que ésta es la clase de problema que se vio en el ejemplo 8.1 (en donde $M = 2$); condujo al uso de una variable indicadora. En forma más general se puede aplicar el método de la suma extra de cuadrados para probar la hipótesis H_0: $\beta_{11} = \beta_{12} = \cdots = \beta_{1M}$. Recuérdese que este procedimiento implica ajustar un **modelo completo** (FM) y un **modelo reducido** (RM) restringido a la hipótesis nula, y calcular la estadística F:

$$F_0 = \frac{[SS_{\text{Res}}(RM) - SS_{\text{Res}}(FM)]/(df_{RM} - df_{FM})}{SS_{\text{Res}}(FM)/df_{FM}} \tag{8.13}$$

Si el modelo reducido es tan satisfactorio como el modelo completo, entonces F_0 será pequeña en comparación con $F_{\alpha, df_{RM} - df_{FM}, df_{FM}}$. Los valores grandes de F_0 implican que el modelo reducido es inadecuado.

Para ajustar el modelo completo (8.12) tan sólo se ajustan M ecuaciones separadas de regresión, a continuación se calcula $SS_{\text{Res}}(FM)$ sumando las sumas de cuadrados de residuales obtenidas en cada regresión separada. Los grados de libertad para $SS_{\text{Res}}(FM)$ son $df_{FM} = \sum_{m=1}^{M}(n_m - 2) = n - 2M$. Para ajustar el modelo reducido se definen $M - 1$ variables indicadoras, $D_1, D_2, \ldots, D_{M-1}$ que corresponden a los M grupos, y entonces se ajusta

$$y = \beta_0 + \beta_1 x + \beta_2 D_1 + \beta_3 D_2 + \cdots + \beta_M D_{M-1} + \varepsilon$$

La suma de cuadrados de residuales de este modelo es $SS_{\text{Res}}(RM)$ con $df_{RM} = n - (M + 1)$ grados de libertad.

Si la prueba F, ecuación (8.13) indica que los M modelos de regresión tienen una pendiente común, entonces $\hat{\beta}_1$ obtenida en el modelo reducido es un estimado de este parámetro, que se determina agrupando o combinando todos los datos, esto se ilustró en el

http://carlos2524.jimdo.com/

ejemplo 8.1. En forma más general, el **análisis de covarianza** se usa para agrupar los datos para estimar la pendiente común. En consecuencia, el análisis de covarianza es un tipo especial de modelo lineal, que es una combinación de un modelo de regresión (con factores cuantitativos) con un modelo de análisis de varianza (con factores cualitativos). Para conocer una introducción al análisis de covarianza, véase Montgomery [1997].

b. Líneas concurrentes. En esta sección, las M ordenadas al origen son iguales, $\beta_{01} = \beta_{02} = \cdots = \beta_{0M}$, pero las pendientes pueden ser distintas. El modelo reducido es

$$y = \beta_0 + \beta_1 x + \beta_2 Z_1 + \beta_2 Z_2 + \cdots + \beta_M Z_{M-1} + \varepsilon$$

en donde $Z_k = xD_k$, $k = 1, 2, \ldots, M - 1$. La suma de cuadrados de residuales de este modelo es $SS_{\text{Res}}(RM)$ y $df_{RM} = n - (M + 1)$ grados de libertad, nótese que se está suponiendo la concurrencia en el origen. El caso más general, de concurrencia en cualquier punto x_0, se describe en Graybill [1976] y en Seber [1977].

c. Líneas coincidentes En este caso las M pendientes y las M ordenadas al origen son iguales, es decir, $\beta_{01} = \beta_{02} = \cdots = \beta_{0M}$, y $\beta_{11} = \beta_{12} = \cdots \beta_{1M}$. El modelo reducido es tan sólo

$$y = \beta_0 + \beta_1 x + \varepsilon$$

y la suma de cuadrados de residuales $SS_{\text{Res}}(RM)$ tiene $df_{RM} = n - 2$ grados de libertad. No son necesarias variables indicadoras en la prueba de coincidencia, pero se incluye este caso para completar la explicación.

8.2 COMENTARIOS SOBRE EL USO DE VARIABLES INDICADORAS

8.2.1 Variables indicadoras en función de la regresión con códigos asignados

Otro método para manejar una variable cualitativa en regresión es medir los niveles de la variable mediante un código asignado. Recuérdese que en el ejemplo 8.3, donde se está investigando el efecto del tamaño de la vivienda y la clase de sistema de acondicionamiento de aire sobre el consumo residencial, por parte de una empresa eléctrica. En lugar de usar tres variables indicadoras para representar los cuatro niveles del factor cualitativo clase de sistema de acondicionamiento de aire, se podría usar un factor cuantitativo, x_2, con el siguiente código asignado:

Tipo de acondicionamiento de aire	x_2
Sin acondicionamiento de aire	1
Unidades de ventana	2
Bombas térmicas	3
Acondicionamiento central de aire	4

Ahora se puede ajustar el modelo de regresión

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{8.14}$$

http://carlos2524.jimdo.com/

en donde x_1 es el tamaño de la vivienda. Este modelo implica que

$$\begin{aligned} E(y \mid x_1, \text{ sin acondicionamiento de aire}) &= \beta_0 + \beta_1 x_1 + \beta_2 \\ E(y \mid x_1, \text{ unidades de ventana}) &= \beta_0 + \beta_1 x_1 + 2\beta_2 \\ E(y \mid x_1, \text{ bomba térmica}) &= \beta_0 + \beta_1 x_1 + 3\beta_2 \\ E(y \mid x_1, \text{ acondicionamiento de aire central}) &= \beta_0 + \beta_1 x_1 + 4\beta_2 \end{aligned}$$

Una consecuencia directa de lo anterior es que

$$\begin{aligned} &E(y \mid x_1, \text{ acondicionamiento de aire central}) - E(y \mid x_1, \text{ bomba térmica}) \\ &\quad = E(y \mid x_1, \text{ bomba térmica}) - E(y \mid x_1, \text{ unidades de ventana}) \\ &\quad = E(y \mid x_1, \text{ unidades de ventana}) - E(y \mid x_1, \text{ sin acondicionamiento de aire}) \\ &\quad = \beta_2 \end{aligned}$$

lo cual puede ser muy poco realista. Los códigos asignados imponen determinada métrica a los niveles del factor cualitativo. Otras opciones del código asignado implicarían diversas distancias entre los niveles del factor cualitativo, pero nadie garantiza que determinado código asignado lleve a un espaciamiento que sea adecuado.

Las variables indicadoras son más informativas para este problema, porque no forzan a determinada métrica sobre los niveles del factor cualitativo, además, la regresión que usa variables indicadoras siempre produce R^2 mayor que la regresión sobre los códigos asignados (por ejemplo, véase Searle y Udell [1970]).

8.2.2 Variables indicadoras como sustituto de un regresor cuantitativo

También se pueden representar regresores cuantitativos por medio de variables indicadoras. A veces esto se hace necesario, porque es difícil reunir información exacta del regresor cuantitativo. Considere el estudio sobre uso de energía eléctrica en el ejemplo 8.3, y supóngase que en el análisis se incluye un segundo regresor cuantitativo, el ingreso familiar. Como es difícil obtener esta información con precisión, se puede recopilar el regresor cuantitativo ingreso, agrupándolo en clases como las siguientes:

0 a 4 999 dólares
5 000 a 9 999 dólares
10 000 a 14 999 dólares
15 000 a 19 999 dólares
20 000 o más dólares

Ahora se puede representar el factor "ingreso" en el modelo usando cuatro variables regresoras.

Una desventaja de este método es que se requieren más parámetros para representar el contenido de información del factor cuantitativo. En general, si el regresor cuantitativo

http://carlos2524.jimdo.com/

se agrupa en a clases, se requerirán $a - 1$ parámetros, mientras que si se usara el regresor cuantitativo original sólo se requeriría un parámetro. Así, el manejar un factor cuantitativo como se hace con uno cualitativo, aumenta la complejidad del modelo. Este método también reduce los grados de libertad del error, aunque si los datos son numerosos, eso no es problema serio. Una ventaja del método de la variable indicadora es que no requiere que el analista haga hipótesis anteriores acerca de la forma funcional de la relación entre la respuesta y la variable regresora.

8.3 MÉTODO DE REGRESIÓN PARA ANÁLISIS DE VARIANZA

El **análisis de varianza** es una técnica que se usa con frecuencia para analizar los datos de **experimentos planeados** o **diseñados**. Aunque se suelen usar procedimientos especiales de cómputo para el análisis de varianza, cualquier problema de análisis de varianza también se puede manejar como un problema de regresión lineal. De ordinario no se recomienda usar métodos de regresión para el análisis de varianza, porque las técnicas de cómputo especializadas suelen ser muy eficientes, sin embargo, hay ciertos casos de análisis de varianza, en especial los que implican diseños no balanceados, donde puede ayudar el método de regresión, además, muchos analistas no se percatan de la estrecha conexión entre los dos procedimientos. En esencia, cualquier problema de análisis de varianza se puede manejar como un problema de regresión en el que todos los regresores son variables indicadoras.

En esta sección ilustraremos la alternativa de regresión para la clasificación unilateral, o análisis de varianza de un solo factor. Para conocer más ejemplos de la relación entre la regresión y el análisis de varianza, véanse Draper y Smith [1998], Montgomery [2001], Schilling [1974a, b] y Seber [1977].

El modelo para el análisis de varianza de clasificación de una vía es

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij},\ i = 1, 2, \ldots, k, \quad j = 1, 2, \ldots, n \tag{8.15}$$

en donde y_{ij} es la j-ésima observación del i-ésimo **tratamiento** o **nivel de factor**, μ es un parámetro común a los k tratamientos (que se suele llamar **media general**), τ_i es un parámetro que representa el efecto del i-ésimo tratamiento y ε_{ij} son los componentes del error, distribuidos normalmente con media cero y varianza σ^2. Se acostumbra definir los efectos del tratamiento en el caso balanceado (es decir, una cantidad igual de observaciones por tratamiento) como sigue:

$$\tau_1 + \tau_2 + \cdots + \tau_k = 0$$

Además, la media del i-ésimo tratamiento es $\mu_i = \mu + \tau_i$, $i = 1, 2, \ldots, k$. En el caso de efectos fijos (o modelo I), el análisis de varianza se usa para probar la hipótesis que todos las k medias poblacionales son iguales, o lo que es lo mismo,

$$\begin{aligned} &H_0\colon \tau_1 = \tau_2 = \cdots = \tau_k = 0 \\ &H_1\colon \tau_i \neq 0 \ \text{ para al menos una } i \end{aligned} \tag{8.16}$$

La tabla 8.4 muestra el análisis estándar de varianza con un solo factor, en este caso se tiene un término de error verdadero, y no un término residual, porque la replicación permite hacer una estimación del error independiente del modelo. La estadística de prueba

http://carlos2524.jimdo.com/

TABLA 8.4 Análisis unilateral de varianza

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0
Tratamientos	$n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2$	$k-1$	$\frac{SS_{\text{Tratamientos}}}{k-1}$	$\frac{MS_{\text{Tratamientos}}}{MS_{\text{Res}}}$
Error	$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$	$k(n-1)$	$\frac{SS_{\text{Res}}}{k(n-1)}$	
Total	$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$	$kn-1$		

F_0 se compara con $F_{\alpha,\,k-1,\,k(n-1)}$. Si F_0 es mayor que este valor crítico, se rechaza la hipótesis nula H_0 de la ecuación (8.16); esto es, se llega a la conclusión que los k promedios de tratamiento no son iguales. Nótese que en la tabla 8.4 se ha empleado la notación estándar de "subíndice de punto", asociada al análisis de varianza, lo que indica que el promedio de las n observaciones en el i-ésimo tratamiento es

$$\bar{y}_{i.}=\frac{1}{n}\sum_{j=1}^{n}y_{ij},\quad i=1,2,\ldots,k$$

y el promedio general es

$$\bar{y}_{..}=\frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n}y_{ij}$$

Para ilustrar la relación entre el análisis de varianza con efectos fijos y un solo factor, y la regresión, supóngase que se tienen $k=3$ tratamientos, de modo que la ecuación (8.15) se transforma en

$$y_{ij}=\mu+\tau_i+\varepsilon_{ij},\qquad i=1,2,3;\qquad j=1,3,\ldots,n \tag{8.17}$$

Esos tres tratamientos se pueden considerar como tres niveles de un **factor cualitativo** y se pueden manejar con variables indicadoras. En forma específica, un factor cualitativo con tres niveles necesitaría dos variables indicadoras, definidas como sigue:

$$x_1=\begin{cases}1 & \text{si la observación procede del tratamiento 1}\\ 0 & \text{en cualquier caso}\end{cases}$$

$$x_2=\begin{cases}1 & \text{si la observación procede del tratamiento 2}\\ 0 & \text{en cualquier caso}\end{cases}$$

Por consiguiente, el modelo de regresión viene a ser

$$y_{ij}=\beta_0+\beta_1x_{1j}+\beta_2x_{2j}+\varepsilon_{ij},\quad i=1,2,3;\quad j=1,2,\ldots,n \tag{8.18}$$

en donde x_{1j} es el valor de la variable indicadora x_1 para la observación j en el tratamiento i, y x_{2j} es el valor de x_2 para la observación j en el tratamiento i.

http://carlos2524.jimdo.com/

La relación entre el parámetro β_u ($u = 0, 1, 2$) en el modelo de regresión, y los parámetros μ y τ_i ($i = 1, 2, \ldots, k$) en el modelo de análisis de varianza se determina con facilidad.

Considérense las observaciones del tratamiento 1, para las cuales

$$x_{1j} = 1 \text{ y } x_{2j} = 0$$

El modelo de regresión (8.18) se transforma en

$$\begin{aligned} y_{1j} &= \beta_0 + \beta_1(1) + \beta_2(0) + \varepsilon_{1j} \\ &= \beta_0 + \beta_1 + \varepsilon_{1j} \end{aligned}$$

Como en el modelo de análisis de varianza una observación del tratamiento 1 se representa con $y_{1j} = \mu + \tau_1 + \varepsilon_{1j} = \mu_1 + \varepsilon_{1j}$, esto implica que

$$\beta_0 + \beta_1 = \mu_1$$

De igual modo, si las observaciones proceden del tratamiento 2, entonces $x_{1j} = 0$, $x_{2j} = 1$ y

$$\begin{aligned} y_{2j} &= \beta_0 + \beta_1(0) + \beta_2(1) + \varepsilon_{2j} \\ &= \beta_0 + \beta_2 + \varepsilon_{2j} \end{aligned}$$

En cuanto al modelo de análisis de varianza, $y_{2j} = \mu + \tau_2 + \varepsilon_{2j} = \mu_2 + \varepsilon_{2j}$, así que

$$\beta_0 + \beta_2 = \mu_2$$

Por último, en cuanto a las observaciones del tratamiento 3, ya que $x_{1j} = x_{2j} = 0$, el modelo de regresión se transforma en

$$\begin{aligned} y_{3j} &= \beta_0 + \beta_1(0) + \beta_2(0) + \varepsilon_{3j} \\ &= \beta_0 + \varepsilon_{3j} \end{aligned}$$

El modelo de análisis de varianza correspondiente es $y_{3j} = \mu + \tau_3 + \varepsilon_{3j} = \mu_3 + \varepsilon_{3j}$, así que

$$\beta_0 = \mu_3$$

Por lo anterior, en la formulación del modelo de regresión del análisis unifactorial de varianza, los coeficientes de regresión describen comparaciones de los dos primeros tratamientos μ_1 y μ_2 con la media μ_3 del tercer tratamiento. Esto es,

$$\begin{aligned} \beta_0 &= \mu_3 \\ \beta_1 &= \mu_1 - \mu_2 \\ \beta_2 &= \mu_2 - \mu_3 \end{aligned}$$

http://carlos2524.jimdo.com/

En general, si hay k tratamientos, el modelo de regresión para el análisis de varianza con un solo factor necesitará $k - 1$ variables indicadoras, por ejemplo,

$$y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_{k-1} x_{k-1,j} + \varepsilon_{ij}$$
$$i = 1, 2, \ldots, k, \; j = 1, 2, \ldots, n \tag{8.19}$$

en donde

$$x_{ij} = \begin{cases} 1 & \text{si la observación } j \text{ procede del tratamiento } i \\ 0 & \text{en cualquier otro caso} \end{cases}$$

La relación entre los parámetros del modelo de regresión y los modelos de análisis de varianza es

$$\beta_0 = \mu_k$$
$$\beta_i = \mu_i - \mu_k, \quad i = 1, 2, \ldots, k - 1$$

Así, β_0 siempre estima la media del k-ésimo tratamiento, y β_i estima las diferencias en las medias del tratamiento i y del tratamiento k.

Ahora se verá el ajuste del modelo de regresión para el análisis de varianza de una vía. De nuevo, supóngase que se tienen $k = 3$ tratamientos, pero ahora sean $n = 3$ observaciones por tratamiento. La matriz $\mathbf{X}$ y el vector $\mathbf{y}$ son los siguientes:

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \end{bmatrix} \qquad \mathbf{X} = \begin{array}{c} \begin{array}{ccc} & x_1 & x_2 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

Nótese que la matriz $\mathbf{X}$ consiste sólo de 0 y 1. Ésta es una característica de la formulación de cualquier modelo de análisis de varianza por regresión. Las ecuaciones normales de mínimos cuadrados son

$$(\mathbf{X}'\mathbf{X})\,\hat{\beta} = \mathbf{X}'\mathbf{y}$$

o sea

$$\begin{bmatrix} 9 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \end{bmatrix}$$

http://carlos2524.jimdo.com/

en donde y_i es el total de todas las observaciones en el tratamiento i, y $y_{..}$ es el gran total de las nueve observaciones (es decir, $y_{..} = y_{1.} + y_{2.} + y_{3.}$). La solución de las ecuaciones normales es

$$\hat{\beta}_0 = \bar{y}_{..} - \bar{y}_{1.} - \bar{y}_{2.} = \bar{y}_{3.}$$

$$\hat{\beta}_1 = \bar{y}_{1.} - \bar{y}_{3.}$$

$$\hat{\beta}_2 = \bar{y}_{2.} - \bar{y}_{3.}$$

Se puede usar el método de la suma extra de cuadrados para probar si hay diferencias en las medias de los tratamientos. Para el modelo completo, la suma de cuadrados de regresión es

$$\begin{aligned} SS_R\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right) &= \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{y} = \left[\bar{y}_{3.}, \bar{y}_{1.} - \bar{y}_{3.}, \bar{y}_{2.} - \bar{y}_{3.}\right] \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \end{bmatrix} \\ &= y_{..}\bar{y}_{3.} + y_{1.}(\bar{y}_{1.} - \bar{y}_{3.}) + y_{2.}(\bar{y}_{2.} - \bar{y}_{3.}) \\ &= (y_{1.} + y_{2.} + y_{3.})\bar{y}_{3.} + y_{1.}(\bar{y}_{1.} - \bar{y}_{3.}) + y_{2.}(\bar{y}_{2.} - \bar{y}_{3.}) \\ &= \bar{y}_{1.}y_{1.} + \bar{y}_{2.}y_{2.} + \bar{y}_{3.}y_{3.} \\ &= \sum_{i=1}^{3} \frac{y_{i.}^2}{3} \end{aligned}$$

con tres grados de libertad. La suma de cuadrados residuales de error, para el modelo completo, es

$$\begin{aligned} SS_{\text{Res}} &= \sum_{i=1}^{3}\sum_{j=1}^{3} y_{ij}^2 - SS_R(\beta_0, \beta_1, \beta_2) \\ &= \sum_{i=1}^{3}\sum_{j=1}^{3} y_{ij}^2 - \sum_{i=1}^{3} \frac{y_{i.}^2}{3} \\ &= \sum_{i=1}^{3}\sum_{j=1}^{3} (y_{ij} - \bar{y}_{i.})^2 \end{aligned} \tag{8.20}$$

con $9 - 3 = 6$ grados de libertad. Nótese que la ecuación (8.20) es la suma de cuadrados de error en la tabla de análisis de varianza (Tabla 8.4) para $k = n = 3$.

La prueba de diferencias en las medias de tratamientos equivale a probar

$$H_0\text{: } \tau_1 = \tau_2 = \tau_3 = 0$$

$$H_1\text{: al menos una } \tau_i \neq 0$$

Si H_0 es cierta, los parámetros en el modelo de regresión vienen a ser

$$\beta_0 = \mu, \quad \beta_1 = 0, \quad \beta_2 = 0$$

http://carlos2524.jimdo.com/

Por consiguiente, el modelo reducido sólo contiene un parámetro, esto es

$$y_{ij} = \beta_0 + \varepsilon_{ij}$$

El estimado de β_0 en el modelo reducido es $\hat{\beta}_0 = \bar{y}_{..}$, y la suma de cuadrados de regresión con un grado de libertad para este modelo es

$$SS_R(\beta_0) = \frac{y_{..}^2}{9}$$

La suma de cuadrados para probar la igualdad de medias de los tratamientos es la diferencia en las sumas de cuadrados de regresión, entre los modelos completo y reducido, o sea

$$\begin{aligned} SS_R(\beta_1, \beta_2|\beta_0) &= SS_R(\beta_0, \beta_1, \beta_2) - SS_R(\beta_0) \\ &= \sum_{i=1}^{3} \frac{y_{i.}^2}{3} - \frac{y_{..}^2}{9} \\ &= 3\sum_{j=1}^{3} (\bar{y}_{i.} - \bar{y}_{..})^2 \end{aligned} \tag{8.21}$$

Esta suma tiene $3 - 1 = 2$ grados de libertad. Nótese que la ecuación (8.21) es la suma de cuadrados de tratamiento, en la tabla 8.4, suponiendo que $k = n = 3$. La estadística de prueba adecuada es

$$\begin{aligned} F_0 &= \frac{SS_R(\beta_1, \beta_2|\beta_0)/2}{SS_{Res}/6} \\ &= \frac{3\sum_{i=1}^{3} (\bar{y}_{i.} - \bar{y}_{..})^2/2}{\sum_{i=1}^{3}\sum_{j=1}^{3} (y_{ij} - \bar{y}_{i.})^2/6} \\ &= \frac{MS_{Tratamientos}}{MS_{Res}} \end{aligned}$$

Si es cierta H_0: $\tau_1 = \tau_2 = \tau_3 = 0$, entonces F_0 sigue la distribución $F_{2,6}$. Es la misma estadística de prueba que aparece en la tabla de análisis de varianza (Tabla 8.4). Por consiguiente, el método de regresión es idéntico al procedimiento de análisis de una vía de varianza que se describió en la tabla 8.4.

PROBLEMAS

8.1 Para el modelo de regresión (8.8) descrito en el ejemplo 8.3, graficar la función de respuesta e indicar el papel que juegan los parámetros del modelo en la determinación de la forma de esta función.

http://carlos2524.jimdo.com/

8.2 Considérense los modelos de regresión descritos en el ejemplo 8.4.
- **a.** Graficar la función de respuesta asociada con la ecuación (8.10).
- **b.** Graficar la función de respuesta asociada con la ecuación (8.11).

8.3 Acerca de los datos del tiempo de entrega del ejemplo 3.1, en la sección 4.2.5 se hizo notar que esas observaciones se tomaron en cuatro ciudades: San Diego, Boston, Austin y Minneapolis.
- **a.** Desarrollar un modelo que relacione el tiempo de entrega y con las cajas x_1, la distancia x_2 y la ciudad donde se hizo la entrega. Estimar los parámetros del modelo.
- **b.** ¿Hay algún indicio de que el lugar de la entrega es una variable importante?
- **c.** Analizar los residuales de este modelo. ¿A qué conclusiones se puede llegar, acerca de la adecuación del modelo?

8.4 Para los datos de rendimiento de gasolina en coches, en la tabla B.3 del apéndice.
- **a.** Formar un modelo de regresión lineal que relacione el rendimiento de gasolina y con la cilindrada del motor x_1 y con el tipo de transmisión x_{11}. ¿Afecta en forma importante el tipo de transmisión al rendimiento de la gasolina?
- **b.** Modificar el modelo desarrollado en la parte a, para incluir una interacción entre la cilindrada del motor y el tipo de transmisión. ¿Qué conclusiones se pueden sacar acerca del efecto del tipo de transmisión sobre el rendimiento de la gasolina? Interpretar los parámetros en este modelo.

8.5 Para los datos de rendimiento de gasolina en coches, de la tabla B.3 del apéndice.
- **a.** Formar un modelo de regresión lineal que relacione el rendimiento de la gasolina y con el peso del vehículo x_{10}, y con el tipo de transmisión x_{11}. El tipo de transmisión ¿afecta el rendimiento de la gasolina en forma importante?
- **b.** Modificar el modelo desarrollado en la parte a, para incluir una interacción entre el peso del vehículo y el tipo de transmisión. ¿A qué conclusiones se puede llegar acerca del efecto del tipo de transmisión sobre el rendimiento de la gasolina? Interpretar los parámetros de este modelo.

8.6 Considérense los datos de la Liga Nacional de Futbol de la tabla B.1 del apéndice. Formular un modelo de regresión lineal que relacione la cantidad de juegos ganados con las yardas por tierra de los contrarios, x_8, el porcentaje de jugadas por tierra x_7 y una modificación del diferencial de pérdidas de balón, x_5. En forma específica, sea el diferencial de pérdidas de balón una variable indicadora cuyo valor se determina si ese diferencial real es positivo, negativo o cero. ¿Qué conclusiones se pueden sacar acerca del efecto de las pérdidas de balón sobre la cantidad de juegos ganados?

8.7 **Regresión lineal por segmentos.** En el ejemplo 7.3 se mostró cómo se puede ajustar un modelo de regresión lineal con un cambio de pendiente en algún punto t ($x_{\text{mín}} < t < x_{\text{máx}}$) usando funciones spline. Desarrollar una formulación del modelo de regresión lineal, por supuestos usando variables indicadoras. Suponer que la función es continua en el punto t.

8.8 **Continuación del problema 8.7**. Indicar cómo se pueden usar variables indicadoras para formar un modelo de regresión lineal por segmentos con una discontinuidad en el punto de unión t.

8.9 Suponer que un análisis unilateral de varianza implica cuatro tratamientos, pero que se ha tomado una cantidad distinta de observaciones (por ejemplo, n_i), para cada tratamiento. Suponer que $n_1 = 3$, $n_2 = 2$, $n_3 = 4$ y $n_4 = 3$, escribir el vector $\mathbf{y}$ y la matriz $\mathbf{X}$ para analizar esos datos en forma de un modelo de regresión múltiple. ¿Hay complicaciones introducidas por la naturaleza desbalanceada de esos datos?

http://carlos2524.jimdo.com/

8.10 **Esquemas alternativos de codificación para el método de regresión de análisis de varianza**. La ecuación (8.18) representa al modelo de regresión correspondiente a un análisis de varianza con tres tratamientos y n observaciones por tratamiento. Suponer que las variables indicadoras x_1 y x_2 se definen como sigue:

$$x_1 = \begin{cases} 1 & \text{si la observación procede del tratamiento 1} \\ -1 & \text{si la observación procede del tratamiento 2} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{si la observación procede del tratamiento 2} \\ -1 & \text{si la observación procede del tratamiento 3} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

a. Demostrar que la relación entre los parámetros en los modelos de regresión y de análisis de varianza es

$$\beta_0 = \frac{\mu_1 + \mu_2 + \mu_3}{3} = \overline{\mu}$$

$$\beta_1 = \mu_1 - \overline{\mu}$$

$$\beta_2 = \mu_2 - \overline{\mu}$$

b. Escribir el vector **y** y la matriz **X**.

c. Desarrollar una suma adecuada de cuadrados para probar la hipótesis H_0: $\tau_1 = \tau_2 = \tau_3 = 0$. ¿Es ésta la suma de cuadrados que se maneja en forma acostumbrada en el análisis de varianza de una vía?

8.11 Montgomery [2001] presenta un experimento acerca de la resistencia a la tensión de las fibras sintéticas con que se fabrican telas para camisas de hombre. Se cree que la resistencia está influida por el porcentaje de algodón en la fibra. Los datos se ven a continuación.

Porcentaje de algodón	Resistencia a la tensión				
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

a. Escribir el factor **y** y la matriz **X** para el modelo de regresión correspondiente.

b. Determinar los estimados de los parámetros del modelo, por mínimos cuadrados.

c. Determinar un estimado de punto para la diferencia en resistencia media con 15% y 25% de algodón.

d. Probar la hipótesis que la resistencia media a la tensión es la misma para los cinco porcentajes de algodón.

8.12 **Análisis de varianza de dos vías**. Suponer que son de interés dos conjuntos distintos de tratamientos. Sea y_{ijk} la k-ésima observación en el nivel i del primer tipo de tratamiento y el nivel j del segundo tipo de tratamiento. El modelo de análisis de varianza de dos vías es

$$y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk}$$

$$i = 1, 2, \ldots, a, \quad j = 1, 2, \ldots, b, \quad k = 1, 2, \ldots, n,$$

http://carlos2524.jimdo.com/

en donde τ_i es el efecto del nivel i sobre el primer tipo de tratamiento, γ_j es el efecto del nivel j sobre el segundo tipo de tratamiento, $(\tau\gamma)_{ij}$ es un efecto de interacción entre los dos tipos de tratamiento y ε_{ijk} es un componente aleatorio de error NID(0, σ^2).

a. Para el caso $a = b = n = 2$, escribir un modelo de regresión que corresponda al análisis de varianza de dos vías.

b. ¿Cuáles son el vector **y** y la matriz **X** para este modelo de regresión?

c. Describir cómo se podría usar el modelo de regresión para probar la hipótesis H_0: $\tau_1 = \tau_2 = 0$ (las medias del tratamiento tipo 1 son iguales), H_0: $\gamma_1 = \gamma_2 = 0$ (las medias del tratamiento tipo 2 son iguales) y H_0: $(\tau\gamma)_{11} = (\tau\gamma)_{12} = (\tau\gamma)_{22} = 0$ (no hay interacción entre los tipos de tratamiento).

8.13 La tabla B.11 del apéndice presenta datos de la calidad del vino Pinot Noir.

a. Formular un modelo de regresión que relacione la calidad y con el aroma x_4, que incorpore la región de información que se ve en la última columna. ¿Tiene algún impacto esa región sobre la calidad del vino?

b. Hacer un análisis de residuales para este modelo, y comentar sobre la adecuación modelo.

c. ¿Hay algunos valores atípicos u observaciones influyentes en este conjunto de dat

d. Modificar el modelo de la parte a para incluir términos de interacción entre el aro y las variables de región. ¿Es mejor este modelo que el que se formó en la parte a?

8.14 Usar los datos de calidad de vino en la tabla B.11 del apéndice, para ajustar un modelo que relacione la calidad del vino y con el aroma x_4, usando la región como código de asignación, que toma los valores que se ven en la tabla: 1, 2 y 3. Describir la interpretación de los parámetros en este modelo. Comparar el modelo con uno que haya formado el lector usando variables indicadoras, en el problema 8.13.

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

CAPÍTULO 9

SELECCIÓN DE VARIABLE Y CONSTRUCCIÓN DEL MODELO

9.1 INTRODUCCIÓN

9.1.1 El problema de la construcción del modelo

En los capítulos anteriores se ha supuesto que se conoce que las variables regresoras incluidas en el modelo son importantes. Nuestro enfoque fue hacia las técnicas que aseguren que la forma funcional del modelo es la correcta y que no se violen las suposiciones básicas. En algunas aplicaciones, las consideraciones teóricas o la experiencia pueden ayudar a seleccionar los regresores que se van a usar en el modelo. Sin embargo, en la mayoría de los problemas prácticos el analista tiene un grupo de **regresores candidatos**, que deberían incluir a todos los factores influyentes, y debe determinar el subconjunto real de regresores que debe usarse en el modelo. La definición de un subconjunto adecuado de regresores para el modelo es lo que se llama **problema de selección de variable.**

La construcción de un modelo de regresión que sólo incluya un subconjunto de los regresores disponibles implica dos objetivos contrapuestos: 1) Se desea que el modelo incluya tantos regresores como sea posible, para que el contenido de información en ellos pueda influir sobre el valor predicho de y. 2) Se desea que el modelo incluya los menos regresores que sea posible, porque la varianza de la predicción $\hat{y}$ aumenta a medida que aumenta la cantidad de regresores. También, mientras más regresores haya en un modelo, los costos de recolección de datos y los de mantenimiento de modelo serán mayores. El proceso de encontrar un modelo que sea un término medio entre los dos objetivos se llama selección de la **"mejor" ecuación de regresión**. Desafortunadamente, como se verá en este capítulo, no hay definición única de "mejor", además hay varios algoritmos que se pueden usar para seleccionar variables, y esos procedimientos especifican como mejores, con frecuencia, subconjuntos distintos de los regresores candidatos.

El problema de selección de variables se suele describir en un entorno idealizado. Por lo general se supone conocida la especificación funcional correcta de los regresores (por ejemplo, $1/x_1$, $\ln x_2$), y que no hay valores atípicos ni observaciones influyentes. En la práctica rara vez se cumplen esas premisas. El **análisis de residuales**, como el que se describió en el capítulo 4, es útil para revelar formas funcionales de regresores, que se debieran investigar, o para señalar nuevos regresores candidatos, así como para identificar defectos en los datos, como son los valores atípicos. El efecto de las **observaciones influyentes** o de **gran balanceo** también se debe determinar. La investigación de la adecuación del modelo está ligada al problema de selección de variable; aunque en el caso ideal esos problemas se deben resolver al mismo tiempo, con frecuencia se usa un método iterativo, en el que 1) se emplea determinada estrategia de selección de variable, y 2) el modelo resultante para el

http://carlos2524.jimdo.com/

subconjunto se verifica para comprobar su especificación funcional correcta, valores atípicos y observaciones influyentes. Esto podrá indicar que se debe repetir el paso 1. Se podrán requerir varias iteraciones para llegar al modelo adecuado.

Ninguno de los procedimientos de selección de variable que se describen en este capítulo tiene garantía de producir la ecuación de regresión óptima para un determinado conjunto de datos, de hecho, en general no hay una sola ecuación óptima, sino más bien varias igualmente buenas. Como los algoritmos de selección de variable dependen mucho del cómputo, a veces el analista tiende a dar mucho énfasis a los resultados de determinado procedimiento; se debe evitar caer en esa tentación. La experiencia, el juicio profesional en el área y las consideraciones subjetivas tienen importancia en el problema de selección de variables. El analista debe usar los procedimientos de selección de variables como métodos para explorar la estructura de los datos. Una buena descripción general de selección de variables en regresión se encuentra en Cox y Snell [1974], Hocking [1972, 1976], Hocking y LaMotte [1973], Myers [1990] y Thompson [1978a, b].

9.1.2 Consecuencias de la mala especificación del modelo

Para motivar la selección de variables se repasará en forma breve las consecuencias de la **especificación incorrecta del modelo**. Supóngase que hay K regresores candidatos, x_1, $x_2, \ldots, x_K$, y que hay $n \geq K + 1$ observaciones de esos regresores y de la respuesta y. El **modelo completo** que contiene a todos los K regresores es

$$y_i = \beta_0 + \sum_{j=1}^{K} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{9.1a}$$

o, lo que es igual

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{9.1b}$$

Se supondrá que la lista de los regresores candidatos contiene a todas las variables importantes. Nótese que la ecuación (9.1) contiene un término de ordenada al origen, β_0. Si bien β_0 también podría ser un candidato para la selección, suele entrar forzado al modelo. Se supondrá que todas las ecuaciones incluyen un término de ordenada al origen, sea r la cantidad de regresores que se eliminan de la ecuación (9.1); entonces, la cantidad de variables que se retienen es $p = K + 1 - r$. Como se incluye la ordenada al origen, el modelo subconjunto contiene $p - 1 = K - r$ de los regresores originales.

El modelo (9.1) se puede escribir en la forma

$$\mathbf{y} = \mathbf{X}_p \beta_p + \mathbf{X}_r \beta_r + \varepsilon \tag{9.2}$$

en donde la matriz $\mathbf{X}$ se ha dividido en $\mathbf{X}_p$, una matriz de $n \times p$ cuyas columnas representan la ordenada al origen y los $p - 1$ regresores que quedan en el modelo de subconjunto, y $\mathbf{X}_r$, una matriz de $n \times r$ cuyas columnas representan los regresores que se van a omitir del modelo completo, sea β dividida respectiva en β_p y β_r. Para el modelo completo, el estimado de β por mínimos cuadrados es

$$\hat{\beta}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{9.3}$$

http://carlos2524.jimdo.com/

y un estimado de la varianza residual, σ^2, es

$$\hat{\sigma}_*^2 = \frac{\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}^{*\prime}\mathbf{X}'\mathbf{y}}{n - K - 1} = \frac{\mathbf{y}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y}}{n - K - 1} \tag{9.4}$$

Los componentes de $\hat{\beta}^*$ se representan por β_p^* y β_r^*, y $\hat{y}_i^*$ representa los valores ajustados. Para el modelo de subconjunto

$$\mathbf{y} = \mathbf{X}_p\boldsymbol{\beta}_p + \varepsilon \tag{9.5}$$

el estimado de β_p por mínimos cuadrados es

$$\hat{\boldsymbol{\beta}}_p = (\mathbf{X}'_p\mathbf{X}_p)^{-1}\mathbf{X}'_p\mathbf{y} \tag{9.6}$$

el estimado de la varianza residual es

$$\hat{\sigma}^2 = \frac{\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'_p\mathbf{X}'_p\mathbf{y}}{n - p} = \frac{\mathbf{y}'\left[\mathbf{I} - \mathbf{X}_p(\mathbf{X}'_p\mathbf{X}_p)^{-1}\mathbf{X}'_p\right]\mathbf{y}}{n - p} \tag{9.7}$$

y los valores ajustados son $\hat{y}_i$.

Varios autores han investigado las propiedades de los estimados $\hat{\beta}_p$ y $\hat{\sigma}^2$ del modelo de subconjunto, incluyendo Hocking [1974, 1976], Narula y Ramberg [1972], Rao [1971], Rosenberg y Levy [1972] y Walls y Weeks [1969]. Los resultados se pueden resumir como sigue:

1. El valor esperado de $\hat{\beta}_p$ es

$$\begin{aligned} E\left(\hat{\boldsymbol{\beta}}_p\right) &= \boldsymbol{\beta}_p + (\mathbf{X}'_p\mathbf{X}_p)^{-1}\mathbf{X}'_p\mathbf{X}_r\boldsymbol{\beta}_r \\ &= \boldsymbol{\beta}_p + \mathbf{A}\mathbf{B}_r \end{aligned}$$

   en donde $\mathbf{A} = (\mathbf{X}'_p\mathbf{X}_p)^{-1}\mathbf{X}'_p\mathbf{X}_r$ (a la matriz A se le llama a veces la matriz alias), entonces, $\hat{\beta}_p$ es un estimado sesgado de β_p, a menos que los coeficientes de regresión que correspondan a las variables eliminadas (β_r) sean cero, o que las variables estimadas sean ortogonales a las variables omitidas ($\mathbf{X}'_p\mathbf{X}_r = \mathbf{0}$).
2. Las varianzas de $\hat{\beta}_p$ y $\hat{\beta}^*$ son $\mathrm{Var}(\hat{\beta}_p) = \sigma^2(\mathbf{X}'_p\mathbf{X}_p)^{-1}$ y $\mathrm{Var}(\hat{\beta}^*) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, respectivamente. También, la matriz $\mathrm{Var}(\hat{\beta}_p^*) - \mathrm{Var}(\hat{\beta}_p)$ es positiva semidefinida, esto es, las varianzas de los estimados, por mínimos cuadrados, de los parámetros del modelo completo son mayores que, o iguales a las varianzas de los parámetros correspondientes en el modelo de subconjunto; en consecuencia, al eliminar las variables nunca se aumentan las varianzas de los estimados de los parámetros restantes.
3. Como $\hat{\beta}_p$ es un estimado sesgado de β_p y $\hat{\beta}_p^*$ no lo es, es más razonable comparar la precisión de los estimados de los parámetros para los modelos completo y de subconjunto en términos del error cuadrático medio; recuérdese que si $\hat{\theta}$ es un estimado del parámetro θ, el error cuadrático medio de $\hat{\theta}$ es

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2$$

http://carlos2524.jimdo.com/

El error cuadrático medio de $\hat{\beta}_p$ es

$$\text{MSE}(\hat{\beta}_p) = \sigma^2(\mathbf{X}'_p\mathbf{X}_p)^{-1} + \mathbf{A}\beta_r\beta_r\mathbf{A}'$$

Si la matriz $\text{Var}(\hat{\beta}^*_p) - \beta_r\beta_r'$ es positiva semidefinida, la matriz $\text{Var}(\hat{\beta}^*_p) - \text{MSE}(\hat{\beta}_p)$ es positiva semidefinida. Eso quiere decir que los estimados por mínimos cuadrados de los parámetros en el modelo de subconjunto tienen menor error cuadrático medio que los correspondientes del modelo completo, cuando las variables eliminadas tienen coeficientes de regresión que son menores que los errores estándar de sus estimados en el modelo completo.

4. El parámetro $\hat{\sigma}^2_*$ del modelo completo es un estimado insesgado de σ^2, sin embargo, para el modelo de subconjunto,

$$E(\hat{\sigma}^2) = \frac{\sigma^2 + \beta_r'\mathbf{X}_r'\left[\mathbf{I} - \mathbf{X}_p(\mathbf{X}_p'\mathbf{X}_p)^{-1}\mathbf{X}_p'\right]\mathbf{X}_r\beta_r}{n-p}$$

Esto es, $\hat{\sigma}^2$ en general es sesgado hacia arriba como estimado de σ^2.

5. Supóngase que se desea predecir la respuesta en el punto $\mathbf{x}' = [\mathbf{x}_p', \mathbf{x}_r']$. Si se usa el modelo completo, el valor predicho es $\hat{y}^* = \mathbf{x}'\hat{\beta}^*$, con media $\mathbf{x}'\beta$ y varianza de la predicción

$$\text{Var}(\hat{y}^*) = \sigma^2[1 + \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}]$$

Sin embargo, si se usa el modelo de subconjunto, $\hat{y} = \mathbf{x}_p'\hat{\beta}_p$, con media

$$E(\hat{y}) = \mathbf{x}_p'\beta_p + \mathbf{x}_p'\mathbf{A}\beta_r$$

y el error cuadrático medio de predicción

$$\text{MSE}(\hat{y}) = \sigma^2[1 + \mathbf{x}_p'(\mathbf{X}'_p\mathbf{X}_p)^{-1}\mathbf{x}_p] + (\mathbf{x}_p'\mathbf{A}\beta_r - \mathbf{x}_r'\beta_r)^2$$

Obsérvese que $\hat{y}$ es un estimado sesgado de y a menos que $\mathbf{x}_p'\mathbf{A}\beta_r = 0$, que sólo es cierto en general si $\mathbf{X}'_p\mathbf{X}_r\beta_r = \mathbf{0}$, además, la varianza de $\hat{y}^*$ procedente del modelo completo no es menor que la de $\hat{y}$ del modelo de subconjunto. En términos de error cuadrático medio, se puede demostrar que

$$\text{Var}(\hat{y}^*) \geq \text{MSE}(\hat{y})$$

siempre y cuando la matriz $\text{Var}(\hat{\beta}^*_p) - \beta_r\beta_r'$ sea positiva semidefinida.

Esta motivación para seleccionar variables se puede resumir así: al eliminar variables del modelo, se puede **mejorar la precisión** de los estimados de los parámetros de las variables retenidas, aun cuando algunas de las variables omitidas no sean despreciables; esto también es cierto para la varianza de una respuesta predicha. El omitir variables introduce **sesgo** potencial en los estimados de los coeficientes de las variables retenidas y la respuesta, sin embargo, si las variables eliminadas tienen efectos pequeños, el MSE de los estimados sesgados será menor que la varianza de los estimados insesgados, ya que, la cantidad de sesgo introducida es menor que la reducción en la varianza. Existe peligro al retener variables despreciables, lo que nos lleva a variables con coeficientes cero, o que sean

http://carlos2524.jimdo.com/

menores que sus errores estándar correspondientes del modelo completo. El peligro es que aumenten las varianzas de los estimados de los parámetros y de la respuesta predicha.

Por último, recuérdese que, según la sección 1.2, los modelos de regresión se forman con frecuencia usando datos retrospectivos, esto es, datos extraídos de registros históricos, esos datos suelen estar saturados por defectos que incluyen valores atípicos, puntos "extraños" e inconsistencias debidas a cambios en el sistema de recolección de datos y del proceso en la información a través del tiempo. Esos defectos en los datos pueden tener gran impacto sobre el proceso de selección de variables, y conducir a la mala especificación del modelo. Un problema muy común en los datos históricos es encontrar que algunos regresores candidatos se han controlado de tal modo que varíen dentro de un intervalo muy limitado, asimismo, dichos datos son las variables más influyentes, y fueron controladas en forma estricta para mantener la respuesta dentro de límites aceptables, sin embargo, debido al intervalo limitado de los datos, el regresor puede parecer poco importante en el ajuste por mínimos cuadrados. Ésta es una especificación grave e incorrecta del modelo, que sólo puede evitar el conocimiento no estadístico, sobre el ambiente del problema, de quien hace el modelo. Cuando se controla mucho el intervalo de las variables que se creen importantes, puede ser que el analista deba recolectar nuevos datos, en forma específica para formar el modelo, a este respecto son útiles los experimentos diseñados.

9.1.3 Criterios para evaluar modelos de regresión con subconjuntos de variables

Dos aspectos clave del problema de selección de variables son la generación de los modelos con subconjuntos y la decisión de si un subconjunto es mejor que otro. En esta sección describiremos criterios para evaluar y comparar los modelos de regresión con subconjuntos. En la sección 9.2 se presentarán métodos de cómputo para seleccionar variables.

Coeficiente de determinación múltiple

Una medida de la adecuación de un modelo de regresión, que se ha usado mucho, es el coeficiente de determinación múltiple, R^2. Sea R_p^2 el coeficiente de determinación múltiple para un modelo de regresión con subconjunto de p términos, esto es, $p - 1$ regresores y un término β_0 de ordenada al origen. Este coeficiente se calcula como sigue:

$$R_p^2 = \frac{SS_{\mathrm{R}}(p)}{SS_{\mathrm{T}}} = 1 - \frac{SS_{\mathrm{Res}}(p)}{SS_{\mathrm{T}}} \tag{9.8}$$

en donde $SS_{\mathrm{R}}(p)$ y $SS_{\mathrm{Res}}(p)$ representan la suma de cuadrados de la regresión y la suma de cuadrados de los residuales, respectivamente, para un modelo de subconjunto de p términos. Nótese que hay $\binom{K}{p-1}$ valores de R_p^2 para cada valor de p, uno para cada posible modelo de subconjunto de tamaño p, ahora bien, R_p^2 aumenta al aumentar p, y es máximo cuando $p = K + 1$; por consiguiente, el analista usa este criterio agregando regresores al modelo hasta el punto en el que una variable adicional ya no es útil porque sólo produce un pequeño aumento de R_p^2. El método general se ve en la figura 9.1, que presenta una gráfica hipotética del valor máximo de R_p^2 para cada subconjunto de tamaño p, en función de p. En el caso característico, se examina una gráfica como ésta y a continuación se especifica la cantidad de regresores para el modelo final, como el punto en el que se hace aparente el "codo" en la curva, es claro que para ello se necesita juicio por parte del analista.

http://carlos2524.jimdo.com/

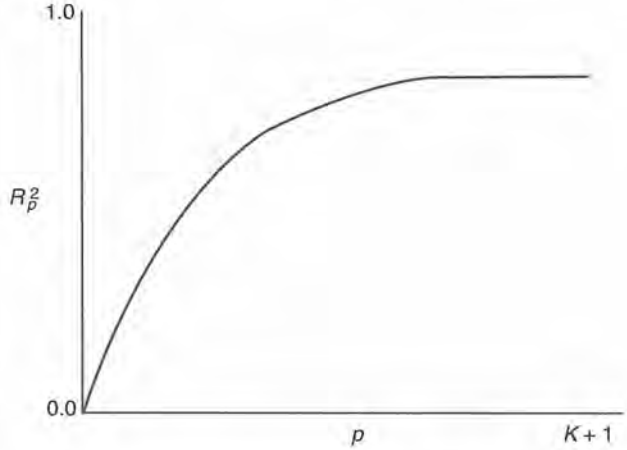


Figura 9.1
Gráfica de R_p^2 en función de p.

Como no se puede determinar un valor "óptimo" de R^2 para un modelo de regresión con subconjunto se debe buscar un valor "satisfactorio". Aitkin [1974] ha propuesto una solución para este problema proponiendo una prueba con la que se pueden identificar todos los modelos de regresión que tienen una R^2 no muy distinta de la R^2 para el modelo completo. Sea

$$R_0^2 = 1 - (1 - R_{K+1}^2)(1 + d_{a,\,n,\,K}) \tag{9.9}$$

en donde

$$d_{a,\,n,\,K} = \frac{KF_{a,\,n,\,n-K-1}}{n - K - 1}$$

y R_{K+1}^2 es el valor de R^2 para el modelo completo. Aitkin llama **subconjunto adecuado** para R^2 (α) a cualquier subconjunto de variables regresoras que produzcan una R^2 mayor que R_0^2.

En general, no es directo el uso de R^2 como criterio para escoger la cantidad de regresores que se van a incluir en el modelo, sin embargo, para una cantidad fija p de variables, se puede usar R_p^2 para comparar los $\binom{K}{p-1}$ modelos de subconjunto generados así; se prefieren los modelos que tengan valores grandes de R_p^2.

R^2 ajustada

Para evitar las dificultades de interpretar a R^2, algunos analistas prefieren usar la estadística R^2 ajustada, definida para una ecuación de p términos como sigue:

$$R_{\text{Aj},\,p}^2 = 1 - \left(\frac{n-1}{n-p}\right)\left(1 - R_p^2\right) \tag{9.10}$$

La estadística $R_{\text{Aj},\,p}^2$ no necesariamente aumenta cuando se introducen más regresores al modelo, de hecho se puede demostrar (Edwards [1969], Haitovski [1969] y Seber [1977]) que si se agregan s regresores al modelo, $R_{\text{Aj},\,p+s}^2$ será mayor que $R_{\text{Aj},\,p}^2$ si, y sólo si la estadística F parcial para probar el significado de los s regresores adicionales es mayor que 1, en consecuencia, un criterio para seleccionar un modelo con subconjunto óptimo es elegir

http://carlos2524.jimdo.com/

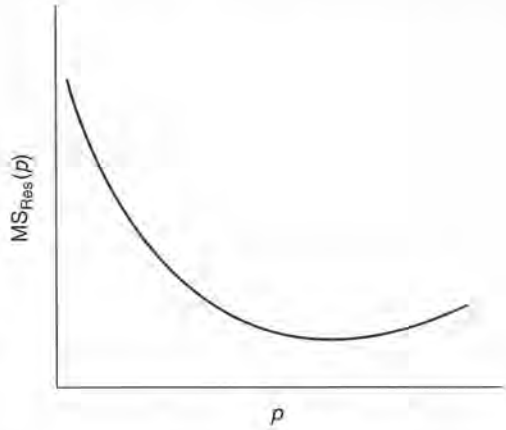


Figura 9.2 Gráfica de $MS_{Res}(p)$ en función de p.

el que tenga una $R^2_{Aj,\,p}$ máxima; sin embargo, esto es equivalente a otro criterio que se presentará a continuación.

Cuadrado medio de residuales

El cuadrado medio de residuales para un modelo de regresión de subconjunto, por ejemplo

$$MS_{Res}(p) = \frac{SS_{Res}(p)}{n-p} \tag{9.11}$$

se puede usar como criterio para evaluación de un modelo. El comportamiento general de $MS_{Res}(p)$ al aumentar p se ve en la figura 9.2. Como $SS_{Res}(p)$ siempre disminuye al aumentar p, al principio $MS_{Res}(p)$ decrece, después se estabiliza y al final puede crecer, el aumento eventual de $MS_{Res}(p)$ se presenta cuando la reducción de $SS_{Res}(p)$ debida a la adición de un regresor al modelo no es suficiente para compensar la pérdida de un grado de libertad en el denominador de la ecuación (9.11), ya que, si se agrega un regresor a un modelo de p términos, $MS_{Res}(p+1)$ se hará mayor que $MS_{Res}(p)$ si la disminución en la suma de cuadrados de residuales es menor que $MS_{Res}(p)$. Los partidarios del criterio $MS_{Res}(p)$ grafican $MS_{Res}(p)$ en función de p, y basan su elección en

1. El $MS_{Res}(p)$ mínimo.
2. El valor de p tal que $MS_{Res}(p)$ sea aproximadamente igual a MS_{Res} para el modelo completo, o bien
3. Un valor de p cercano al punto en donde crece el $MS_{Res}(p)$ mínimo.

El modelo de regresión para subconjunto, que minimiza $MS_{Res}(p)$ también maximizará $R^2_{Aj,\,p}$. Para visualizar esto, nótese que

$$\begin{aligned} R^2_{Aj,p} &= 1 - \frac{n-1}{n-p}\left(1 - R^2_p\right) \\ &= 1 - \frac{n-1}{n-p}\,\frac{SS_{Res}(p)}{SS_T} \end{aligned}$$

http://carlos2524.jimdo.com/

$$= 1 - \frac{n-1}{SS_T}\frac{SS_{Res}(p)}{n-p}$$

$$= 1 - \frac{MS_{Res}(p)}{SS_T/(n-1)}$$

Así, los criterios de $MS_{Res}(p)$ mínimo y de R^2 ajustada máxima son equivalentes.

Estadística C_p de Mallows

Mallows [1964, 1966, 1973, 1995] ha propuesto un criterio que se relaciona con el error cuadrático medio de un valor ajustado, esto es,

$$E[\hat{y}_i - E(y_i)]^2 = [E(y_i) - E(\hat{y}_i)]^2 + \mathrm{Var}(\hat{y}_i) \tag{9.12}$$

Nótese que $E(y_i)$ es la respuesta esperada de la ecuación verdadera de regresión, y que $E(\hat{y}_i)$ es la respuesta esperada con el modelo de subconjunto de p términos. Entonces, $E(y_i) - E(\hat{y}_i)$ es el sesgo en el i-ésimo punto de dato, en consecuencia, los dos términos del lado derecho de la ecuación (9.12) son los componentes **sesgo al cuadrado** y **varianza**, respectivamente, del error cuadrático medio, sea el sesgo total al cuadrado, para una ecuación de p términos

$$SS_B(p) = \sum_{i=1}^{n}\left[E(y_i) - E(\hat{y}_i)\right]^2$$

y se define el error cuadrático medio total normalizado como sigue:

$$\begin{aligned} T_p &= \frac{1}{\sigma^2}\left\{\sum_{i=1}^{n}\left[E(y_i) - E(\hat{y}_i)\right]^2 + \sum_{i=1}^{n}\mathrm{Var}(\hat{y}_i)\right\} \\ &= \frac{SS_B(p)}{\sigma^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathrm{Var}(\hat{y}_i) \end{aligned} \tag{9.13}$$

Se puede demostrar que

$$\sum_{i=1}^{n}\mathrm{Var}(\hat{y}_i) = p\sigma^2$$

y que el valor esperado de la suma de cuadrados de residuales de una ecuación de p términos es

$$E[SS_{Res}(p)] = SS_B(p) + (n-p)\sigma^2$$

Se sustituyen $\Sigma_{i=1}^{n}\mathrm{Var}(\hat{y}_i)$ y $SS_B(p)$ en la ecuación (9.13), para obtener

$$\Gamma_p = \frac{1}{\sigma^2}\left\{E\left[SS_{Res}(p)\right] - (n-p)\sigma^2 + p\sigma^2\right\} = \frac{E\left[SS_{Res}(p)\right]}{\sigma^2} - n + 2p \tag{9.14}$$

http://carlos2524.jimdo.com/

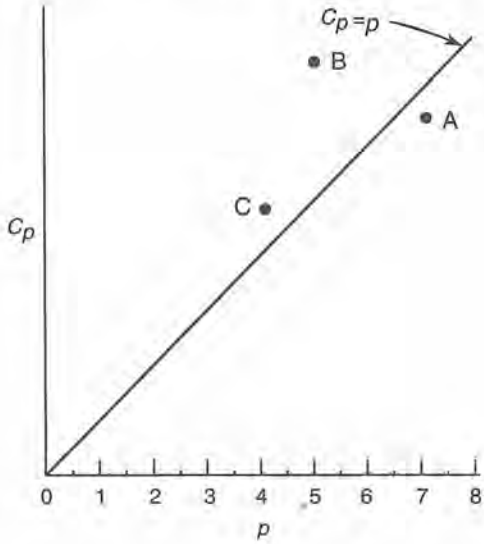


Figura 9.3
Una gráfica de C_p.

Supóngase que $\hat{\sigma}^2$ es un buen estimado de σ^2, entonces al sustituir $E[SS_{\text{Res}}(p)]$ por el valor observado $SS_{\text{Res}}(p)$, se produce un estimado de Γ_p, por ejemplo

$$C_p = \frac{SS_{\text{Res}}(p)}{\hat{\sigma}^2} - n + 2p \tag{9.15}$$

Si el modelo con p términos tiene sesgo despreciable, entonces $SS_B(p) = 0$, en consecuencia, $E[SS_{\text{Res}}(p)] = (n-p)\sigma^2$ y

$$E\left[C_p \,|\, \text{Sesgo} = 0\right] = \frac{(n-p)\sigma^2}{\sigma^2} - n + 2p = p$$

Cuando se usa el criterio C_p puede ser útil el visualizar la gráfica de C_p en función de p para cada ecuación de regresión, como se ve en la figura 9.3. Las ecuaciones de regresión con poco sesgo tendrán valores de C_p que caen cerca de la recta $C_p = p$ (punto A en la Fig. 9.3), mientras que las que las que tengan sesgo apreciable quedarán arriba de esa recta (punto B de la Fig. 9.3). En general, **se prefieren valores pequeños de C_p**, por ejemplo, aunque el punto C de la figura 9.3 está arriba de la recta $C_p = p$, está más abajo del punto A y por consiguiente representa a un modelo con menor error total. Puede ser mejor aceptar cierto sesgo en la ecuación, para reducir el error promedio de predicción.

Para calcular C_p se necesita un estimado insesgado de σ^2, con frecuencia se usa el cuadrado medio de residuales para la ecuación completa, sin embargo, esto hace que $C_p = p = K + 1$ para la ecuación completa. Al usar $MS_{\text{Res}}(K+1)$ del modelo completo como estimado de σ^2 se supone que el modelo completo tiene sesgo despreciable. Si el modelo completo tiene varios regresores que no contribuyen de forma significativa al modelo (coeficientes de regresión cero), entonces $MS_{\text{Res}}(K+1)$ sobreestimará σ^2 con frecuencia, y por ello los valores de C_p serán pequeños. Si la estadística C_p ha de funcionar en forma correcta, se debe usar un buen estimado de σ^2, como alternativa a $MS_{\text{Res}}(K+1)$ se podría basar la estimación de σ^2 en pares de puntos que sean "vecinos cercanos" en el espacio de x, como se explicó en la sección 4.5.2.

http://carlos2524.jimdo.com/

Usos de la regresión y criterios de evaluación del modelo

Como se ha visto hay varios criterios que se pueden aplicar para evaluar los modelos de regresión de subconjuntos. El criterio que se usará para seleccionar el modelo se debería relacionar con el uso pretendido del modelo. Hay varios usos posibles de la regresión, que incluyen 1) la descripción de los datos, 2) predicción y estimación; 3) estimación de parámetros, y 4) control.

Si el objetivo es obtener una buena descripción de determinado proceso, o modelar un sistema complejo, lo indicado es buscar ecuaciones de regresión con pequeñas sumas de cuadrados de residuales. Como SS_{Res} se minimiza usando los K regresores candidatos, se suele preferir eliminar algunas variables sólo si se obtiene un pequeño aumento en SS_{Res}; en general, lo deseable es describir el sistema con el mínimo posible de regresores, pero a la vez explicar la parte importante de la variabilidad de y.

Con frecuencia se usan ecuaciones de regresión para predecir observaciones en el futuro, o estimación de la respuesta promedio, en general, se desea seleccionar los regresores de tal modo que el error cuadrático medio de la predicción se reduzca al mínimo, esto suele implicar que se deben eliminar del modelo los regresores con efectos pequeños, también se podría usar la estadística PRESS que se explicó en el capítulo 4, para evaluar ecuaciones candidato, obtenidas con un procedimiento de generación de subconjuntos, recuérdese que para un modelo de regresión de p términos,

$$\begin{aligned} \text{PRESS}_p &= \sum_{i=1}^{n} \left[y_i - \hat{y}_{(i)} \right]^2 \\ &= \sum_{i=1}^{n} \left(\frac{e_i}{1 - h_{ii}} \right)^2 \end{aligned} \tag{9.16}$$

A continuación se selecciona el modelo de regresión de subconjuntos basado en un valor pequeño de PRESS_p. Si bien PRESS_p tiene un atractivo intuitivo, en especial para el problema de predicción, no es una función sencilla de la suma de cuadrados de residuales, y no es obvio e inmediato el desarrollo de un algoritmo para seleccionar variables basado en este criterio, sin embargo, esta estadística es potencialmente útil para discriminar entre modelos alternativos, como se ilustrará.

Si el interés es de estimar parámetros, entonces es claro que se deben considerar tanto el sesgo que resulta al eliminar variables, como las varianzas de los coeficientes estimados. Cuando los regresores son muy multicolineales, los estimados de los coeficientes individuales de regresión, por mínimos cuadrados, serán muy malos. En el capítulo 10 se describirán varios métodos que pueden mejorar los estimados de los parámetros cuando la matriz $\mathbf{X}$ es deteriorada. Esos métodos conducen, con frecuencia, a una ecuación de predicción que es más efectiva en la extrapolación que la basada en los mínimos cuadrados, porque en la extrapolación son esenciales los buenos estimados de los parámetros.

Cuando un modelo de regresión se usa para control, lo importante son estimados exactos de los parámetros. Eso implica que los errores estándar de los coeficientes de regresión deben ser pequeños, además, como los ajustes hechos en las x para controlar y serán proporcionales a las $\hat{\beta}$ los coeficientes de regresión deben representar con fidelidad los efectos de los regresores. Si los regresores son muy multicolineales, las $\hat{\beta}$ pueden ser estimadas muy pobres de los efectos de los regresores individuales.

http://carlos2524.jimdo.com/

9.2 TÉCNICAS COMPUTACIONALES PARA SELECCIONAR VARIABLES

Se ha visto que es preferible considerar modelos de regresión que usen un subconjunto de las variables regresoras candidatos. Para determinar el conjunto de variables que se van a usar en la ecuación final, es natural la consideración de ajuste de modelos con varias combinaciones de los regresores candidatos. En esta sección se describirán varias técnicas computacionales para generar modelos de regresión con subconjuntos, e ilustrar criterios para evaluar esos modelos.

9.2.1 Todas las regresiones posibles

Este procedimiento requiere que el analista ajuste todas las ecuaciones de regresión, que tengan un regresor candidato, dos regresores candidato, etc. Esas ecuaciones se evalúan de acuerdo con algún criterio adecuado y se selecciona el "mejor" modelo de regresión. Si se supone que el término de ordenada al origen β_0 se incluye en todas las ecuaciones, entonces, si hay K regresores candidato, hay 2^K ecuaciones en total por estimar y examinar, por ejemplo si $K = 4$, hay $2^4 = 16$ ecuaciones posibles, mientras que si $K = 10$, hay $2^{10} = 1\ 024$ ecuaciones posibles de regresión. Se ve que la cantidad de ecuaciones por examinar aumenta con rapidez a medida que aumenta la cantidad de regresores candidato. Antes del desarrollo de programas eficientes de cómputo, era impráctico generar todas las regresiones posibles, para problemas donde intervenían más que unos pocos regresores. La disponibilidad de computadoras de alta velocidad ha motivado el desarrollo de varios algoritmos muy eficientes para obtener todas las regresiones posibles. A continuación se describirán algunos de esos algoritmos.

Ejemplo 9.1 *Datos de Hald sobre cemento*

Hald [1952][1] presentó datos acerca del calor producido (y), en calorías por gramos de cemento, en función de la cantidad de cada uno de los cuatro ingredientes en la mezcla: aluminato tricálcico (x_1), silicato tricálcico (x_2), aluminoferrito tetracálcico (x_3) y silicato dicálcico (x_4), esos datos aparecen en la tabla 9.1; y se usarán para ilustrar el método de todas las regresiones posibles para seleccionar variables.

Como hay $K = 4$ regresores candidatos, hay $2^4 = 16$ ecuaciones posibles de regresión, si se incluye siempre la ordenada al origen β_0. En la tabla 9.2 se muestran los resultados de ajustar esas 16 ecuaciones, también en esa tabla aparecen las estadísticas R_p^2, $R_{\text{Aj},p}^2$, $MS_{\text{Res}}(p)$ y C_p.

La tabla 9.3 muestra los estimados de los coeficientes de regresión por mínimos cuadrados. La naturaleza parcial de los coeficientes de regresión se ve con facilidad al examinar la tabla, por ejemplo, se considerará x_2. Cuando el modelo sólo contiene a x_2, el estimado por mínimos cuadrados para el efecto de x_2 es 0.789; si se agrega al modelo x_4, el efecto de x_2 es 0.311, una reducción de más del 50%; la inclusión posterior de x_3 cambia el efecto de x_2 a –0.923. Es claro que el estimado de un coeficiente de regresión por mínimos cuadrados depende mucho de los **demás** regresores en el modelo. Los grandes

[1] Estos son datos "clásicos" para ilustrar los problemas inherentes a la selección de variables. Véanse otros análisis en Daniel y Wood [1980], Draper y Smith [1998] y Seber [1977].

http://carlos2524.jimdo.com/

TABLA 9.1 Datos de cemento obtenidos por Hald para el ejemplo 9.1

Observación i	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}
1	78.5	7	26	6	60
2	74.3	1	29	15	52
3	104.3	11	56	8	20
4	87.6	11	31	8	47
5	95.9	7	52	6	33
6	109.2	11	55	9	22
7	102.7	3	71	17	6
8	72.5	1	31	22	44
9	93.1	2	54	18	22
10	115.9	21	47	4	26
11	83.8	1	40	23	34
12	113.3	11	66	9	12
13	109.4	10	68	8	12

TABLA 9.2 Resumen de todas las regresiones posibles para el ejemplo 9.1

Cantidad de regresores en el modelo	p	Regresores en el modelo	$SS_{\text{Res}}(p)$	R_p^2	$R_{\text{Aj},p}^2$	$MS_{\text{Res}}(p)$	C_p
Ninguno	1	Ninguno	2715.7635	0	0	226.3136	442.92
1	2	x_1	1265.6867	0.53395	0.49158	115.0624	202.55
1	2	x_2	906.3363	0.66627	0.63593	82.3942	142.49
1	2	x_3	1939.4005	0.28587	0.22095	176.3092	315.16
1	2	x_4	883.8669	0.67459	0.64495	80.3515	138.73
2	3	$x_1 x_2$	57.9045	0.97868	0.97441	5.7904	2.68
2	3	$x_1 x_3$	1227.0721	0.54817	0.45780	122.7073	198.10
2	3	$x_1 x_4$	74.7621	0.97247	0.96697	7.4762	5.50
2	3	$x_2 x_3$	415.4427	0.84703	0.81644	41.5443	62.44
2	3	$x_2 x_4$	868.8801	0.68006	0.61607	86.8880	138.23
2	3	$x_3 x_4$	175.7380	0.93529	0.92235	17.5738	22.37
3	4	$x_1 x_2 x_3$	48.1106	0.98228	0.97638	5.3456	3.04
3	4	$x_1 x_2 x_4$	47.9727	0.98234	0.97645	5.3303	3.02
3	4	$x_1 x_3 x_4$	50.8361	0.98128	0.97504	5.6485	3.50
3	4	$x_2 x_3 x_4$	73.8145	0.97282	0.96376	8.2017	7.34
4	5	$x_1 x_2 x_3 x_4$	47.8636	0.98238	0.97356	5.9829	5.00

cambios en los coeficientes de regresión que se observan con los datos de cemento de Hald, cuando se agregan o quitan variables, indican que hay una apreciable correlación entre los cuatro regresores (es el problema de multicolinealidad que se explicó en el Cap. 3).

Se describirá la evaluación de los modelos con subconjunto con el criterio de R_p^2. Una gráfica de R_p^2 en función de p se ve en la figura 9.4. Al examinar esta figura se ve que después de que hay dos regresores en el modelo, hay poca ganancia en términos de R^2 cuando se introducen variables adicionales. Los dos modelos de dos regresores (x_1, x_2) y (x_1, x_4) tienen en esencia los mismos valores de R^2, y en lo que respecta a este criterio, sería poca la

http://carlos2524.jimdo.com/

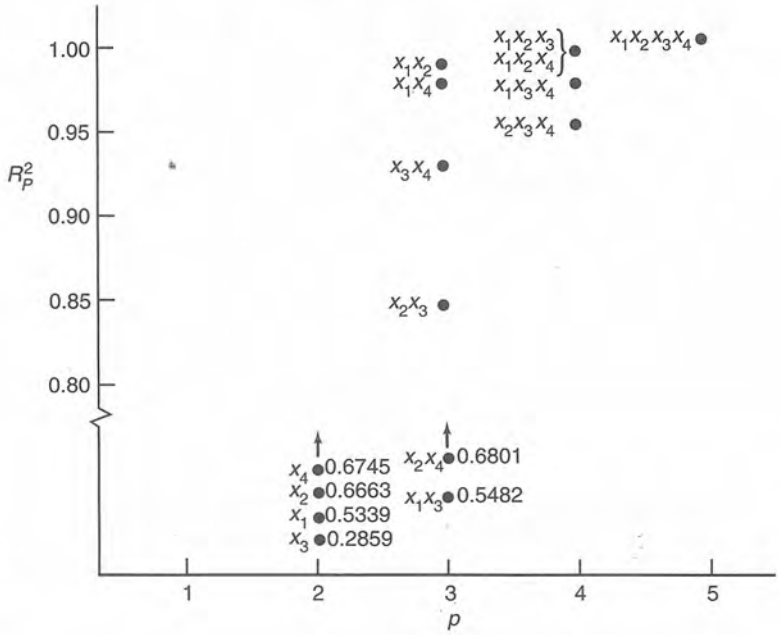


Figura 9.4 Gráfica de R_p^2 en Función de p, ejemplo 9.1.

TABLA 9.3 Estimados de todas las regresiones posibles por mínimos cuadrados (datos de cemento de Hald)

Variables en el modelo	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
x_1	81.479	1.869			
x_2	57.424		0.789		
x_3	110.203			– 1.256	
x_4	117.568				– 0.738
$x_1 x_2$	52.577	1.468	0.662		
$x_1 x_3$	72.349	2.312		0.494	
$x_1 x_4$	103.097	1.440			– 0.614
$x_2 x_3$	72.075		0.731	– 1.008	
$x_2 x_4$	94.160		0.311		– 0.457
$x_3 x_4$	131.282			– 1.200	– 0.724
$x_1 x_2 x_3$	48.194	1.696	0.657	0.250	
$x_1 x_2 x_4$	71.648	1.452	0.416		– 0.237
$x_2 x_3 x_4$	203.642		– 0.923	– 1.448	– 1.557
$x_1 x_3 x_4$	111.684	1.052		– 0.410	– 0.643
$x_1 x_2 x_3 x_4$	62.405	1.551	0.510	0.102	– 0.144

TABLA 9.4 Matriz de correlaciones simples para los datos de Hald, en el ejemplo 9.1

	x_1	x_2	x_3	x_4	y
x_1	1.0				
x_2	0.229	1.0			
x_3	– 0.824	– 0.139	1.0		
x_4	– 0.245	– 0.973	0.030	1.0	
y	0.731	0.816	– 0.535	– 0.821	1.0

http://carlos2524.jimdo.com/

diferencia si se escogiera cualquiera de los modelos como la ecuación final de regresión; podría ser mejor usar (x_1, x_4) porque x_4 produce el mejor modelo con un regresor. De acuerdo con la ecuación (9.9) se ve que si $\alpha = 0.05$,

$$\begin{aligned} R_0^2 &= 1 - \left(1 - R_5^2\right)\left(1 + \frac{4F_{0.5,4,8}}{8}\right) \\ &= 1 - 0.01762\left[1 + \frac{4(3.84)}{8}\right] \\ &= 0.94855 \end{aligned}$$

Por consiguiente, todo el subconjunto de regresión para el que $R_p^2 > R_0^2 = 0.94855$ es adecuado R^2 (0.05), es decir, su R^2 no es en forma significativa, diferente de R_{k+1}^2, es claro que varios modelos de la tabla 9.2 satisfacen este criterio, por lo que la elección del modelo final no está clara.

Es ilustrativo examinar los pares de correlaciones entre x_i y x_j y entre x_i y y. Esas correlaciones sencillas se ven en la tabla 9.4, nótese que los pares de regresores (x_1, x_3) y (x_2, x_4) están muy correlacionados, porque

$$r_{13} = -0.824 \qquad \text{y} \qquad r_{24} = -0.973$$

En consecuencia, si se agregan más regresores cuando x_1 y x_2 o cuando x_1 y x_4 ya están en el modelo será de poco provecho, porque el contenido de información en los regresores excluidos está presente, en esencia, en los regresores que ya están en el modelo. Esta estructura de correlación es responsable, parcialmente, de los grandes cambios de los coeficientes de regresión que se nota en la tabla 9.3.

En la figura 9.5 se muestra una gráfica de $MS_{\text{Res}}(p)$ en función de p; el modelo con cuadrado medio residual mínimo es el (x_1, x_2, x_4), para el que $MS_{\text{Res}}(4) = 5.3303$, nótese que, como era de esperarse, el modelo que minimiza a $MS_{\text{Res}}(p)$ también maximiza la R^2 ajustada, sin embargo, dos de los otros modelos con tres regresores, (x_1, x_2, x_3) y (x_1, x_3, x_4), y los dos modelos de dos regresores (x_1, x_2) y (x_1, x_4), tienen valores comparables del cuadrado medio residual; si en el modelo está ya sea (x_1, x_2) o (x_1, x_4), hay poca reducción en el cuadrado medio de residuales cuando se agregan más regresores. El modelo con subconjunto (x_1, x_2) puede ser más adecuado que el (x_1, x_4), porque tiene menor valor del cuadrado medio de residuales.

En la figura 9.6 se ve una gráfica de C_p; para ilustrar los cálculos, supóngase que se toma $\hat{\sigma} = 5.9829$ (MS_{Res} para el modelo completo) y se calcula C_3 para el modelo (x_1, x_4). De acuerdo con la ecuación (9.15),

$$\begin{aligned} C_3 &= \frac{SS_{\text{Res}}(3)}{\hat{\sigma}^2} - n + 2p \\ &= \frac{74.7621}{5.9829} - 13 + 2(3) \\ &= 5.50 \end{aligned}$$

Al examinar esta gráfica se ve que hay cuatro modelos que podrían ser aceptables: (x_1, x_2), (x_1, x_2, x_3), (x_1, x_2, x_4) y (x_1, x_3, x_4). Sin considerar factores adicionales, como la información técnica acerca de los regresores, o los costos de la recolección de datos, podría ser adecuado elegir el modelo más sencillo (x_1, x_2), como modelo final, porque tiene el C_p mínimo.

http://carlos2524.jimdo.com/

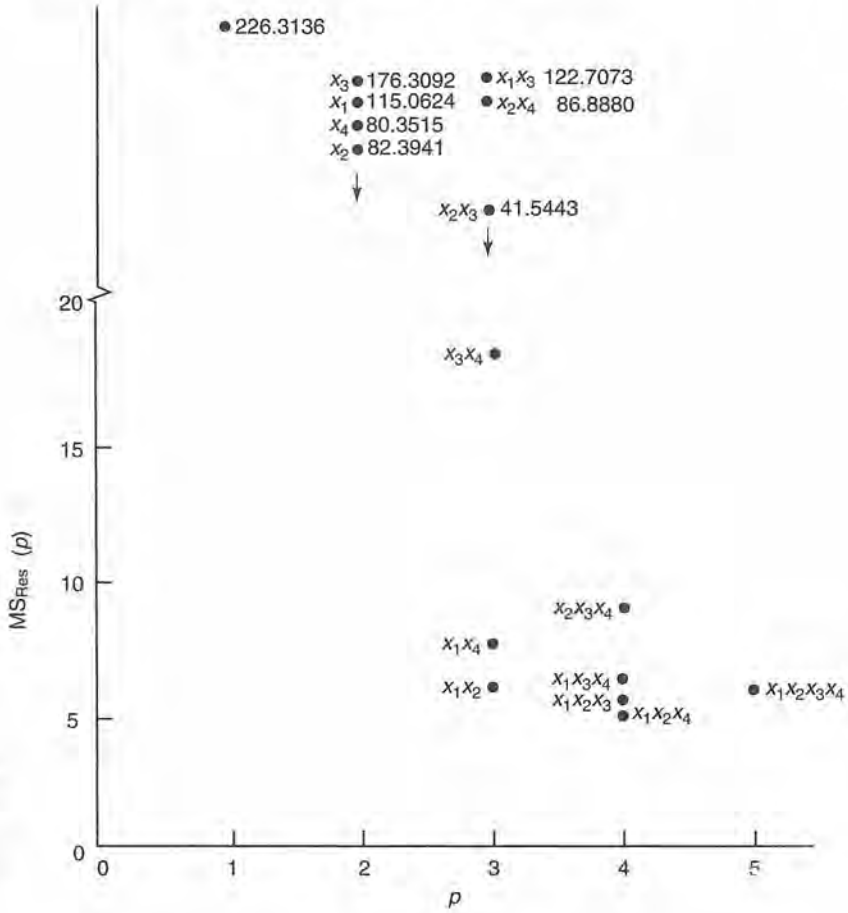


Figura 9.5 Gráfica de $MS_{Res}(p)$ en función de p, ejemplo 9.1.

TABLA 9.5 Comparaciones de dos modelos para los datos de cemento de Hald

Observación	$\hat{y} = 52.58 + 1.468x_1 + 0.662x_2$ [a]			$\hat{y} = 71.65 + 1.452x_1 + 0.416x_2 - 0.237x_4$ [b]		
i	e_i	h_{ii}	$[e_i/(1-h_{ii})]^2$	e_i	h_{ii}	$[(e_i/(1-h_{ii})]^2$
1	−1.5740	0.25119	4.4184	0.0617	0.52058	0.0166
2	−1.0491	0.26189	2.0202	1.4327	0.27670	3.9235
3	−1.5147	0.11890	2.9553	−1.8910	0.13315	4.7588
4	−1.6585	0.24225	4.7905	−1.8016	0.24431	5.6837
5	−1.3925	0.08362	2.3091	0.2562	0.35733	0.1589
6	4.0475	0.11512	20.9221	3.8982	0.11737	19.5061
7	−1.3031	0.36180	4.1627	−1.4287	0.36341	5.0369
8	−2.0754	0.24119	7.4806	−3.0919	0.34522	22.2977
9	1.8245	0.17195	4.9404	1.2818	0.20881	2.6247
10	1.3625	0.55002	9.1683	0.3539	0.65244	1.0368
11	3.2643	0.18402	16.0037	2.0977	0.32105	9.5458
12	0.8628	0.19666	1.1535	1.0556	0.20040	1.7428
13	−2.8934	0.21420	13.5579	−2.2247	0.25923	9.0194
		PRESS x_1, x_2 =	93.8827		PRESS x_1, x_2, x_4 =	85.3516

[a] $R^2_{\text{predicción}} = 0.9654$, $\text{FIV}_1 = 1.05$, $\text{FIV}_2 = 1.06$

[b] $R^2_{\text{predicción}} = 0.9684$, $\text{FIV}_1 = 1.07$, $\text{FIV}_2 = 18.78$, $\text{FIV}_4 = 18.94$.

http://carlos2524.jimdo.com/

Este ejemplo ha ilustrado el proceso de cómputo relacionado con la formación de modelo con todas las regresiones posibles; nótese que no hay elección definida de la mejor ecuación de regresión. Con mucha frecuencia se ve que los distintos criterios sugieren distintas ecuaciones, por ejemplo, la ecuación con C_p mínimo es la que tiene (x_1, x_2), y la de MS_{Res} mínimo es la que tiene (x_1, x_2, x_4). Todos los modelos candidatos "finales" deben someterse a las pruebas usuales de adecuación, incluyendo la investigación de los puntos de balanceo, de influencia y la multicolinealidad, por ejemplo, la tabla 9.5 examina los dos modelos (x_1, x_2) y (x_1, x_2, x_4) con respecto a PRESS y a sus factores de inflación de varianza (FIV). Ambos modelos tienen valores de PRESS muy parecidos (más o menos el doble de la suma de cuadrados de residuales para la ecuación con MS_{Res} mínima), y la R^2 para predicción, calculada a partir de PRESS es parecida en ambos modelos, sin embargo, x_2 y x_4 muy multicolineales, lo que se ve por los mayores factores de inflación de varianza en x_2, x_4). Ya que ambos modelos tienen estadísticas PRESS equivalentes, se recomenda- modelo con (x_1, x_2), con base en la falta de colinealidad en él.

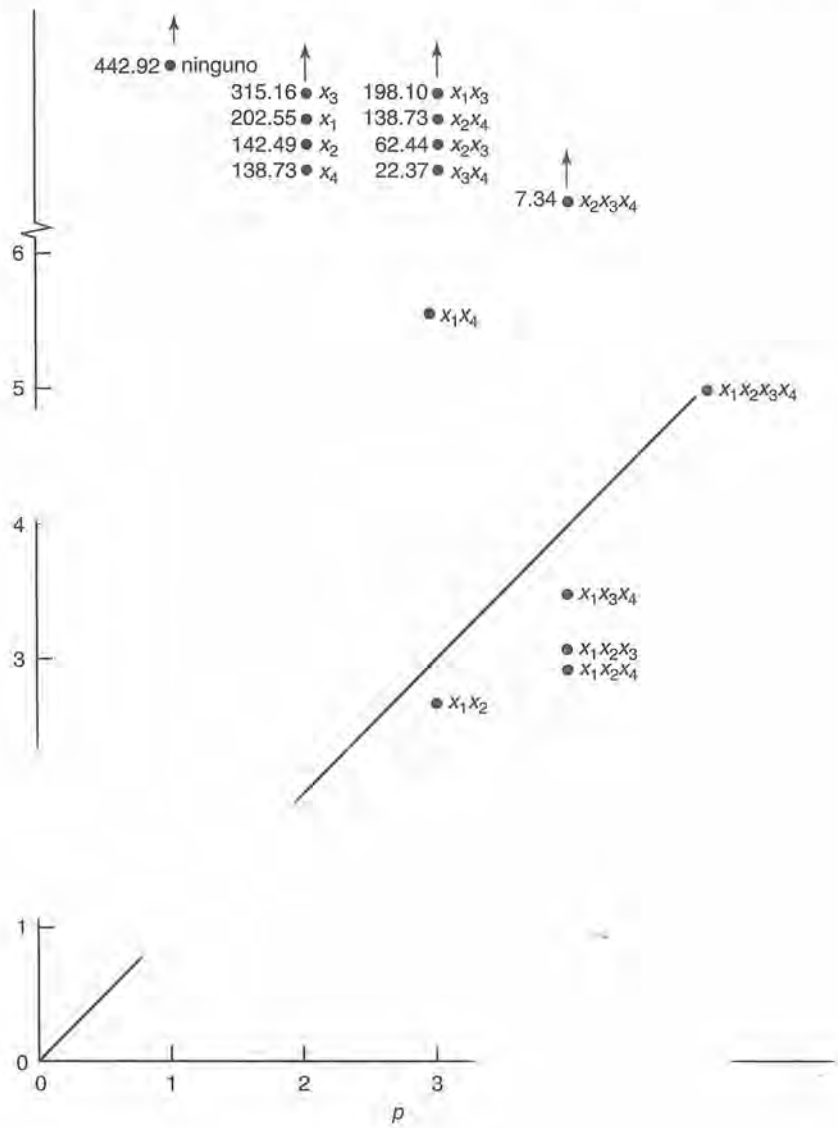


Figura 9.6 La gráfica de C_p para el ejemplo 9.1.

http://carlos2524.jimdo.com/

Mejor regresión de subconjunto: y en función de x1, x2, x3, x4

Response is y

Vars	R-Sq	R-Sq(adj)	C-p	S	x 1	x 2	x 3	x 4
1	67.5	64.5	138.7	8.9639				X
1	66.6	63.6	142.5	8.0771		X		
1	53.4	49.2	202.5	10.727	X			
1	28.6	22.1	315.2	13.278			X	
2	97.9	97.4	2.7	2.4063	X	X		
2	97.2	96.7	5.5	2.7343	X			X
2	93.5	92.2	22.4	4.1921			X	X
2	84.7	81.6	62.4	6.4455		X	X	
2	68.0	61.6	138.2	9.3214		X		X
3	98.2	97.6	3.0	2.3087	X	X		X
3	98.2	97.6	3.0	2.3121	X	X	X	
3	98.1	97.5	3.5	2.3766	X		X	X
3	97.3	96.4	7.3	2.8638		X	X	X
4	98.2	97.4	5.0	2.4460	X	X	X	X

Figura 9.7 Resultado de cómputo (Minitab) para el algoritmo de todas las regresiones posibles de Furnival y Wilson.

Generación eficiente de todas las regresiones posibles

Hay varios algoritmos potencialmente útiles para generar todas las regresiones posibles. Véanse, por ejemplo, Furnival [1971], Furnival y Wilson [1974], Garside [1965, 1971], Morgan y Tatar [1972] y Schatzoff, Tsao y Fienberg [1968]. La idea básica en todos esos algoritmos es efectuar los cálculos para los 2^K modelos de subconjunto posibles, de tal modo que esos modelos secuenciales difieran sólo en una variable. Eso permite usar métodos numéricos muy eficientes para hacer los cálculos. Esos métodos se suelen basar ya sea en la reducción de Gauss-Jordan o en el operador de barrido (véanse Beaton [1964] o Seber [1977]). Algunos de esos algoritmos están disponibles comercialmente. Por ejemplo, el de Furnival y Wilson [1974] son opciones disponibles en los programas de cómputo Minitab y SAS.

En la figura 9.7 se ve un reporte de resultado de Minitab, aplicado a los datos de cemento de Hald; este programa permite que el usuario seleccione el mejor modelo de regresión con subconjuntos para cada tamaño de $1 \leq p \leq K + 1$, y presenta los criterios C_p, R_p^2 y $MS_{\text{Res}}(p)$; también muestra los valores de las estadísticas C_p, R_p^2, $R_{\text{Aj},p}^2$ y $S = \sqrt{MS_{\text{Res}}(p)}$ para varios (pero no todos) los modelos para cada valor de p. El programa tiene la capacidad de identificar los m modelos mejores de regresión de subconjunto (para $m \leq 5$).

Todos los procedimientos actuales posibles de regresión procesan con mucha eficiencia hasta unos 30 regresores candidatos, con tiempos de cómputo comparables a los de los algoritmos usuales de regresión del tipo por etapas, que se describirán en la sección 9.2.2. Nuestra experiencia indica que los problemas con 30 regresores candidatos o menos se pueden resolver, normalmente, con relativa facilidad con el método de todas las regresiones posibles.

9.2.2 Métodos de regresión por segmentos

Como la evaluación de todos los regresores posibles puede ser computacionalmente difícil, se han desarrollado varios métodos para evaluar sólo una pequeña cantidad de modelos de

http://carlos2524.jimdo.com/

regresión con subconjunto, agregando o eliminando regresores uno por uno. A esos métodos se les llama en general **procedimientos del tipo por segmentos**. Pueden clasificarse en tres categorías principales: 1) **selección hacia adelante**, 2) **eliminación hacia atrás**, y 3) **regresión por segmentos**, que es una combinación muy usada de los procedimientos 1 y 2. A continuación se describirán e ilustrarán en forma breve esos procedimientos.

Selección hacia adelante

Este procedimiento comienza con la hipótesis que **no hay regresores en el modelo** además de la ordenada al origen. Se trata de determinar un subconjunto óptimo insertando regresores, uno por uno, en el modelo. El primer regresor que se selecciona para entrar en la ecuación es el que tenga la **máxima correlación simple** con la variable de respuesta y. Supóngase que ese regresor es x_1, éste también es el regresor que producirá el máximo valor de la estadística F en la prueba de significancia de la regresión. El regresor se introduce si la estadística F es mayor que un valor predeterminado de F, por ejemplo F_{IN} (o F para quien entra). El segundo regresor que se escoge para entrar es el que ahora tenga la máxima correlación con y, después de ajustar y por el efecto del primer regresor que se introdujo, x_1. A esas correlaciones se les llama **correlaciones parciales**, que son las correlaciones sencillas entre los residuales de la regresión $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$ y los residuales de las regresiones de los demás regresores candidatos sobre x_1, por ejemplo $\hat{x}_j = \hat{\alpha}_{0j} + \hat{\alpha}_{1j} x_1, j = 2, 3, \ldots, K$.

Supóngase que en el paso 2 el regresor con la máxima correlación parcial con y es x_2, eso implica que la estadística parcial F con mayor valor es

$$F = \frac{SS_R(x_2 | x_1)}{MS_{Res}(x_1, x_2)}$$

Si este valor de F es mayor que F_{IN}, entonces se agrega x_2 al modelo, en general, en cada paso se agrega al modelo el regresor que tenga la máxima correlación parcial con y (o, lo que es igual, la máxima estadística parcial F, dados los demás regresores que ya estén en el modelo), si su estadística parcial F es mayor que el valor preseleccionado para entrar, F_{IN}; el procedimiento termina cuándo la estadística parcial F en determinado paso no es mayor que F_{IN}, o cuándo se ha agregado el último regresor candidato al modelo.

Ejemplo 9.2 *Selección hacia adelante - datos de cemento de Hald*

Se aplicará el procedimiento de selección hacia adelante a los datos de cemento de Hald del ejemplo 9.1. La figura 9.8 muestra los resultados obtenidos cuando se aplicó a esos datos un programa de cómputo, el algoritmo de selección hacia adelante de SAS PROC REG. En este programa, el usuario especifica el valor de corte F_{IN}, eligiendo una tasa de error α tipo I, para que el regresor con la máxima correlación parcial con y se agregue al modelo si su estadística parcial F es mayor que $F_{\alpha,1,n-p}$, en este ejemplo se usará $\alpha = 0.05$ para determinar F_{IN}. En algunos programas de cómputo se requiere seleccionar un valor numérico de F_{IN}, y los que más se acostumbran van de 2.0 a 4.0.

Se ve en la tabla 9.4 que el regresor más correlacionado con y es x_4, porque ($r_{4y} = -0.821$), y como la estadística F asociada con el modelo que usa x_4 es $F = 22.80 > F_{0.05,1,11} = 4.84$, se agrega x_4 a la ecuación. En el paso 2, el regresor que tiene la máxima

http://carlos2524.jimdo.com/

The REG Procedure
Dependent Variable: y

Forward Selection: Step 1 Variable x4 Entered: R-Square = 0.6473 and C(p) = 145.4119

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1620.83817	1620.83817	18.35	0.0016
Error	10	883.29100	88.32910		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	117.36967	6.03895	33365	377.74	< .0001
x4	– 0.73369	0.17128	1620.83817	18.35	0.0016

Bounds on condition number: 1, 1

Forward Selection: Step 2 Variable x1 Entered: R-Square = 0.9704 and C(p) = 6.8727

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2430.01276	1215.00638	147.54	< .0001
Error	9	74.11641	8.23516		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	103.28516	2.32787	16212	1968.60	< .0001
x1	1.44219	0.14549	809.17459	98.26	< .0001
x4	– 0.61850	0.05357	1097.63523	133.29	< .0001

Bounds on condition number: 1.0494, 4.1975

Forward Selection: Step 3 Variable x2 Entered: R-Square = 0.9835 and C(p) = 3.1716

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2462.83753	820.94584	159.05	< .0001
Error	8	41.29164	5.16145		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	67.01422	14.50041	110.24162	21.36	0.0017
x1	1.46153	0.11544	827.35355	160.29	< .0001
x2	– 0.48585	0.19266	32.82477	6.36	0.0357
x4	– 0.48871	0.17563	5.95877	1.15	0.3139

Bounds on condition number: 17.994, 110.63

No other variable met the 0.0500 significance level for entry into the model.

Summary of Forward Selection

Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x4	1	0.6473	0.6473	145.412	18.35	0.0016
2	x1	2	0.3231	0.9704	6.8727	98.26	< .0001
3	x2	3	0.0131	0.9835	3.1716	6.36	0.0357

Figura 9.8 Resultados obtenidos con el algoritmo de selección hacia adelante de SAS PROC REG, ejemplo 9.2.

correlación parcial con y (o la máxima estadística parcial F, dado que x_4 está en el modelo) es x_1, y como la estadística parcial F para este regresor es

$$F = \frac{SS_R(x_1 | x_4)}{MS_{Res}(x_1, x_4)} = \frac{809.1048}{7.4762} = 108.22$$

que es mayor que $F_{IN} = F_{0.05,\,1,\,10} = 4.96$, se agrega x_1 al modelo. En el tercer paso x_2 es el que muestra la máxima correlación parcial con y; la estadística parcial F es

$$F = \frac{SS_R(x_2 | x_1, x_4)}{MS_{Res}(x_1, x_2, x_4)} = \frac{27.7894}{5.3303} = 5.03$$

que es mayor que $F_{IN} = F_{0.05,\,1,\,9} = 5.12$, por lo que x_2 se agrega al modelo, en este punto, el único regresor candidato que queda es x_3, para el cual la estadística parcial F no es mayor que $F_{IN} = R_{0.05,\,1,\,8} = 5.32$, por lo que termina el procedimiento de selección en avance con

$$\hat{y} = 71.6483 + 1.4519x_1 + 0.4161x_2 - 0.2365x_4$$

como el modelo final.

http://carlos2524.jimdo.com/

Eliminación hacia atrás

La selección hacia adelante comienza **sin regresores en el modelo** y trata de introducir variables hasta que se obtiene un modelo adecuado. En la **eliminación hacia atrás** se trata de determinar un buen modelo trabajando en dirección contraria, esto es, se comienza con un modelo que incluya todos los K regresores candidato, a continuación se calcula la estadística parcial F para cada regresor, como si fuera la última variable que entró al modelo. La mínima de esas estadísticas parciales F se compara con un valor preseleccionado, F_{SAL} o F_{OUT} (es decir, F que sale), por ejemplo, y si el valor mínimo de F parcial es menor que F_{OUT}, se quita ese regresor del modelo, ahora se ajusta un modelo de regresión con $K - 1$ regresores, se calculan las estadísticas F parciales para ese nuevo modelo, y se repite el procedimiento. El algoritmo de eliminación en reversa termina cuando el valor mínimo de F parcial no es menor que F_{OUT}, el valor preseleccionado de corte.

La eliminación en reversa suele ser un procedimiento muy bueno de selección de variables. Lo prefieren en especial los analistas que desean ver el efecto de incluir todos los regresores candidato, para que no se pierda nada "obvio".

Ejemplo 9.3 *Eliminación hacia atrás - datos de cemento de Hald*

Se ilustrará la eliminación hacia atrás con los datos de cemento de Hald, del ejemplo 9.1. La figura 9.9 presenta los resultados de aplicar la versión de eliminación hacia atrás del programa SAS PROC REG a esos datos, en esta corrida se ha seleccionado el valor de corte F_{OUT} haciendo que $\alpha = 0.05$. Así, un regresor se elimina si su estadística parcial F es menor que $F_{0.05,\,1,\,n-p}$. El primer paso (Step 0) muestra el resultado de ajustar el modelo completo. El valor mínimo de F parcial es $F = 0.02$, y se asocia con x_3, así, como $F = 0.02 < F_{OUT} = F_{0.05,\,1,\,8} = 3.23$, se elimina x_3 del modelo. En el paso 1 (Step 1), en la figura 9.9, se ven los resultados de ajustar el modelo de tres variables donde intervienen (x_1, x_2, x_4). La F parcial mínima en este modelo, $F = 1.86$, se asocia con x_4, como $F = 1.86 < F_{OUT} = F_{0.05,\,1,\,9} = 5.12$, sale x_4 del modelo. En el paso 2 (Step 2) se ven los resultados de ajustar el modelo con dos variables donde intervienen (x_1, x_2), la estadística F parcial mínima en este modelo es $F = 146.52$, asociada con x_1, y como es mayor que $F_{OUT} = F_{0.05,\,1,\,10} = 4.96$, ya no se pueden quitar más regresores del modelo, por consiguiente, la eliminación en reversa termina y se obtiene el modelo final

$$\hat{y} = 52.5773 + 1.4683x_1 + 0.6623x_2$$

Nótese que es un modelo distinto al que se determinó con la selección hacia adelante, además, es el mismo modelo que se identificó como el mejor, en forma tentativa, con el procedimiento de todas las regresiones posibles.

Regresión por segmentos

Los dos procedimientos que se acaban de describir sugieren varias combinaciones posibles, una de las más usadas es el **algoritmo de regresión por segmentos**, de Efroymson [1960]. La regresión por segmentos es una modificación de la selección hacia adelante, en la que a cada paso se reevalúan todos los regresores que habían entrado antes al modelo, mediante sus estadísticas parciales F. Un regresor agregado en una etapa anterior puede

http://carlos2524.jimdo.com/

The REG Procedure
Dependent Variable: y

Backward Elimination: Step 0 All Variable Entered: R-Square = 0.9839 and C(p) = 5.0000

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2463.82566	615.95642	106.98	< .0001
Error	17	40.30350	5.75764		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	37.87077	6.03895	1.59306	0.28	0.6151
x1	– 1.77030	0.75524	31.63523	5.49	0.0515
x2	0.78462	0.74934	6.31260	1.10	0.3299
x3	0.31636	0.76365	0.98814	0.17	0.6911
x4	0.10303	0.72823	0.11525	0.02	0.8915

Bounds on condition number: 277.33, 2431.4

Backward Elimination: Step 1 Variable x4 Removed: R-Square = 0.9839 and C(p) = 3.0200

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2463.71042	821.23681	162.55	< .0001
Error	8	40.41875	5.05234		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	48.04042	3.80647	804.75231	159.28	< .0001
x1	1.66782	0.20019	350.68522	69.41	< .0001
x2	0.67884	0.04653	1075.38866	212.85	< .0001
x3	0.21188	0.18221	6.83166	1.35	0.2784

Bounds on condition number: 3.2382, 22.304

Backward Elimination: Step 2 Variable x3 Removed: R-Square = 0.9811 and C(p) = 2.2066

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2456.87876	1228.43938	233.99	< .0001
Error	9	47.25041	5.25005		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	51.62412	2.27741	2697.66345	513.84	< .0001
x1	– 1.47601	0.11563	855.48424	162.95	< .0001
x2	0.68673	0.04692	1224.50123	214.19	< .0001

Bounds on condition number: 1.0397, 4.1587

All variables left in the model are sig nificant at the 0.0500 level.

Summary of Backward Elimination

Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x4	3	0.0000	0.9839	3.0200	0.02	0.8915
2	x3	2	0.0027	0.9811	2.2066	1.35	0.2784

Figura 9.9 Resultados obtenidos con el algoritmo de eliminación hacia atrás de SAS PROC REG, ejemplo 9.2.

volverse redundante, debido a las relaciones entre él y los regresores que ya estén en la ecuación. Si la estadística parcial F de una variable es menor que F_{OUT}, esa variable se elimina del modelo.

En la regresión por segmentos se requieren dos valores de corte, F_{IN} y F_{OUT}, algunos analistas prefieren definir $F_{IN} = F_{OUT}$, aunque eso no es necesario, con frecuencia se opta por $F_{IN} > F_{OUT}$, con lo que se hace algo más difícil agregar un regresor que eliminar uno.

Ejemplo 9.4 *Regresión por segmentos - datos de cemento de Hald*

La figura 9.10 presenta los resultados de usar el algoritmo de regresión por segmentos de SAS PROC REG para los datos de cemento de Hald. Se especificó el nivel $\alpha = 0.05$ para agregar o eliminar un regresor. En el paso 1 (Step 1) el procedimiento comienza sin variables en el modelo, y trata de agregar x_4. Como la estadística parcial F en este paso es mayor que $F_{IN} = F_{0.05,\,1,\,11} = 4.84$, x_4 se agrega al modelo. En el paso 2 (Step 2) se agrega x_1 al modelo, si el valor de F parcial para x_4 es menor que $F_{OUT} = F_{0.05,\,1,\,10} = 4.96$, x_4 se eliminaría; sin embargo, el valor de F parcial para x_4 en el paso 2 es $F = 159.30$, por lo que x_4 se conserva. En el paso 3 (Step 3) de regresión adiciona el algoritmo x_2 para el modelo, entonces se comparan los estadísticos parciales F para x_1 y x_4 con $F_{SAL} = F_{0.05,\,1,\,9} = 5.12$.

http://carlos2524.jimdo.com/

The REG Procedure
Dependent Variable: y

Stepwise Selection: Step 1 Variable x4 Entered: R-Square = 0.6473 and C(p) = 145.4119

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1620.83817	1620.83817	18.35	0.0016
Error	10	883.29100	88.32910		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	117.36967	6.03895	33365	377.74	< .0001
x4	– 0.73369	0.17128	1620.83817	18.35	0.0016

Bounds on condition number: 1, 1

Stepwise Selection: Step 2 Variable x1 Entered: R-Square = 0.9704 and C(p) = 6.8727

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2430.01276	1215.00638	147.54	< .0001
Error	9	74.11641	8.2316		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	103.28516	2.32787	16212	1968.60	< .0001
x1	1.44219	0.14549	809.17459	98.26	< .0001
x4	– 0.61850	0.05357	1097.63523	133.29	< .0001

Bounds on condition number: 1, 0494, 4.1975

Stepwise Selection: Step 3 Variable x2 Entered: R-Square = 0.99835 and C(p) = 3.1716

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2462.83753	820.94584	159.05	< .0001
Error	8	41.29164	5.16145		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	67.01422	14.50041	110.24162	21.36	0.0017
x1	1.46153	0.11544	827.35355	160.29	< .0001
x2	1.48585	0.19266	32.82477	6.36	0.0357
x4	– 0.18871	0.17563	5.95877	1.15	0.3139

Bounds on condition number: 17.994, 110.63

Stepwise Selection: Step 4 Variable x4 Removed: R-Square = 0.9811 and C(p) = 2.2066

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2456.87876	1228.43938	233.99	< .0001
Error	9	47.25041	5.25005		
Corrected Total	11	2504.12917			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	51.62412	2.27741	2697.66345	513.84	< .0001
x1	1.47601	0.11563	855.48424	162.95	< .0001
x2	0.68673	0.04692	1124.50123	214.19	< .0001

Bounds on condition number: 1.0397, 4.1587

All variables left in the model are significant at the 0.0500 level.
No other variable met the 0.0500 significance level for entry into the model.

Summary of Stepwise Selection.

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x4		1	0.6473	0.6473	145.412	18.35	0.0016
2	x1		2	0.3231	0.9704	6.8727	98.26	< .0001
3	x2		3	0.0131	0.9835	3.1716	6.36	0.0357
4		x4	2	0.0024	0.9811	2.2066	1.15	0.3139

Figura 9.10 Resultados obtenidos con la regresión por segmentos de SAS PROC REG, ejemplo 9.4.

Como se ve que x_4 tiene un valor F parcial de 1.86, que es menor que $F_{OUT} = 3.36$, x_4 se elimina. El paso 4 (Step 4) muestra el resultado de eliminar x_4 del modelo, en este punto, el único regresor candidato que queda es x_3, que no se puede agregar porque su valor de F parcial no es mayor que el de F_{IN}, por consiguiente, la regresión por segmentos termina con el modelo

$$\hat{y} = 52.5773 + 1.4683x_1 + 0.6623x_2$$

Es la misma ecuación que se identificó con los procedimientos de todas las regresiones posibles y de eliminación hacia atrás.

http://carlos2524.jimdo.com/

Comentarios generales sobre los procedimientos por etapas

Los algoritmos de regresión por segmentos que se describieron arriba han sido criticados en diversos medios, el más común es que ninguno de ellos garantiza en general que se identifique el mejor modelo de subconjunto para tamaño alguno, además, como todos los procedimientos por etapas terminan con una ecuación final, los analistas sin experiencia podrán llegar a la conclusión que se ha determinado un modelo que es óptimo en algún sentido; parte del problema es que es probable que no haya un mejor modelo de subconjunto, sino que haya varios igualmente buenos.

El analista también debe tener en cuenta que el orden en el que entran o salen los regresores del modelo no necesariamente implica un orden de importancia de los regresores. No es raro ver que un regresor que entró al modelo al principio se vuelve sin importancia en un paso posterior; esto se ve con los datos de cemento de Hald, para los cuales la selección hacia adelante escoge a x_4 como el primer regresor en entrar, sin embargo, cuando se agrega x_2 en un paso posterior, x_4 ya no se requiere, por la gran intercorrelación entre x_2 y x_4, esto, de hecho, es un problema **general** con el procedimiento de selección hacia adelante, una vez agregado un regresor, no se puede eliminar en un paso posterior.

Nótese que la selección hacia adelante, la eliminación hacia atrás y la regresión por segmentos no necesariamente conducen a la **misma** elección del modelo final. La intercorrelación entre los regresores afecta el orden de entrada y la eliminación, por ejemplo, al usar los datos de cemento de Hald se vio que los regresores seleccionados por cada procedimiento fueron los siguientes:

Selección hacia adelante	x_1	x_2	x_4
Eliminación hacia atrás	x_1	x_2	
Regresión por segmentos	x_1	x_2	

Algunos usuarios han recomendado que se apliquen todos los procedimientos, con la esperanza de encontrar alguna concordancia o aprender algo acerca de la estructura de los datos, que pudiera haberse escapado si se usa sólo un procedimiento de selección, además, no necesariamente hay concordancia entre cualquiera de los procedimientos del tipo por segmentos y todas las regresiones posibles, sin embargo, Berk [1978] hace notar que la selección hacia adelante tiende a concordar con todas las regresiones posibles para tamaños de subconjunto pequeños, pero no para los grandes, mientras que la eliminación hacia atrás tiende a concordar con todas las regresiones posibles para tamaños de subconjunto grandes, pero no para los pequeños.

Por esas razones, se deben usar con precaución los procedimientos de selección de variable del tipo en etapas. Nosotros preferimos el algoritmo de regresión por segmentos seguido de la eliminación para atrás. Frecuentemente el algoritmo de eliminación hacia atrás se afecta menos, por la estructura correlativa de los regresores, que la selección hacia adelante (véase Mantel [1970)].

Reglas de parada para procedimientos en etapas

Se puede imaginar que la elección de valores de corte F_{IN} y/o F_{OUT} en los procedimientos del tipo en etapas es la especificación de una regla para que los algoritmos se detengan. Algunos programas de cómputo permiten que el analista especifique en forma directa esos números, mientras que en otros se requiere escoger una tasa α de error tipo I, para generar F_{IN} y/o F_{OUT}, sin embargo, como el valor de F parcial examinado en cada etapa es el

http://carlos2524.jimdo.com/

máximo de varias variables F de correlación parcial, es engañoso imaginar que α sea un nivel de significancia de la tasa de errores tipo I; algunos autores (por ejemplo, Draper, Guttman y Kanemasa [1971], y Pope y Webster [1972]) han investigado este problema, y se ha avanzado poco en encontrar las condiciones bajo las que tiene un significado el nivel "anunciado" de significación de F, o en desarrollar la distribución exacta de las estadísticas F para entrar y F para eliminar.

Algunos usuarios prefieren elegir valores de F_{IN} y F_{OUT} relativamente pequeños, por lo que de ordinario se rechazarían varios regresores más en comparación a cuando se usan valores de F más conservadores. En el caso extremo se pueden escoger F_{IN} y F_{OUT} de tal modo que todos los regresores entren en la selección hacia adelante, o se eliminen en la eliminación para atrás, dejando sólo un modelo de subconjunto de cada tamaño $p = 2, 3, \ldots, K + 1$. Esos modelos de subconjunto se pueden evaluar entonces mediante criterios tales como C_p o MS_{Res}, para determinar el modelo final. Esta estrategia extrema no la recomendamos porque el analista puede pensar que los subconjuntos así determinados son óptimos en algún sentido, cuando lo probable es que se haya pasado por alto el mejor modelo de subconjunto. Un procedimiento muy empleado es igualar $F_{IN} = F_{OUT} = 4$, porque eso corresponde más o menos al punto superior de 5% en la distribución F, otra posibilidad es hacer varias corridas usando distintos valores de F_{IN} y F_{OUT}, y observar el efecto de la elección de criterios sobre los subconjuntos obtenidos.

Hay varios estudios con el fin de proporcionar lineamientos prácticos en la elección de las reglas de paro. Bendel y Afifi [1974] recomiendan $\alpha = 0.25$ en la selección en avance, eso daría como resultado, en el caso usual, un valor numérico de F_{IN} de 1.3 a 2. Kennedy y Bancroft [1971] también sugieren que $\alpha = 0.25$ para la selección hacia adelante y recomiendan que $\alpha = 0.10$ para la eliminación hacia atrás. La elección de los valores de F_{IN} y F_{OUT} es, principalmente, cosa de preferencias personales del analista, y en este respecto con frecuencia se toma un gran margen de seguridad.

9.3 ALGUNAS RECOMENDACIONES FINALES PARA LA PRÁCTICA

En este capítulo se han descrito varios procedimientos de selección de variable en regresión lineal. Los métodos se pueden clasificar, en general, como del tipo por etapas o de todas las regresiones posibles (con sus variaciones). Las ventajas principales de los métodos del tipo por etapas son el ser rápidos, fáciles de implementar en computadoras digitales y que se consiguen con facilidad para casi todos los sistemas de cómputo. Sus desventajas son que no producen modelos de subconjunto que sean necesariamente los mejores con respecto a algún criterio común, y además, como se orientan hacia la producción de una sola ecuación final, el usuario no versado puede ser conducido a creer que ese modelo sea óptimo, en algún sentido. Por otra parte, todas las regresiones posibles identificarán los modelos de subconjunto con respecto a cualquier criterio que imponga el analista. Cuando los regresores candidatos son hasta unos 20 o 30, el costo de cómputo, con todas las regresiones posibles, es más o menos igual que con los procedimientos por etapas, sin embargo, los programas de todas las regresiones posibles no se consiguen con tanta facilidad como los de los métodos en etapas, en especial para computadoras personales (al menos en la actualidad).

Cuando la cantidad de regresores candidato es demasiado grande como para emplear de entrada el método de todas las regresiones posibles, recomendamos una **estrategia de dos etapas**. Se pueden usar los métodos por etapas para "filtrar" los regresores candidato y eliminar los que tengan efectos despreciables, para obtener una lista menor de regresores candidato; este conjunto reducido de candidatos regresores se puede investigar a continuación

http://carlos2524.jimdo.com/

para todas las regresiones posibles. El analista siempre debe recurrir a sus conocimientos del entorno del problema y al sentido común para evaluar los regresores candidato; cuando se tiene una gran lista de regresores candidato, suele ser provechoso dedicar una meditación seria antes de ir a la computadora. Con frecuencia se ve que algunas variables se pueden eliminar con base en la lógica o en el sentido técnico.

Se han descrito varios criterios formales para evaluar modelos de regresión de subconjuntos como el de la estadística C_p de Mallows y el cuadrado medio residual, sin embargo, en general no es clara la elección de un modelo final, además de los criterios formales de evaluación, sugerimos que el analista se pregunte lo siguiente:

1. ¿Es razonable la ecuación? Esto es, ¿tienen sentido los regresores en el modelo, considerando el entorno del problema?
2. ¿Es útil el modelo para el objetivo que se pretendía? Por ejemplo, un modelo necesario para predicción que contenga un regresor que sea inobservable cuando se requiera la predicción, no se puede usar. Si el costo de reunir datos de un regresor es prohibitivo, eso también haría que no se pudiera usar el modelo.
3. ¿Son razonables los coeficientes de regresión? ¿Son realistas los signos y las magnitudes de los coeficientes, y los errores estándar son relativamente pequeños?
4. ¿Son satisfactorias las comprobaciones comunes de diagnóstico de adecuación de modelo? Por ejemplo, ¿indican las gráficas de residuales la estructura o los valores atípicos, o hay uno o más puntos de gran balanceo que estén controlando el ajuste?

Si se piensa en serio sobre esas preguntas y se aplican las respuestas en forma estricta, en algunos casos (quizá en muchos) **no** haya una ecuación final de regresión satisfactoria. Es claro que se requieren el juicio y la experiencia del entorno de operación del modelo pretendido. Por último, aunque la ecuación ajuste bien a los datos, y pase las pruebas normales de diagnóstico, no hay seguridad de que vaya a predecir con exactitud nuevas observaciones. Recomendamos evaluar la **capacidad predictiva** de un modelo, observando su funcionamiento con datos nuevos que no se hayan usado para formar el modelo. Si eso no se puede hacer con facilidad, se deben apartar algunos de los datos originales para este fin. Esto se explicará con más detalle en el capítulo 15.

PROBLEMAS

9.1 Para los datos de la Liga Nacional de Futbol en la tabla B.1 del apéndice:

a. Usar el algoritmo de selección en avance para seleccionar un modelo de regresión de subconjunto.

b. Usar el algoritmo de selección en reversa para seleccionar un modelo de regresión de subconjunto.

c. Usar la regresión por pasos para seleccionar un modelo de regresión de subconjunto.

d. Comentar el modelo final escogido por estos tres procedimientos.

9.2 Para los datos de la Liga Nacional de Futbol de la tabla B.1 en el apéndice, restringir la atención a los regresores x_1 (yardas por tierra), x_2 (yardas por aire), x_4 (porcentaje de goles de campo), x_7 (porcentaje de carreras), x_8 (yardas por tierra del contrario) y x_9 (yardas por aire del contrario); aplicar el procedimiento de todas las regresiones posibles. Evaluar R_p^2, C_p y MS_{Res} para cada modelo. ¿Cuál modelo de subconjunto recomienda el lector?

http://carlos2524.jimdo.com/

9.3 En la regresión por segmentos se especifica que $F_{IN} \geq F_{OUT}$. Justificar esta elección de valores de corte.

9.4 Para los datos de la prueba de energía solar en la tabla B.2 del apéndice:

a. Usar la selección en avance para especificar un modelo de regresión de subconjunto.

b. Aplicar la eliminación en reversa para especificar un modelo de regresión de subconjunto.

c. Usar regresión por pasos para especificar un modelo de regresión de subconjunto.

d. Aplicar el método de todas las regresiones posibles a los datos. Evaluar R_p^2, C_p y MS_{Res} para cada modelo. ¿Cuál modelo de subconjunto recomienda el lector?

e. Comparar y describir las diferencias de los modelos producidos por las estrategias de selección de variables en las partes a a d.

9.5 Examinar los datos de rendimiento de gasolina en la tabla B.3 del apéndice.

a. Aplicar el método de todas las regresiones posibles para determinar un modelo adecuado de regresión.

b. Aplicar la regresión por segmentos para especificar un modelo de subconjunto de regresión. ¿Se llega así al mismo modelo que se determinó en la parte a?

9.6 Para los datos de valuación catastral de la tabla B.4 del apéndice:

a. Aplicar el método de todas las regresiones posibles para calcular el "mejor" conjunto de regresores.

b. Usar la regresión por etapas para seleccionar un modelo de sobconjuntos de regresión. ¿Concuerda este modelo con el determinado en la parte a?

9.7 Usar la regresión por segmentos con $F_{IN} = F_{OUT} = 4.0$ para determinar el "mejor" conjunto de variables regresoras de los datos de licuefacción en Belle Ayr, en la tabla B.5 del apéndice. Repetir el análisis con $F_{IN} = F_{OUT} = 2.0$. ¿Hay diferencias sustanciales en los modelos que se obtuvieron?

9.8 Usar el método de todas las regresiones posibles para seleccionar un modelo de regresión de subconjunto para los datos de licuefacción en Belle Ayr, en la tabla B.5 del apéndice. Evaluar los modelos de subconjunto con el criterio C_p. Justificar la elección del modelo final aplicando las pruebas normales de adecuación del modelo.

9.9 Analizar los datos del reactor de tubo de flujo, en la tabla B.6 del apéndice, mediante todas las regresiones posibles. Evaluar los modelos de subconjunto con los criterios R_p^2, C_p y MS_{Res}. Justificar la elección del modelo final con las comprobaciones comunes de adecuación de modelo.

9.10 Gorman y Toman [1966] presentan datos acerca de la profundidad de rodada de 31 pavimentos de asfalto, preparados bajo distintas condiciones especificadas por cinco regresores. Como variable indicadora para separar los datos en dos conjuntos de corridas se usa un sexto regresor, las variables son y = log(cambio de profundidad de rodada/millón de pasos de rueda), x_1 = log(viscosidad del asfalto), x_2 = porcentaje de asfalto en el tendido superficial, x_3 = porcentaje de asfalto en la capa de base, x_4 = variable indicadora, x_5 = porcentaje de finos en la capa superficial y x_6 = porcentaje de huecos en la capa superficial. Los datos se ven en la siguiente tabla.

http://carlos2524.jimdo.com/

Observación i	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}
1	0.82930	0.44716	4.68	4.87	– 1.00	8.4	4.916
2	1.11394	0.14613	5.19	4.50	– 1.00	6.5	4.563
3	1.16879	0.14613	4.82	4.73	– 1.00	7.9	5.321
4	1.10037	0.51851	4.85	4.76	– 1.00	8.3	4.865
5	0.91645	0.23045	4.86	4.95	– 1.00	8.4	3.776
6	1.02816	0.46240	5.16	4.45	– 1.00	7.4	4.397
7	0.86213	0.56820	4.82	5.05	– 1.00	6.8	4.867
8	1.10278	0.23045	4.86	4.70	– 1.00	8.6	4.828
9	1.09968	– 0.03621	4.78	4.84	– 1.00	6.7	4.865
10	1.31387	– 0.16749	5.16	4.76	– 1.00	7.7	4.034
11	0.55388	0.77815	4.57	4.82	– 1.00	7.4	5.450
12	0.84510	0.63347	4.61	4.65	– 1.00	6.7	4.853
13	1.41830	– 0.22185	5.07	5.10	– 1.00	7.5	4.257
14	1.06707	0.25537	4.66	5.09	– 1.00	8.2	5.144
15	0.88479	0.77815	5.42	4.41	– 1.00	5.8	3.718
16	1.08814	0.64345	5.01	4.74	– 1.00	7.1	4.715
17	– 0.11919	1.94448	4.97	4.66	1.00	6.5	4.625
18	0.13033	1.79239	4.01	4.72	1.00	8.0	4.977
19	0.15836	1.69897	4.96	4.90	1.00	6.8	4.322
20	0.20412	1.76343	5.20	4.70	1.00	8.2	5.087
21	0.04139	1.95424	4.80	4.60	1.00	6.6	5.971
22	0.07058	1.81954	4.98	4.69	1.00	6.4	4.647
23	0.07918	2.14613	5.35	4.76	1.00	7.3	5.115
24	0.25181	2.38021	5.04	4.80	1.00	7.8	5.939
25	0.14267	2.62325	4.80	4.80	1.00	7.4	5.916
26	0.32790	2.69897	4.83	4.60	1.00	6.7	5.471
27	0.48149	2.25527	4.66	4.72	1.00	7.2	4.602
28	0.58503	2.43136	4.67	4.50	1.00	6.3	5.043
29	0.11919	2.23045	4.72	4.70	1.00	6.8	5.075
30	0.09691	1.99123	5.00	5.07	1.00	7.2	4.334
31	0.30102	1.54407	4.70	4.80	1.00	7.7	5.705

a. Aplicar el método de todas las regresiones posibles para determinar el mejor modelo de subconjuntos para la profundidad de rodada. Usar C_p como criterio.

b. Repetir la parte a, usando MS_{Res} como criterio. ¿Se llegó al mismo modelo?

c. Usar la regresión por segmentos para determinar el mejor modelo de subconjunto. ¿Se llegó a la misma ecuación que la encontrada en la parte a o en la b anteriores?

9.11 Acerca del análisis de todas las regresiones posibles para los datos de cemento de Hald, en el ejemplo 9.1, si el objetivo es desarrollar el modelo para predecir observaciones nuevas, ¿cuál ecuación se recomendaría, y por qué?

9.12 Para el análisis de los datos de la Liga Nacional de Futbol en el problema 9.2, por el método de todas las regresiones posibles, identificar los modelos de regresión de subconjunto que sean adecuados (0.05) en R^2.

9.13 Suponer que el modelo completo es $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$, $i = 1, 2, \ldots, n$, donde se han codificado x_{i1} y x_{i2} de modo que $S_{11} = S_{22} = 1$. También se considerará el ajuste de un modelo de subconjunto, por ejemplo $y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$.

http://carlos2524.jimdo.com/

a. Sea $\hat{\beta}_1^*$ el estimado de β_1 por mínimos cuadrados del modelo completo. Demostrar que Var $(\hat{\beta}_1^*) = \sigma^2/(1 - r_{12}^2)$, donde r_{12} es la correlación entre x_1 y x_2.

b. Sea $\hat{\beta}_1$ el estimado de β_1, por mínimos cuadrados, del modelo de subconjunto. Demostrar que Var $(\hat{\beta}_1) = \sigma^2$. ¿Se estima β_1 con más precisión a partir del modelo de subconjunto, o a partir del modelo completo?

c. Demostrar que $E(\hat{\beta}_1) = \beta_1 + r_{12}\beta_2$. ¿Bajo qué condiciones $\hat{\beta}_1$ es un estimador insesgado de β_1?

d. Calcular el error cuadrático medio para el estimador $\hat{\beta}_1$ de subconjunto. Comparar MSE$(\hat{\beta}_1)$ con Var$(\hat{\beta}_1^*)$. ¿Bajo qué condiciones $\hat{\beta}_1$ es un estimador preferible desde el punto de vista de MSE?
Podrá ser útil repasar la sección 9.12.

9.14 En la tabla B.11 del apéndice se presentan datos de la calidad del vino Pinot Noir.

a. Formar un modelo adecuado de regresión con el método de todas las regresiones posibles. Usar C_p como criterio de selección del modelo, e incorporar la información de la región mediante variables indicadoras.

b. Para los dos mejores modelos obtenidos, en términos de C_p, investigar la adecuación del modelo usando gráficas de residuales. ¿Hay alguna base práctica para seleccionar entre esos modelos?

c. ¿Hay diferencia entre los dos modelos de la parte b, en lo que se refiere a la estadística PRESS?

9.15 Usar los datos de calidad de vino en la tabla B.11 del apéndice para formar un modelo de regresión para la calidad, con el método de regresión por segmentos. Comparar este modelo con el que se determinó en la parte a del problema 9.14.

9.16 Repetir el problema 9.14, pero excluir la información de la región.

a. Comentar la diferencia en los modelos encontrados. ¿Hay indicios de que la información de la región mejora el modelo en forma apreciable?

b. Calcular los intervalos de confianza como medida de calidad para todos los puntos del conjunto de datos, usando los modelos de la parte a de este problema y la parte a del problema 9.14. De acuerdo con este análisis, ¿cuál modelo sería preferible?

9.17 La tabla B.12 del apéndice presenta datos sobre un proceso de tratamiento térmico para carburizar engranajes. El espesor de la capa cementada es un factor crítico en la fiabilidad general de los engranajes. La variable de respuesta y = PITCH es el resultado de un análisis de carbono en el flanco de los dientes, en una parte cortada transversalmente. Usar todas las regresiones posibles y el criterio de C_p para determinar un modelo adecuado de regresión para esos datos. Investigar la adecuación del modelo mediante gráficas de residuales.

9.18 Volver a los datos de tratamiento térmico de la tabla B.12 en el apéndice. Ajustar un modelo a la respuesta PITCH con las variables

$$x_1 = SOAKTIME \times SOAKPCT \text{ y } x_2 = DIFFTIME \times DIFFPCT$$

como regresores. ¿Cómo se compara este modelo al encontrado que se determinó con el método de todas las regresiones posibles en el problema 9.17?

9.19 Repetir el problema 9.17 usando las dos variables de productos cruzados definidas en el problema 9.18, como regresores candidato adicionales. Comentar el modelo que se determinó.

http://carlos2524.jimdo.com/

9.20 Comparar los modelos que se encontraron en los problemas 9.17, 9.18 y 9.19, calculando los intervalos de confianza de la media de la respuesta PITCH para todos los puntos del conjunto original de datos. De acuerdo con una comparación de esos intervalos de confianza, ¿cuál modelo se prefiere? Ahora calcular la estadística PRESS para esos modelos. ¿Cuál modelo indicado por PRESS es probablemente el mejor para predecir nuevas observaciones de PITCH?

9.21 La tabla B.13 del apéndice contiene datos sobre el empuje de una turbina de reacción, y seis regresores candidato. Usar todas las regresiones posibles y el criterio C_p para determinar un modelo de regresión adecuado para esos datos. Investigar la adecuación del modelo empleando gráficas de residuales.

9.22 Regresar a los datos de empuje del motor de turbina de reacción, en la tabla B.13 del apéndice. Aplicar la regresión por segmentos para determinar un modelo adecuado de regresión para esos datos. Investigar la adecuación del modelo usando gráficas de residuales. Comparar este modelo con el encontrado con el método de todas las regresiones posibles, del problema 9.21.

9.23 Comparar los dos mejores modelos que se hayan encontrado en el problema 9.21, respecto a C_p, calculando los intervalos de confianza de la media del empuje (la respuesta) para todos los puntos del conjunto original de datos. Con base en una comparación de esos intervalos de confianza, ¿cuál es el modelo que se prefiere? Ahora calcular la estadística PRESS para esos modelos. ¿Cuál modelo indicado por PRESS es probable que sea el mejor para predecir nuevas observaciones del empuje?

9.24 La tabla B.14 del apéndice presenta datos de los puntos transitorios de un inversor electrónico. Usar todas las regresiones posibles y el criterio C_p para determinar un modelo adecuado de regresión para esos datos. Investigar la adecuación del modelo mediante gráficas de residuales.

9.25 De nuevo para los datos del inversor electrónico, en la tabla B.14 del apéndice, aplicar la regresión por segmentos para determinar un modelo adecuado de regresión para esos datos. Investigar la adecuación del modelo usando gráficas de residuales. Comparar este modelo con el encontrado con el método de todas las regresiones posibles en el problema 9.24.

9.26 Comparar los dos mejores modelos que se determinaron en el problema 9.24 con el criterio C_p, calculando los intervalos de confianza del promedio de la respuesta para todos los puntos en el conjunto original de datos. De acuerdo con una comparación de esos intervalos de confianza, ¿cuál modelo se prefiere? Ahora calcular la estadística PRESS para esos modelos. ¿Cuál modelo de los que indica PRESS probablemente sea el mejor para predecir nuevas respuestas de observaciones?

9.27 Volver a los datos del inversor electrónico en la tabla B.14 del apéndice. En los problemas 9.24 y 9.25 se formaron modelos de regresión para esos datos, con distintos algoritmos de selección de variable. Suponer que ahora se conoce que la segunda observación fue anotada en forma incorrecta, y que se debe pasar por alto.

a. Ajustar un modelo a los datos modificados, usando todas las regresiones posibles y C_p como criterio. Comparar este modelo con el determinado en el problema 9.24.

b. Usar la regresión por segmentos para determinar un modelo adecuado para los datos modificados. Comparar este modelo al determinado en el problema 9.25.

c. Calcular los intervalos de confianza de la respuesta media para todos los puntos en el conjunto modificado de datos. Comparar estos resultados con los intervalos de confianza del problema 9.26. Comentar lo encontrado.

http://carlos2524.jimdo.com/

9.28 Para los datos del inversor electrónico de la tabla B.14 del apéndice, omitir la observación 2 de esos datos. De acuerdo con la teoría eléctrica, se deberían definir nuevas variables como sigue: $y^* = \ln y$, $x_1^* = 1/\sqrt{x_1}$, $x_2^* = \sqrt{x_2}$, $x_3^* = 1/\sqrt{x_3}$ y $x_4^* = \sqrt{x_4}$.

a. Determine un modelo de subconjunto adecuado para esos datos, usando todas las regresiones posibles y el criterio C_p.

b. Graficar los residuales en función de $\hat{y}^*$ para este modelo. Comentar las gráficas.

c. Describir cómo se podría comparar este modelo con los que se formaron con la respuesta y los regresores originales en el problema 9.27.

http://carlos2524.jimdo.com/

CAPÍTULO 10

MULTICOLINEALIDAD

10.1 INTRODUCCIÓN

El uso y la interpretación de un modelo de regresión múltiple dependen, con frecuencia, en forma explícita o implícita, de los estimados de los coeficientes individuales de regresión. Entre los ejemplos de las inferencias que se hacen a menudo están:

1. Identificación de los efectos relativos de las variables regresoras.
2. Predicción y/o estimación.
3. Selección de un conjunto adecuado de variables para el modelo.

Si no hay relación lineal entre los regresores, se dice que éstos son **ortogonales**. Cuando los regresores son ortogonales se pueden hacer con relativa facilidad inferencias como las de arriba, desafortunadamente, en la mayor parte de las aplicaciones de regresión, los regresores no son ortogonales. A veces no es grave la falta de ortogonalidad, sin embargo, en algunos casos los regresores tienen una relación lineal casi perfecta, y en esos casos, las inferencias basadas en el modelo de regresión pueden ser engañosas o erróneas. Cuando hay **dependencias casi lineales** entre los regresores, se dice que existe el problema de **multicolinealidad**.

En este capítulo se ampliará la descripción preliminar de la multicolinealidad que comenzó en el capítulo 3, y se describirá una diversidad de problemas y técnicas relacionados con este problema. En forma específica, se examinarán las causas de la multicolinealidad, algunos de sus efectos específicos sobre la inferencia, los métodos para detectar la presencia de multicolinealidad y algunas técnicas para enfrentar el problema.

10.2 FUENTES DE MULTICOLINEALIDAD

Se escribirá el modelo de regresión múltiple en la forma

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{10.1}$$

en donde $\mathbf{y}$ es un vector de $n \times 1$ de respuestas, $\mathbf{X}$ es una matriz de $n \times p$ de las variables regresoras, β es un vector de $p \times 1$ de las constantes desconocidas y ε es un vector de $n \times 1$ de los errores aleatorios, siendo $\varepsilon_i \sim \text{NID}(0, \sigma^2)$. Será conveniente suponer que las variables regresoras y la respuesta se han centrado y escalado a longitud unitaria, como en la sección 3.7, en consecuencia, $\mathbf{X}'\mathbf{X}$ es una matriz de correlaciones[1] de $p \times p$, entre los

[1] Se acostumbra llamar coeficientes de correlación a los elementos no diagonales de $\mathbf{X}'\mathbf{X}$, aunque los regresores no necesariamente son variables aleatorias.

http://carlos2524.jimdo.com/

regresores, y $\mathbf{X}'\mathbf{y}$ es un vector de $p \times 1$, de correlaciones entre los regresores y la respuesta.

Sea $\mathbf{X}_j$ la j-ésima columna de la matriz $\mathbf{X}$, de modo que $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p]$, entonces, $\mathbf{X}_j$ contiene los n niveles de la j-ésima variable regresora. Se definirá, formalmente, la multicolinealidad en términos de la dependencia lineal de las columnas de $\mathbf{X}$. Los vectores $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p$ son linealmente dependientes si hay un conjunto de constantes $t_1, t_2, \ldots, t_p$, no todas cero, tales que[2]

$$\sum_{j=1}^{p} t_j \mathbf{X}_j = \mathbf{0} \tag{10.2}$$

Si la ecuación (10.2) es exactamente válida para un subconjunto de las columnas de $\mathbf{X}$, el rango de la matriz $\mathbf{X}'\mathbf{X}$ es menor que p, y no existe $(\mathbf{X}'\mathbf{X})^{-1}$. Sin embargo, supóngase que la ecuación (10.2) es aproximadamente válida para algún subconjunto de las columnas de $\mathbf{X}$. En ese caso habrá una dependencia casi lineal en $\mathbf{X}'\mathbf{X}$, y se dice que existe el problema de multicolinealidad, nótese que la multicolinealidad es una forma de deterioramiento en la matriz $\mathbf{X}'\mathbf{X}$, además, el problema es de grado, indicando con esto que cada conjunto de datos sufrirá cierto grado de multicolinealidad, a menos que las columnas de $\mathbf{X}$ sean ortogonales (que $\mathbf{X}'\mathbf{X}$ sea una matriz diagonal), en general, eso sólo sucederá en un experimento diseñado. Como se verá, la presencia de multicolinealidad puede hacer que el análisis del modelo de regresión, por mínimos cuadrados, sea terriblemente inadecuado.

Hay cuatro **fuentes de multicolinealidad** principales:

1. El método de recolección de datos que se empleó.
2. Restricciones en el modelo o en la población.
3. Especificación del modelo.
4. Un modelo sobredefinido.

Es importante comprender las diferencias entre estas fuentes de multicolinealidad, porque los datos y la interpretación del modelo resultante dependen, en cierto grado, de la causa del problema (véase Mason, Gunst y Webster [1975], para más descripción de las fuentes de multicolinealidad).

El **método de obtención de datos** puede originar problemas de multicolinealidad cuando el analista sólo muestrea un subespacio de la región de los regresores definidos (en forma aproximada) por la ecuación (10.2), por ejemplo, considérense los datos de tiempo de entrega de bebidas gaseosas, que se describieron en el ejemplo 3.1. El espacio de las variables regresoras "cajas" y "distancia", así como el subespacio de esa región que se muestreó, se ven en la matriz de diagramas de dispersión, figura 3.4. Nótese que los pares (cajas, distancia) de muestra caen en forma aproximada a lo largo de una recta; en general, si hay más de dos regresores, los datos estarán, en forma aproximada, a lo largo de un hiperespacio definido por la ecuación (10.2). En este ejemplo, las observaciones con una cantidad pequeña de cajas frecuentemente tienen una distancia pequeña, mientras que las observaciones con una cantidad grande de cajas tienen, por lo general, una distancia grande. Así, cajas y distancia tienen correlación positiva, y si esa correlación es suficientemente fuerte, habrá problema de multicolinealidad. La multicolinealidad causada por la técnica

[2] Si los regresores no están centrados, entonces el $\mathbf{0}$ de la ecuación (10.2) se transforma en un vector de constantes $\mathbf{m}$, no todas necesariamente iguales a 0.

http://carlos2524.jimdo.com/

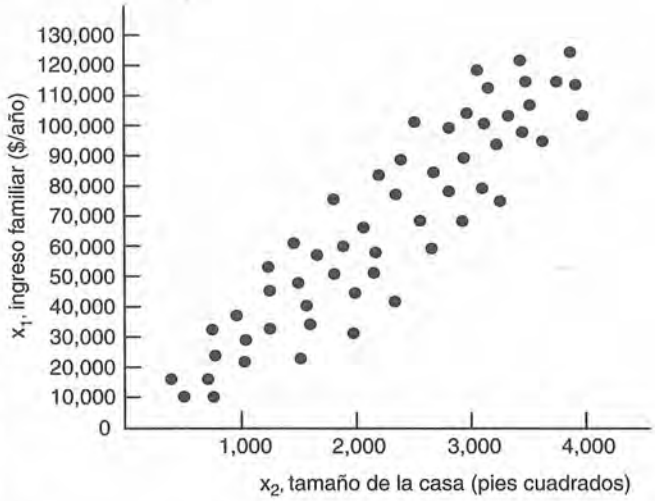


Figura 10.1 Niveles de ingreso familiar y tamaño de vivienda, para un estudio de consumo residencial de electricidad.

de muestreo no es inherente al modelo o a la población que se muestrea, por ejemplo, en el problema del tiempo de entrega se podrían reunir datos con una cantidad pequeña de cajas y una distancia grande; no hay nada en la estructura física del problema que lo evite.

Las **restricciones** en el modelo o en la población que se muestrea pueden causar multicolinealidad; por ejemplo, supóngase que una empresa eléctrica está investigando el efecto del ingreso familiar (x_1) y el tamaño de la vivienda (x_2) sobre el consumo eléctrico residencial. Los niveles de las dos variables regresoras, obtenidos en los datos de muestreo, se ven en la figura 10.1, nótese que los datos están más o menos a lo largo de una recta, e indican que hay un problema potencial de multicolinealidad. En este ejemplo, una restricción física en la población fue lo que causó este fenómeno: las familias que tienen ingresos mayores en general tienen casas mayores que las familias de menores ingresos, cuando hay restricciones físicas como ésta, habrá multicolinealidad **independientemente** del método de muestreo que se emplee. Con frecuencia se presentan restricciones en problemas donde intervienen procesos de producción o químicos, cuando los regresores son los componentes de un producto y ésos suman una constante.

También se puede inducir la multicolinealidad por la **elección del modelo**. Por ejemplo, en el capítulo 7 se vio que al agregar términos polinomiales a un modelo de regresión se produce un deterioramiento en $\mathbf{X}'\mathbf{X}$, además, si el rango de x es pequeño, al agregar un término en x^2 puede producirse una multicolinealidad importante. Con frecuencia se encuentran casos como ésos, cuando dos o más regresores tienen dependencia casi lineal, y el retener esos regresores puede contribuir a la multicolinealidad, en esos casos suele ser preferible algún subconjunto de regresores, desde el punto de vista de la multicolinealidad.

Un **modelo sobredefinido** tiene más variables regresoras que observaciones. A veces se encuentran esos modelos en la investigación médica y conductual, cuando puede que sólo haya una pequeña cantidad de personas (unidades de muestra) disponibles, y se reúne información de una gran cantidad de regresores en cada persona. El método común para manejar la multicolinealidad en este contexto es eliminar algunas de las variables regresoras. Mason, Gunst y Webster [1975] presentan tres recomendaciones específicas: 1) redefinir el modelo en términos de un conjunto menor de regresores, 2) hacer estudios preliminares usando sólo subconjuntos de los regresores originales y 3) usar métodos de regresión del tipo de componentes principales, para decir cuáles regresores se van a quitar al modelo. Los dos primeros métodos no tienen en cuenta las interrelaciones entre los regresores y, por consiguiente,

http://carlos2524.jimdo.com/

pueden conducir a resultados no satisfactorios. La regresión por componentes principales se describirá en la sección 10.5.5, aunque no en el contexto de modelos sobredefinidos.

10.3 EFECTOS DE LA MULTICOLINEALIDAD

La presencia de multicolinealidad tiene una gran cantidad de efectos graves sobre los estimados de coeficientes de regresión por mínimos cuadrados. Supóngase que sólo hay dos variables regresoras, x_1 y x_2. El modelo, suponiendo que se escalan x_1, x_2 y y a longitud unitaria, es

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

y las ecuaciones normales de mínimos cuadrados son

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

en donde r_{12} es la correlación simple entre x_1 y x_2, y r_{jy} es la correlación simple entre x_j y y, $j = 1, 2$. Ahora bien, la inversa de $(\mathbf{X}'\mathbf{X})$ es

$$\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \dfrac{1}{(1 - r_{12}^2)} & \dfrac{-r_{12}}{(1 - r_{12}^2)} \\ \dfrac{-r_{12}}{(1 - r_{12}^2)} & \dfrac{1}{(1 - r_{12}^2)} \end{bmatrix} \tag{10.3}$$

y los estimados de los coeficientes de regresión son

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12}r_{2y}}{(1 - r_{12}^2)}, \qquad \hat{\beta}_2 = \frac{r_{2y} - r_{12}r_{1y}}{(1 - r_{12}^2)} \tag{10.4}$$

Si hay fuerte multicolinealidad entre x_1 y x_2, el coeficiente de correlación r_{12} será grande. De acuerdo con la ecuación (10.3) se ve que cuando $|r_{12}| \to 1$, $\mathrm{Var}(\hat{\beta}_j) = C_{jj}\sigma^2 \to \infty$, y $\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) = C_{12}\sigma^2 \to \pm\infty$, dependiendo de si $r_{12} \to +1$ o $r_{12} \to -1$, por consiguiente, la fuerte multicolinealidad entre x_1 y x_2 da como resultado **grandes varianzas y covarianzas** de los estimadores de coeficientes de regresión por mínimos cuadrados.[3] Esto implica que distintas muestras tomadas con los mismos valores de x podrían ocasionar estimaciones muy diferentes de los parámetros del modelo.

Cuando hay más de dos variables regresoras, la multicolinealidad produce efectos parecidos. Se puede demostrar que los elementos diagonales de la matriz $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$ son

[3] La multicolinealidad no es la única causa de grandes varianzas y covarianzas de los coeficientes de regresión.

http://carlos2524.jimdo.com/

$$C_{jj} = \frac{1}{1 - R_j^2}, \quad j = 1, 2, \ldots, p \tag{10.5}$$

en donde R_j^2 es el coeficiente de determinación múltiple de la regresión de x_j respecto a las demás $p - 1$ variables regresoras. Si hay fuerte multicolinealidad entre x_j y cualquier subconjunto de los demás $p - 1$ regresores, el valor de R_j^2 será cercano a la unidad. Como la varianza de $\hat{\beta}_j$ es $\text{Var}(\hat{\beta}_j) = C_{jj}\sigma^2 = (1 - R_j^2)^{-1}\sigma^2$, una fuerte multicolinealidad implica que la varianza del estimado del coeficiente de regresión β_j por mínimos cuadrados es muy grande, por lo general, la covarianza de $\hat{\beta}_i$ y $\hat{\beta}_j$ también será grande, si los regresores x_i y x_j intervienen en una relación multicolineal.

La multicolinealidad tiende también a producir estimados de $\hat{\beta}_j$ que son **demasiado grandes** en valor absoluto; para visualizar eso, se examina la distancia de $\hat{\beta}$ al vector β del parámetro real, elevada al cuadrado, por ejemplo

$$L_1^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta) \tag{10.6}$$

La distancia esperada, elevada al cuadrado, $E(L_1^2)$ es

$$\begin{aligned} E(L_1^2) &= E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) \\ &= \sum_{j=1}^{p} E(\hat{\beta}_j - \beta_j)^2 \\ &= \sum_{j=1}^{p} \text{Var}(\hat{\beta}_j) \\ &= \sigma^2 \text{Tr}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \tag{10.7}$$

en donde la traza de la matriz (que se abrevia Tr) es justo la suma de los elementos de la diagonal principal. Cuando hay multicolinealidad, algunos de los eigenvalores de $\mathbf{X}'\mathbf{X}$ serán pequeños; como la traza de una matriz también es igual a la suma de sus eigenvalores, la ecuación (10.7) se transforma en

$$E(L_1^2) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \tag{10.8}$$

en donde $\lambda_j > 0$, $j = 1, 2, \ldots, p$ son los eigenvalores de $\mathbf{X}'\mathbf{X}$. Así, si la matriz $\mathbf{X}'\mathbf{X}$ está mal acondicionada por la multicolinealidad, al menos una de las λ_j será pequeña, y la ecuación (10.8) implica que la distancia del estimado de $\hat{\beta}$ por mínimos cuadrados a los parámetros β puede ser grande. En forma equivalente se puede demostrar que

$$\begin{aligned} E(L_1^2) &= E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) \\ &= E(\hat{\beta}'\hat{\beta} - 2\hat{\beta}'\beta + \beta'\beta) \end{aligned}$$

o sea

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \text{Tr}(\mathbf{X}'\mathbf{X})^{-1} \tag{10.9}$$

http://carlos2524.jimdo.com/

TABLA 10.1 Datos de acetileno para el ejemplo 10.1

Observación	Conversión de n-heptano en acetileno (%)	Temperatura del reactor (°C)	Relación (molar) de H_2 a n-heptano	Tiempo de contacto (seg)
1	49.0	1300	7.5	0.0120
2	50.2	1300	9.0	0.0120
3	50.5	1300	11.0	0.0115
4	48.5	1300	13.5	0.0130
5	47.5	1300	17.0	0.0135
6	44.5	1300	23.0	0.0120
7	28.0	1200	5.3	0.0400
8	31.5	1200	7.5	0.0380
9	34.5	1200	11.0	0.0320
10	35.0	1200	13.5	0.0260
11	38.0	1200	17.0	0.0340
12	38.5	1200	23.0	0.0410
13	15.0	1100	5.3	0.0840
14	17.0	1100	7.5	0.0980
15	20.5	1100	11.0	0.0920
16	29.5	1100	17.0	0.0860

En general, el vector $\hat{\beta}$ es más largo que el vector β, esto implica que el método de los mínimos cuadrados produce coeficientes de regresión estimados que son demasiado grandes en valor absoluto.

Si bien el método de los mínimos cuadrados producirá en general malos estimados de los parámetros individuales del modelo cuando hay una fuerte multicolinealidad, eso no necesariamente implica que el modelo ajustado sea mal predictor. Si las predicciones se confinan a regiones del espacio de x donde la multicolinealidad es aproximadamente válida, con frecuencia el modelo ajustado produce predicciones satisfactorias, lo que puede suceder porque se logra estimar bastante bien la combinación lineal $\Sigma_{j=1}^{p} \beta_j x_{ij}$, aun cuando los parámetros individuales β_j se estimen mal. Indica con esto que, si los datos originales están aproximadamente sobre el hiperplano definido por la ecuación (10.2), las observaciones futuras que también estén cerca de este hiperplano se podrán predecir con precisión, muy a menudo, a pesar de los estimados inadecuados de los parámetros individuales del modelo.

Ejemplo 10.1 *Datos sobre acetileno*

La tabla 10.1 presenta datos sobre la conversión de n-heptano en acetileno, y tres variables explicativas (Himmelblau [1970], Kunugi, Tamura y Naito [1961] y Marquardt y Snee [1975]). Son datos característicos de un proceso químico, para los que se suele considerar que una superficie totalmente cuadrática de respuesta en los tres regresores es un modelo tentativo adecuado. En la figura 10.2 se ve una gráfica del tiempo de contacto en función de la temperatura del reactor; como esos dos regresores están muy correlacionados, hay un problema potencial de multicolinealidad en estos datos.

El modelo cuadrático completo para los datos del acetileno es

$$P = \gamma_0 + \gamma_1 T + \gamma_2 H + \gamma_3 C + \gamma_{12} TH + \gamma_{13} TC + \gamma_{23} HC + \gamma_{11} T^2 + \gamma_{22} H^2 + \gamma_{33} C^2 + \varepsilon$$

http://carlos2524.jimdo.com/

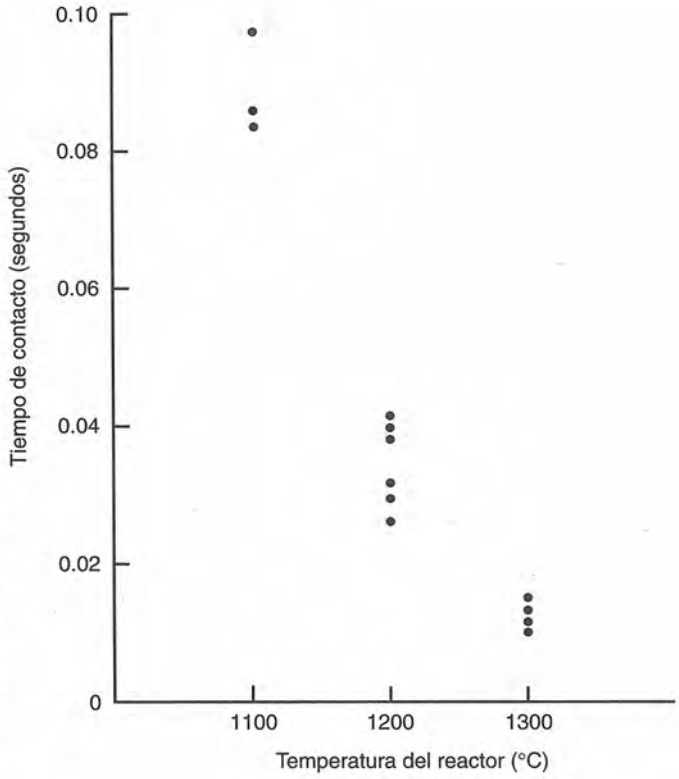


Figura 10.2 Tiempo de contacto en función de la temperatura del reactor, datos del acetileno. (Tomado de Marquardt y Snee (1975) con autorización del editor.)

En donde

$$P = \text{porcentaje de conversión}$$

$$T = \frac{\text{temperatura} - 1\,212.50}{80.623}$$

$$H = \frac{H_2\,(n\text{- heptano}) - 12.44}{5.662}$$

y

$$C = \frac{\text{tiempo de contacto} - 0.0403}{0.03164}$$

Cada uno de los regresores originales se escaló con el escalamiento normal unitario de la sección 3.7 [restando el promedio (es decir, centrándolo) y dividiéndolo entre la desviación estándar]. Los términos cuadráticos y los de producto cruzados generan a partir de los términos lineales escalados. Como se dijo en el capítulo 7, es conveniente centrar los términos lineales para eliminar el deterioramiento no esencial cuando se ajusten los polinomios. El ajuste por mínimos cuadrados es

$$\hat{P} = 35.897 + 4.019T + 2.781H - 8.031C - 6.457TH - 26.982TC$$
$$- 3.768HC - 12.524T^2 - 0.973H^2 - 11.594C^2$$

http://carlos2524.jimdo.com/

TABLA 10.2 Estadísticos de resumen para el modelo del acetileno, por mínimos cuadrados

Término	Coeficiente de regresión	Error estándar	t_0	Coeficiente de regresión estandarizado
Ordenada al origen	35.8971	1.0903	32.93	
T	4.0187	4.5012	0.89	0.3377
H	2.7811	0.3074	9.05	0.2337
C	– 8.0311	6.0657	– 1.32	– 0.6749
TH	– 6.4568	1.4660	– 4.40	– 0.4799
TC	– 26.9818	21.0224	– 1.28	– 2.0344
HC	– 3.7683	1.6554	– 2.28	– 0.2657
T^2	– 12.5237	12.3239	– 1.02	– 0.8346
H^2	– 0.9721	0.3746	– 2.60	– 0.0904
C^2	– 11.5943	7.7070	– 1.50	– 1.0015

$MS_{Res} = 0.8126$, $R^2 = 0.998$, $F_0 = 289.72$
Cuando la respuesta está estandarizada, $MS_{Res} = 0.00038$ para el modelo de mínimos cuadrados.

Las estadísticas de resumen para este modelo se muestran en la tabla 10.2. Los coeficientes de regresión aparecen en términos tanto de los regresores originales centrados, como los regresores normalizados.

Los valores ajustados de los seis puntos (A, B, E, F, I y J) que definen el contorno del casco de contacto de la variable regresora de tiempo de contacto y temperatura del reactor se ven en la figura 10.3, junto con los valores correspondientes observados del porcentaje

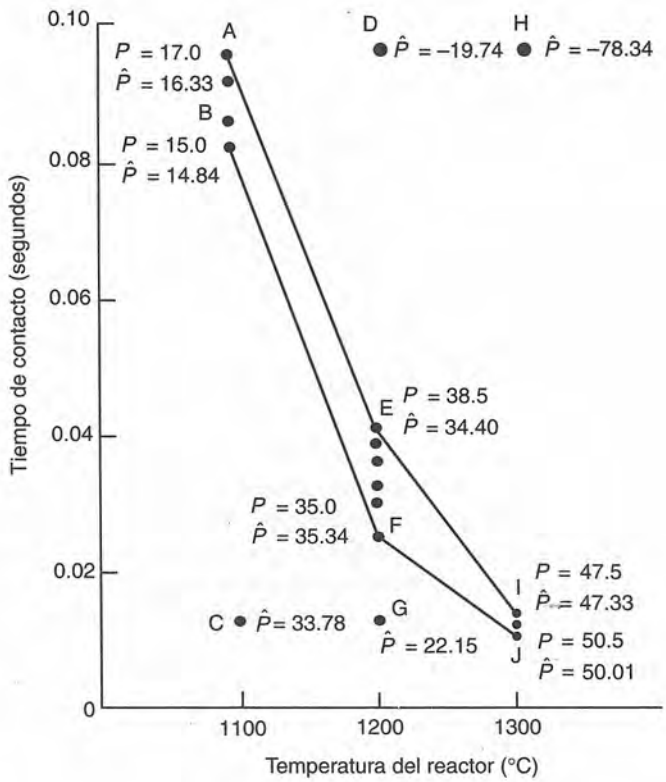


Figura 10.3 Predicciones de conversión porcentual dentro del recorrido de los datos, y extrapolación, para el modelo del acetileno por mínimos cuadrados. (Adaptado de Marquardt y Snee [1975] con autorización del editor.)

http://carlos2524.jimdo.com/

de conversión. Los valores predichos y observados concuerdan mucho y, en consecuencia, el modelo parece adecuado para interpolación dentro del intervalo de los datos originales; ahora se examinará el uso del modelo para extrapolación. También en la figura 10.3 (puntos C, D, G y H) se ven predicciones hechas en las esquinas de la región definida por los márgenes de los datos originales; esos puntos representan extrapolaciones relativamente suaves, porque los intervalos originales de los regresores no se han rebasado. Las conversiones predichas en tres de los cuatro puntos de extrapolación son negativas, lo que es una imposibilidad obvia. Parece que el modelo de mínimos cuadrados se ajusta a los datos en forma razonablemente buena, pero que es muy malo para extrapolaciones; una causa probable de eso, en vista de la fuerte correlación aparente entre el tiempo de contacto y la temperatura del reactor, es la multicolinealidad. En general, si un modelo debe extrapolar bien, se requieren buenos estimados de los coeficientes individuales. Cuando se sospecha que hay multicolinealidad, los estimados de los coeficientes de regresión, por mínimos cuadrados, pueden ser muy malos; eso puede limitar mucho la utilidad del modelo de regresión para hacer inferencias y predicciones.

10.4 DIAGNÓSTICO DE MULTICOLINEALIDAD

Se han propuesto varias técnicas para **detectar la multicolinealidad**. A continuación se describirán e ilustrarán algunas de esas medidas de diagnóstico. Las características deseables en un método de diagnóstico son que refleje el grado del problema de multicolinealidad, y que proporcione información de utilidad para determinar qué regresores están implicados.

10.4.1 Examen de la matriz de correlación

Una medida muy sencilla de la multicolinealidad es la inspección de los elementos r_{ij} no diagonales en $\mathbf{X}'\mathbf{X}$. Si los regresores x_i y x_j son casi linealmente dependientes $|r_{ij}|$ será próximo a la unidad, para ilustrar este procedimiento, considérense los datos del acetileno del ejemplo 10.1. La tabla 10.3 muestra las nueve variables regresoras y la respuesta, en forma normalizada en la que cada una de las variables se ha centrado, restando el promedio de la variable y dividiendo entre la raíz cuadrada de la suma corregida de cuadrados para esa variable. La matriz $\mathbf{X}'\mathbf{X}$ en la forma de correlación, para los datos del acetileno es

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1.000 & 0.224 & -0.958 & -0.132 & 0.443 & 0.205 & -0.271 & 0.031 & -0.577 \\ & 1.000 & -0.240 & 0.039 & 0.192 & -0.023 & -0.148 & 0.498 & -0.224 \\ & & 1.000 & 0.194 & -0.661 & -0.274 & 0.501 & -0.018 & 0.765 \\ & & & 1.000 & -0.265 & -0.975 & 0.246 & 0.398 & 0.274 \\ & & & & 1.000 & 0.323 & -0.972 & 0.126 & -0.972 \\ & & & & & 1.000 & -0.279 & -0.374 & 0.358 \\ & & & & & & 1.000 & -0.124 & 0.874 \\ & & & & & & & 1.000 & -0.158 \\ & & & & & & & & 1.000 \\ \text{simétrica} & & & & & & & & \end{bmatrix}$$

La matriz $\mathbf{X}'\mathbf{X}$ deja ver la gran correlación entre la temperatura del reactor (x_1), y el tiempo de contacto (x_3), que se sospechaba antes, al inspeccionar la figura 10.2, porque $r_{13} = -0.958$,

http://carlos2524.jimdo.com/

TABLA 10.3 Datos estandarizados del acetileno[a]

Observación i	y	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2
1	.27979	.28022	−.22554	−.23106	−.33766	−.02085	.30952	.07829	−.04116	−.03452
2	.30583	.28022	−.15704	−.23106	−.25371	−.02085	.23659	.07829	−.13270	−.03452
3	.31234	.28022	−.06584	−.23514	−.14179	−.02579	.14058	.07829	−.20378	−.02735
4	.26894	.28022	.04817	−.22290	.00189	−.01098	.01960	.07829	−.21070	−.04847
5	.24724	.28022	.20777	−.21882	.19398	−.00605	−.14065	.07829	−.06745	−.05526
6	.18214	−.04003	.48139	−.23106	.52974	−.02085	−.44415	.07829	.59324	−.03452
7	.17590	−.04003	−.32577	−.00255	−.00413	.25895	.07300	−.29746	.15239	−.23548
8	−.09995	−.04003	−.22544	−.01887	−.02171	.26177	.08884	−.29746	−.04116	−.23418
9	−.03486	−.04003	−.06584	−.06784	−.04970	.27023	.08985	−.29746	−.20378	−.21822
10	−.02401	−.04003	.04817	−.11680	−.06968	.27869	.04328	−.29746	−.21070	−.18419
11	.04109	−.04003	.20777	−.05152	−.09766	.26741	.01996	−.29746	−.06745	−.22554
12	.05194	−.04003	.48139	.00561	−.14563	.25754	.08202	−.29746	.59329	−.23538
13	.45800	−.36029	−.32577	.35653	.45252	−.29615	−.46678	.32879	.15239	.24374
14	.41460	−.36029	−.22544	.47078	.29423	−.47384	−.42042	.32879	−.04116	.60000
15	−.33865	−.36029	−.06584	.42187	.04240	−.39769	−.05859	.32879	−.20378	.43527
16	−.14335	−.36029	.20777	.37285	−.38930	−.32153	−.42738	.32879	−.06745	.28861

[a] Los datos estandarizados se obtuvieron a partir de la forma centrada y escalada de los datos originales de la tabla 10.1.

http://carlos2524.jimdo.com/

TABLA 10.4 Regresor y variables de respuesta no estandarizados, de Webster, Gunst y Mason [1974]

Observación i	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}
1	10.006	8.000	1.000	1.000	1.000	0.541	−0.099
2	9.737	8.000	1.000	1.000	0.000	0.130	0.070
3	15.087	8.000	1.000	1.000	0.000	2.116	0.115
4	8.422	0.000	0.000	9.000	1.000	−2.397	0.252
5	8.625	0.000	0.000	9.000	1.000	−0.046	0.017
6	16.289	0.000	0.000	9.000	1.000	0.365	1.504
7	5.958	2.000	7.000	0.000	1.000	1.996	−0.865
8	9.313	2.000	7.000	0.000	1.000	0.228	−0.055
9	12.960	2.000	7.000	0.000	1.000	1.380	0.502
10	5.541	0.000	0.000	0.000	10.000	−0.798	−0.399
11	8.756	0.000	0.000	0.000	10.000	0.257	0.101
12	10.937	0.000	0.000	0.000	10.000	0.440	0.432

además hay otros coeficientes de correlación grandes, entre x_1x_2 y x_2x_3, x_1x_3 y x_1^2 y entre x_1^2 y x_3^2. Eso no debe sorprender, porque esas variables se generan a partir de los términos lineales, e intervienen en ellas los regresores x_1 y x_3, muy correlacionados entre sí, así, por inspección de la matriz de correlación se ve que hay varias dependencias casi lineales en los datos del acetileno.

Es útil examinar las correlaciones simples r_{ij} entre los regresores, para detectar la dependencia casi lineal, sólo entre **pares de regresores**, desafortunadamente, cuando intervienen más de dos regresores en una dependencia casi lineal, no hay la seguridad de que alguna de las correlaciones apareadas r_{ij} sea grande, por ejemplo, considérense los datos de la tabla 10.4 que fueron generados por Webster, Gunst y Mason [1974], en forma artificial y que establecieron que $\Sigma_{j=1}^{4} x_{ij} = 10$ para las observaciones 2 a 12, mientras que $\Sigma_{j=1}^{4} x_{ij} = 11$ para la observación 1. Los regresores 5 y 6 se obtuvieron de una tabla de números aleatorios normales. Las respuestas y_i fueron generadas con la relación

$$y_i = 10 + 2.0x_{i1} + 1.0x_{i2} + 0.2x_{i3} - 2.0x_{i4} + 3.0x_{i5} + 10.0x_{i6} + \varepsilon_i$$

en donde $\varepsilon_i \sim N(0, 1)$. La matriz $\mathbf{X}'\mathbf{X}$ en la correlación con esos datos es

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1.000 & 0.052 & -0.343 & -0.498 & 0.417 & -0.192 \\ & 1.000 & -0.432 & -0.371 & 0.485 & -0.317 \\ & & 1.000 & -0.355 & -0.505 & 0.494 \\ & & & 1.000 & -0.215 & -0.087 \\ & & & & 1.000 & -0.123 \\ \text{Simétrica} & & & & & 1.000 \end{bmatrix}$$

Ninguna de las correlaciones apareadas r_{ij} es sospechosamente grande, y en consecuencia no hay indicios de dependencia casi lineal entre los regresores, en general, la inspección de los r_{ij} no es suficiente para detectar cosas más complejas que la multicolinealidad por pares.

http://carlos2524.jimdo.com/

10.4.2 Factores de inflación de varianza

Se hizo notar en el capítulo 3 que los elementos diagonales de la matriz $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$ son muy útiles para detectar la multicolinealidad. Recuérdese, de la ecuación (10.5), que C_{jj}, el j-ésimo elemento diagonal de $\mathbf{C}$ se puede escribir en la forma $C_{jj} = (1 - R_j^2)^{-1}$, siendo R_j^2 el coeficiente de determinación obtenido cuando se hace la regresión de x_j respecto a los demás $p - 1$ regresores. Si x_j es casi ortogonal a los regresores restantes, R_j^2 es pequeño y C_{jj} es cercano a la unidad, mientras que si x_j es casi linealmente dependiente en algún subconjunto de los regresores restantes, R_j^2 es casi lineal y C_{jj} es grande. Como la varianza de los j-ésimos coeficientes de regresión es $C_{jj}\sigma^2$ se puede considerar que C_{jj} es el factor en el que aumenta la varianza de $\hat{\beta}_j$ debido a dependencias casi lineales entre los regresores. En el Capítulo 3 se definió

$$\text{VIF}_j = C_{jj} = (1 - R_j^2)^{-1}$$

como **factor de inflación de varianza**, esta terminología se debe a Marquardt [1970]. El factor VIF (de *variance inflation factor*) para cada término del modelo mide el efecto combinado que tienen las dependencias entre los regresores sobre la varianza de ese término. Si hay uno o más VIF grandes, hay multicolinealidad. La experiencia indica que si cualquiera de los VIF es mayor que 5 o 10, es indicio de que los coeficientes asociados de regresión están mal estimados debido a la multicolinealidad.

Los VIF tienen otra interpretación interesante. La longitud del intervalo de confianza en la teoría normal, para el j-ésimo coeficiente de regresión, se puede escribir en la forma

$$L_j = 2(C_{jj}\ \hat{\sigma}^2)^{1/2} t_{\alpha/2,\, n-p-1}$$

y la longitud del intervalo correspondiente, basada en un **diseño de referencia ortogonal,** con el mismo tamaño de muestra y valores cuadráticos medios (rms) [es decir, rms = $\Sigma_{i=1}^{n}(x_{ij} - \bar{x}_j)^2/n$ es una medida de la dispersión del regresor x_j] que en el diseño original, es

$$L^* = 2\ \hat{\sigma} t_{\alpha/2,\, n-p-1}$$

La relación entre estos dos intervalos de confianza es $L_j/L^* = C_{jj}^{1/2}$, así, la raíz cuadrada del j-ésimo VIF indica cuánto más largo es el intervalo de confianza del j-ésimo coeficiente de regresión, a causa de la multicolinealidad.

En la parte A de la tabla 10.5 se ven los VIF para los datos del acetileno, esos VIF son los elementos de la diagonal principal de $(\mathbf{X}'\mathbf{X})^{-1}$, suponiendo que los términos lineales en el modelo están centrados, y que los términos de segundo orden se generan en forma directa a partir de los términos lineales. El VIF máximo es 6 565.91, por lo que se llega a la conclusión que existe un problema de multicolinealidad, además, los VIF de varias de las otras variables de producto cruzado y cuadráticas, donde intervienen x_1 y x_3, son grandes. Por consiguiente, los VIF pueden ayudar a identificar cuáles regresores intervienen en la multicolinealidad, y también, que los VIF en los modelos polinomiales se afectan al centrar los términos lineales. La parte B de la tabla 10.5 muestra los VIF de los datos del acetileno, suponiendo que los términos lineales no están centrados; esos VIF son mucho mayores que los de los datos centrados, así, al centrar los términos lineales en un modelo polinomial se elimina algo del deterioramiento no esencial, causado por la elección del origen de los regresores.

Los VIF de los datos de Webster, Gunst y Mason, se ven en la parte C de la tabla 10.5; como el VIF máximo es 297.14, es claro que hay multicolinealidad, también esta

http://carlos2524.jimdo.com/

TABLA 10.5 Factores de inflación de varianza (VIF) para los datos del acetileno y los datos de Webster Gunst y Mason

Datos del acetileno, centrados (A)	Datos del acetileno, sin centrar (B)	Datos de Webster, Gunst y Mason (C)
VIF del término	VIF del término	VIF del término
$x_1 = 374$	$x_1 = 2{,}856{,}749$	$x_1 = 181.83$
$x_2 = 1.74$	$x_2 = 10{,}956.1$	$x_2 = 161.40$
$x_3 = 679.11$	$x_3 = 2{,}017{,}163$	$x_3 = 265.49$
$x_1x_2 = 31.03$	$x_1x_2 = 2{,}501{,}945$	$x_4 = 297.14$
$x_1x_3 = 6565.91$	$x_1x_3 = 65.73$	$x_5 = 1.74$
$x_2x_3 = 35.60$	$x_2x_3 = 12{,}667.1$	$x_6 = 1.44$
$x_1^2 = 1762.58$	$x_1^2 = 9802.9$	
$x_2^2 = 3.17$	$x_2^2 = 1{,}428{,}092$	
$x_3^2 = 1158.13$	$x_3^2 = 240.36$	
VIF máximo = 6565.91	VIF máximo = 2,856,749	VIF máximo = 297.14

vez, nótese que los VIF correspondientes a los regresores que intervienen en la multicolinealidad son mucho mayores que para x_5 y x_6.

10.4.3 Análisis del eigensistema de X′X

Las raíces características, o **eigenvalores** o valores propios de $\mathbf{X'X}$, por ejemplo $\lambda_1, \lambda_2, \ldots, \lambda_p$ se pueden usar para medir el grado de multicolinealidad en los datos.[4] Si hay una o más dependencias casi lineales en los datos, una o más de las raíces características será pequeña. Uno o más eigenvalores pequeños implican que hay dependencias casi lineales entre las columnas de $\mathbf{X}$. Algunos analistas prefieren examinar el **número de condición** de $\mathbf{X'X}$, que se define como

$$\kappa = \frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}} \tag{10.10}$$

Esto no es más que una medida de la dispersión en el espectro de eigenvalores de $\mathbf{X'X}$. En general, si el número de condición es menor que 100, no hay problema grave de multicolinealidad. Los números de condición de 100 a 1 000 implican multicolinealidad de moderada a fuerte, y si κ es mayor que 1 000, es indicio de una fuerte multicolinealidad.

Los **índices de condición** de la matriz $\mathbf{X'X}$ son

$$\kappa_j = \frac{\lambda_{\text{máx}}}{\lambda_j} \qquad j = 1, 2, \ldots, p$$

Es claro que el máximo índice de condición es el número de condición definido en la ecuación (10.10). La cantidad de índices de condición que son grandes (digamos, $\geq 1\,000$) es una medida útil de la cantidad de dependencias casi lineales en $\mathbf{X'X}$.

[4] Recuérdese que los eigenvalores de una matriz $\mathbf{A}$ de $p \times p$ son las p raíces de la ecuación $|\mathbf{A} - \lambda\mathbf{I}| = 0$. Los eigenvalores se calculan casi siempre con rutinas de cómputo. Los métodos para calcular eigenvalores y eigenvectores se describen en Smith *et al.* [1974], Stewart [1973] y Wilkinson [1965].

http://carlos2524.jimdo.com/

Los eigenvalores de $\mathbf{X}'\mathbf{X}$ para los datos del acetileno son $\lambda_1 = 4.2048$, $\lambda_2 = 2.1626$, $\lambda_3 = 1.1384$, $\lambda_4 = 1.0413$, $\lambda_5 = 0.3845$, $\lambda_6 = 0.0495$, $\lambda_7 = 0.0136$, $\lambda_8 = 0.0051$ y $\lambda_9 = 0.0001$. Hay cuatro eigenvalores muy pequeños, lo cual es síntoma de datos deteriorados en grado extremo. El número de condición es

$$\kappa = \frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}} = \frac{4.2048}{0.0001} = 42\,048$$

que indica una multicolinealidad grave. Los índices de condición son

$$\kappa_1 = \frac{4.2048}{4.2048} = 1, \quad \kappa_2 = \frac{4.2048}{2.1626} = 1.94, \quad \kappa_3 = \frac{4.2048}{1.1384} = 3.69,$$

$$\kappa_4 = \frac{4.2048}{1.0413} = 4.04, \quad \kappa_5 = \frac{4.2048}{0.3845} = 10.94, \quad \kappa_6 = \frac{4.2048}{0.0495} = 84.96,$$

$$\kappa_7 = \frac{4.2048}{0.0136} = 309.18, \quad \kappa_8 = \frac{4.2048}{0.0051} = 824.47, \quad \kappa_9 = \frac{4.2048}{0.0001} = 42\,048$$

Como uno de los índices de condición es mayor que 1 000 (y dos más son mayores que 100), se llega a la conclusión de que al menos hay una fuerte dependencia casi lineal en los datos del acetileno. En vista de que x_1 está muy correlacionado con x_3, y que el modelo contiene términos cuadráticos y de producto cruzado en x_1 y x_3, no debe sorprender esa conclusión.

Los eigenvalores para los datos de Webster, Gunst y Mason son $\lambda_1 = 2.4288$, $\lambda_2 = 1.5462$, $\lambda_3 = 0.9221$, $\lambda_4 = 0.7940$, $\lambda_5 = 0.3079$ y $\lambda_6 = 0.0011$. El eigenvalor pequeño indica que hay dependencia casi lineal en los datos. El número de condición es

$$\kappa = \frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}} = \frac{2.4288}{0.0011} = 2\,188.11$$

que también indica una fuerte multicolinealidad. Sólo hay un índice de condición mayor que 1 000, por lo que se llega a la conclusión que sólo hay una dependencia casi lineal en los datos.

También se puede aplicar el **análisis del eigensistema** para identificar la naturaleza de las dependencias casi lineales en los datos. La matriz $\mathbf{X}'\mathbf{X}$ se puede descomponer como sigue:

$$\mathbf{X}'\mathbf{X} = \mathbf{T}\Lambda\mathbf{T}'$$

donde Λ es una matriz diagonal de $p \times p$, cuyos elementos de diagonal principal son los **eigenvalores** λ_j $(j = 1, 2, \ldots, p)$ de $\mathbf{X}'\mathbf{X}$ y $\mathbf{T}$ es una matriz ortogonal de $p \times p$, cuyas columnas son los eigenvectores de $\mathbf{X}'\mathbf{X}$. Sean $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_p$ las columnas de $\mathbf{T}$. Si el eigenvalor λ_j es cercano a cero, indicando una dependencia casi lineal en los datos, los elementos del eigenvector $\mathbf{t}_j$ describen la naturaleza de esa dependencia lineal. En forma específica, los elementos del vector $\mathbf{t}_j$ son los coeficientes $t_1, t_2, \ldots, t_p$ en la ecuación (10.2).

La tabla 10.6 muestra los eigenvectores para los datos de Webster, Gunst y Mason. El eigenvalor más pequeño es $\lambda_6 = 0.0011$, por lo que los elementos del eigenvector $\mathbf{t}_6$ son los coeficientes de los regresores en la ecuación (10.2). Esto implica que

$$-0.44768\,x_1 - 0.42114\,x_2 - 0.54169\,x_3 - 0.57337\,x_4$$
$$-0.00605\,x_5 - 0.00217\,x_6 = 0$$

http://carlos2524.jimdo.com/

TABLA 10.6 Eigenvectores para los datos de Webster, Gunst y Mason

$\mathbf{t}_1$	$\mathbf{t}_2$	$\mathbf{t}_3$	$\mathbf{t}_4$	$\mathbf{t}_5$	$\mathbf{t}_6$
−.39072	−.33968	.67980	.07990	−.25104	−.44768
−.45560	−.05392	−.70013	.05769	−.34447	−.42114
.48264	−.45333	−.16078	.19103	.45364	−.54169
.18766	.73547	.13587	−.27645	.01521	−.57337
−.49773	−.09714	−.03185	−.56356	.65128	−.00605
.35195	−.35476	−.04864	−.74818	−.43375	−.00217

do términos, se obtiene

$$x_1 \simeq -0.941x_2 - 1.210x_3 - 1.281x_4$$

Esto es, los cuatro primeros regresores aproximadamente suman igual a una constante, así, los elementos de $\mathbf{t}_6$ reflejan en forma directa la relación que se usó para generar x_1, x_2, x_3 y x_4.

Belsley, Kuh y Welsch [1980] proponen un método parecido para diagnosticar la multicolinealidad. La matriz $\mathbf{X}$ de $n \times p$ se puede descomponer como sigue:

$$\mathbf{X} = \mathbf{UDT}'$$

en donde $\mathbf{U}$ es de $n \times p$, $\mathbf{T}$ es de $p \times p$, $\mathbf{U}'\mathbf{U} = \mathbf{I}$, $\mathbf{T}'\mathbf{T} = \mathbf{I}$ y $\mathbf{D}$ es una matriz diagonal de $p \times p$, con elementos μ_j diagonales, $j = 1, 2, \ldots, p$, no negativos. Las μ_j se llaman **valores singulares** de $\mathbf{X}$, y $\mathbf{X} = \mathbf{UDT}'$ se llama la **descomposición de X en valores singulares**. La descomposición en valores singulares se relaciona mucho con los conceptos de eigenvalores y eigenvectores, porque $\mathbf{X}'\mathbf{X} = (\mathbf{UDT}')\mathbf{UDT}' = \mathbf{TD}^2\mathbf{T}' = \mathbf{T}\Lambda\mathbf{T}'$, por lo que los cuadrados de los valores singulares de $\mathbf{X}$ son los eigenvalores de $\mathbf{X}'\mathbf{X}$; en este caso, $\mathbf{T}$ es la matriz de eigenvectores de $\mathbf{X}'\mathbf{X}$ definida antes, y $\mathbf{U}$ es una matriz cuyas columnas son los eigenvectores asociados con los p eigenvalores no cero de $\mathbf{XX}'$.

El deterioramiento de $\mathbf{X}$ se refleja en el tamaño de los valores singulares. Habrá un valor singular pequeño por cada dependencia casi lineal. El grado de deterioramiento depende de lo pequeño que sea el valor singular, en relación con el máximo valor singular, $\mu_{\text{máx}}$. Belsley, Kuh y Welsch [1980] definen los **índices de condición** de la matriz $\mathbf{X}$ como sigue:

$$\eta_j = \frac{\mu_{\text{máx}}}{\mu_j}, \qquad j = 1, 2, \ldots, p$$

El máximo valor de η_j es el número de condición de $\mathbf{X}$, nótese que este método maneja en forma directa la matriz de datos $\mathbf{X}$, que es la que nos ocupa principalmente, y no la matriz de la suma de cuadrados y productos cruzados, $\mathbf{X}'\mathbf{X}$. Una ventaja más de este método es que los algoritmos para generar la descomposición en valores singulares son numéricamente más estables que los del análisis del eigensistema, aunque en la práctica puede que eso no sea una grave limitación, si se prefiere el método del eigensistema.

La matriz de covarianza de $\hat{\beta}$ es

$$\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{T}\Lambda^{-1}\mathbf{T}'$$

http://carlos2524.jimdo.com/

y la varianza del j-ésimo coeficiente de regresión es el j-ésimo elemento diagonal de esta matriz, es decir

$$\operatorname{Var}\left(\hat{\beta}_j\right) = \sigma^2 \sum_{i=1}^{p} \frac{t_{ji}^2}{\mu_i^2} = \sigma^2 \sum_{i=1}^{p} \frac{t_{ji}^2}{\lambda_i}$$

También obsérvese que además de σ^2, el j-ésimo elemento diagonal de $\mathbf{T}\Lambda^{-1}\mathbf{T}'$ es el j-ésimo factor de inflación de varianza, es decir,

$$\mathrm{VIF}_j = \sum_{i=1}^{p} \frac{t_{ji}^2}{\mu_i^2} = \sum_{i=1}^{p} \frac{t_{ji}^2}{\lambda_i}$$

Es claro que uno o más valores singulares pequeños (o eigenvalores pequeños) pueden inflar en forma dramática la varianza de $\hat{\beta}_j$. Belsley, Kuh y Welsch sugieren usar **proporciones de descomposición de varianza**, definidas como sigue:

$$\pi_{ij} = \frac{t_{ji}^2/\mu_i^2}{\mathrm{VIF}_j}, \qquad j = 1, 2, \ldots, p$$

como medidas de multicolinealidad. Si se agrupan las π_{ij} en una matriz π de $p \times p$, los elementos de cada columna de π no son más que las proporciones de la varianza de cada $\hat{\beta}_j$ (o de cada factor de inflación de varianza), debidas al i-ésimo valor singular (o eigenvalor). Si una gran proporción de la varianza de dos o más coeficientes de regresión se relaciona con un valor singular pequeño, quiere decir que hay multicolinealidad, por ejemplo, si π_{32} y π_{34} son grandes, el tercer valor singular está asociado con una multicolinealidad que está inflando las varianzas de $\hat{\beta}_2$ y $\hat{\beta}_4$. Índices de condición mayores que 30, y proporciones de descomposición de varianza mayores que 0.5 son las reglas que se recomiendan utilizar.

La tabla 10.7 muestra los índices (η_j) de condición de $\mathbf{X}$ y las proporciones de descomposición de varianza (π_{ij}) para los datos de Webster, Gunst y Mason. En la parte A de la tabla se han centrado los regresores, para que esas variables sean $(x_{ij} - \bar{x}_j)$, $j = 1, 2, \ldots, 6$. En la sección 10.4.2 se observó que los factores de inflación de varianza VIF, en un modelo polinomial, se afectan al centrar los términos lineales en el modelo, antes de generar los términos de orden mayor del polinomio. El centrado también afecta las proporciones de descomposición de varianza (así como a los eigenvalores y eigenvectores). En esencia, el centrado elimina todo deterioramiento no esencial, debido a la ordenada al origen.

Obsérvese que sólo hay un índice de condición grande ($\eta_6 = 46.86 > 30$), por lo que hay una dependencia en las columnas de $\mathbf{X}$, Además, las proporciones de descomposición de varianza π_{61}, π_{62}, π_{63} y π_{64} son todas mayores que 0.5, lo que indica que los cuatro primeros regresores están implicados en una relación multicolineal. Es, en esencia, la misma información que se obtuvo antes, al examinar los eigenvalores.

Belsley, Kuh y Welsch [1980] sugieren que se deben escalar los regresores a longitud unitaria, pero que no se deben centrar al calcular las proporciones de descomposición de varianza, con objeto de diagnosticar el papel de la ordenada al origen en las dependencias casi lineales. Esta opción se muestra en la parte B de la tabla 10.7, nótese que el efecto de lo anterior es aumentar la dispersión de los eigenvalores y aumentar los índices de condición.

Hay cierta controversia acerca de si se deben centrar los datos de la regresión para diagnosticar la multicolinealidad, sea con el análisis del eigensistema, o con el método de la proporción de descomposición de varianza. El centrado ortogonaliza la ordenada al origen con los demás regresores, por lo que se puede considerar que el centrado es una operación que

http://carlos2524.jimdo.com/

TABLA 10.7 Proporciones de descomposición de varianza para los datos de Webster, Gunst y Mason [1974]

A. Regresores centrados

Número	Eigenvalor	Índices de condición	Proporciones de descomposición de varianza					
			X_1	X_2	X_3	X_4	X_5	X_6
1	2.42879	1.00000	0.0003	0.0005	0.0004	0.0000	0.0531	0.0350
2	1.54615	1.25334	0.0004	0.0000	0.0005	0.0012	0.0032	0.0559
3	0.92208	1.62297	0.0028	0.0033	0.0001	0.0001	0.0006	0.0018
4	0.79398	1.74900	0.0000	0.0000	0.0002	0.0003	0.2083	0.4845
5	0.30789	2.80864	0.0011	0.0024	0.0025	0.0000	0.7175	0.4199
6	0.00111	46.86052	0.9953	0.9937	0.9964	0.9984	0.0172	0.0029

B. Regresores no centrados

Número	Eigenvalor	Índices de condición	Ordenada al origen	Proporciones de descomposición de varianza					
				X_1	X_2	X_3	X_4	X_5	X_6
1	2.63287	1.00000	0.0001	0.0003	0.0003	0.0001	0.0001	0.0217	0.0043
2	1.82065	1.20255	0.0000	0.0001	0.0002	0.0005	0.0000	0.0523	0.0949
3	1.03335	1.59622	0.0000	0.0002	0.0000	0.0002	0.0013	0.0356	0.1010
4	0.65826	1.99994	0.0000	0.0005	0.0000	0.0005	0.0003	0.1906	0.3958
5	0.60573	2.08485	0.0000	0.0025	0.0035	0.0001	0.0001	0.0011	0.0002
6	0.24884	3.25280	0.0000	0.0012	0.0023	0.0028	0.0000	0.6909	0.4003
7	0.00031	92.25341	0.9999	0.9953	0.9936	0.9959	0.9983	0.0178	0.0034

gen no tiene interpretación física (como es el caso en muchas aplicaciones de la regresión en ingeniería y en las ciencias físicas), entonces el deterioramiento causado por el término constante es realmente "no esencial" y por ello es totalmente adecuado centrar los regresores, sin embargo, si la ordenada al origen tiene un valor interpretativo, el centrado no será el mejor método. Es claro que la respuesta a esta pregunta depende del problema; para conocer unas excelentes descripciones de este argumento, véanse Brown [1977] y Myers [1990].

10.4.4 Otros diagnósticos

Hay otras técnicas que a veces se usan para diagnosticar la colinealidad. Se puede usar el **determinante** de $\mathbf{X}'\mathbf{X}$ como índice de colinealidad; ya que la matriz $\mathbf{X}'\mathbf{X}$ está en forma de correlación, el intervalo de posibles valores del determinante es $0 \leq |\mathbf{X}'\mathbf{X}| \leq 1$. Si $|\mathbf{X}'\mathbf{X}| = 1$, los regresores son ortogonales, mientras que si $|\mathbf{X}'\mathbf{X}| = 0$, hay una dependencia lineal exacta entre ellos. El grado de multicolinealidad se agrava a medida que $|\mathbf{X}'\mathbf{X}|$ tiende a cero. Si bien esta medida de multicolinealidad es fácil de aplicar, no proporciona información alguna sobre el origen de la multicolinealidad.

Willam y Watts [1978] sugieren otra interpretación de este diagnóstico. La región de intervalo conjunto de confianza $100(1 - \alpha)$ por ciento de β, basada en los datos observados, es

$$(\beta - \hat{\beta})'\mathbf{X}'\mathbf{X}(\beta - \hat{\beta}) \leq p\ \hat{\sigma}^2 F_{\alpha,\, p,\, n-p-1}$$

http://carlos2524.jimdo.com/

mientras que la región correspondiente de confianza de $\hat{\beta}$ basada en el diseño ortogonal de referencia que se describió antes es

$$(\beta - \hat{\beta})'(\beta - \hat{\beta}) \leq p\ \hat{\sigma}^2 F_{\alpha,\, p,\, n-p-1}$$

El diseño ortogonal de referencia produce la mínima región de confianza conjunta para un tamaño fijo y valores rms fijos, y una α determinada. La relación de los volúmenes de las dos regiones de confianza es $|\mathbf{X}'\mathbf{X}|^{1/2}$, por lo que $|\mathbf{X}'\mathbf{X}|^{1/2}$ mide la pérdida de poder de estimación debida a la multicolinealidad. Planteado en otra forma, $100(|\mathbf{X}'\mathbf{X}|^{1/2} - 1)$ refleja el aumento de porcentaje en el volumen de la región de confianza conjunta de β, debido a las dependencias casi lineales en $\mathbf{X}$, por ejemplo, si $|\mathbf{X}'\mathbf{X}| = 0.25$, el volumen de la región de confianza conjunta es $100[(0.25)^{-1/2} - 1] = 50\%$ mayor que lo que sería si se hubiera usado un diseño ortogonal.

La estadística F para el significado de la regresión y las estadísticas t individuales (o F parciales) a veces pueden indicar la presencia de multicolinealidad. En forma específica, si la estadística F general es significativa, pero las estadísticas t individuales son todas no significativas, habrá multicolinealidad, desafortunadamente, muchos conjuntos de datos tienen multicolinealidad importante, pero no tienen este comportamiento, por lo que es cuestionable la utilidad de esta medida de multicolinealidad.

Los **signos** y las **magnitudes** de los coeficientes de regresión a veces proporcionan una indicación de la presencia de multicolinealidad, en particular, si al agregar o eliminar un regresor se producen grandes cambios en los estimados de los coeficientes de regresión, quiere decir que hay multicolinealidad. Si la eliminación de uno o más puntos de datos da como resultado grandes cambios en los coeficientes de regresión, puede haber multicolinealidad, por último, si los signos o las magnitudes de los coeficientes de regresión en el modelo son contrarios a lo esperado, se debe estar alerta por posible multicolinealidad, por ejemplo, el modelo de mínimos cuadrados para los datos del acetileno tiene coeficientes normalizados de regresión grandes para la interacción x_1x_3, y para los términos cuadráticos x_1^2 y x_3^2. A veces es raro que los modelos cuadráticos tengan grandes coeficientes de regresión para los términos de orden superior, por lo que ello puede ser una indicación de multicolinealidad, sin embargo, se debe tener cuidado al usar los signos y las magnitudes de los coeficientes de regresión como síntomas de multicolinealidad, porque muchos conjuntos de datos muy deteriorados no se comportan en forma que sea obviamente inusual en este respecto.

Creemos que los VIF (factores de inflación de varianza) y los procedimientos basados en los eigenvalores de $\mathbf{X}'\mathbf{X}$ son los mejores diagnósticos con los que se dispone en la actualidad para la multicolinealidad. Son fáciles de calcular, de interpretación directa y son útiles para investigar la naturaleza específica de la multicolinealidad. Para tener más información sobre éstos y otros métodos para detectar la multicolinealidad, véanse Belsley, Kuh y Welsch [1980], Farrar y Glauber [1967] y Willan y Watts [1978].

10.5 MÉTODOS PARA MANEJAR LA MULTICOLINEALIDAD

Se han propuesto varias técnicas para manejar los problemas causados por la multicolinealidad. Entre los métodos generales están el reunir más datos, la reespecificación del modelo y el uso de métodos de estimación distintos de los mínimos cuadrados, diseñados en forma específica para combatir los problemas inducidos por la multicolinealidad.

http://carlos2524.jimdo.com/

10.5.1 Recolección de datos adicionales

Se ha sugerido la recolección de datos adicionales como el mejor método para combatir la multicolinealidad (véase, por ejemplo, Farrar y Glauber [1967], y Silvey [1969]). Los datos adicionales se deben reunir en una forma diseñada para eliminar la multicolinealidad en los datos actuales, por ejemplo, en los datos de tiempo de entrega que se presentaron primero en el ejemplo 3.1. Una gráfica de los regresores cajas (x_1) en función de distancia (x_2) se ve en la matriz de diagramas de dispersión, figura 3.4. Ya se hizo notar antes que la mayor parte de esos datos están a lo largo de una recta desde valores pequeños de cajas y distancia, hasta altos valores de cajas y distancia, y en consecuencia puede haber algún problema con la multicolinealidad. Eso se podría evitar reuniendo algunos datos adicionales en tal forma que se rompa cualquier multicolinealidad potencial, esto es, en puntos donde las cajas son pequeñas y la distancia es grande, y en puntos donde las cajas sean grandes y la distancia sea pequeña.

Desafortunadamente no siempre es posible coleccionar más datos, por **restricciones económicas** o porque el proceso que se estudia **ya no está disponible** para su muestreo. Aun cuando hayan más datos disponibles, puede ser inadecuado usarlos, si amplían el recorrido de las variables regresoras mucho más allá de la región de interés para el analista, además, si los nuevos puntos de datos son raros o atípicos en el proceso que se estudia, su presencia en la muestra podría ser muy influyente sobre el modelo ajustado. Por último, nótese que la reunión de datos adicionales no es una solución viable del problema de multicolinealidad cuando ésta se debe a restricciones sobre el modelo o en la población, por ejemplo, considérense los factores "ingreso familiar" (x_1), y "tamaño de casa" (x_2), graficados en la figura 10.1. La recolección de más datos sería de poco valor en este caso, porque la relación entre el ingreso familiar y el tamaño de la casa es una característica estructural de la población. Casi todos los datos en la población tendrán este comportamiento.

10.5.2 Reespecificación del modelo

Con frecuencia, la multicolinealidad se debe a la elección del modelo, como cuando dos regresores muy correlacionados se usan en la ecuación de regresión, en estos casos la **reespecificación** de la ecuación de regresión puede aminorar el impacto de la multicolinealidad. Un método para reespecificar el modelo es redefinir los regresores, por ejemplo si x_1, x_2 y x_3 son casi linealmente dependientes, podrá determinarse alguna función, como $x = (x_1 + x_2)/x_3$, o $x = x_1x_2x_3$, que preserve el contenido de información de los regresores originales, pero que reduzca el deterioramiento.

Otro método muy usado para reespecificar el modelo es la **eliminación de variable**. Esto es, si x_1, x_2 y x_3 son casi linealmente dependientes, la eliminación de un regresor, por ejemplo x_3, puede ayudar a combatir la multicolinealidad. La eliminación de variable es frecuentemente una técnica muy efectiva, sin embargo, podrá no producir una solución satisfactoria si los regresores eliminados del modelo tienen un gran poder de explicación en relación de la respuesta y, lo que significa que la eliminación de los regresores para reducir la multicolinealidad puede dañar el poder predictivo del modelo. Se debe tener cuidado al seleccionar las variables, porque muchos de los procedimientos de selección se distorsionan mucho por la multicolinealidad y no hay seguridad de que el modelo final muestre menor grado de multicolinealidad que la que había en los datos originales.

http://carlos2524.jimdo.com/

TABLA 10.8 Modelos de regresión de subconjunto, para los datos del acetileno

	Modelo A		Modelo B		Modelo C		Modelo D		Modelo E	
Regresor	$\hat{\beta}$	VIF	$\hat{\beta}$	VIF	$\hat{\beta}$	VIF	$\hat{\beta}$	VIF	$\hat{\beta}$	VIF
x_1	0.8595	12.23	0.8930	1.14					0.8355	3.09
x_2	0.1657	1.06	0.0265	1.45	0.1369	1.48	0.2365	1.72	0.2280	1.46
x_3	−0.0506	12.33			0.7225	2.64	−1.1281	5.74		
x_1x_2			0.0451	1.37			−0.4805	31.03	−0.4788	30.37
x_1x_3							−3.3111	1221.64		
x_2x_3					0.0555	1.52	−0.2728	35.43	−0.2529	33.35
x_1^2			0.005	1.20			−1.4902	353.20	0.2161	12.39
x_2^2			0.0294	1.72	0.0238	1.86	−0.0919	3.16	−0.0844	1.84
x_3^2					0.0420	2.98	−1.5068	320.71	−0.1879	19.70
	$R^2 = 0.9198$		$R^2 = 0.9946$		$R^2 = 0.9807$		$R^2 = 0.9974$		$R^2 = 0.9970$	
	$MS_{\text{Res}} = 0.00668$		$MS_{\text{Res}} = 0.00054$		$MS_{\text{Res}} = 0.00193$		$MS_{\text{Res}} = 0.00037$		$MS_{\text{Res}} = 0.00037$	
	$C_p = 201.5626$		$C_p = 9.9948$		$C_p = 46.3566$		$C_p = 8.7971$		$C_p = 7.7876$	

La fuerte multicolinealidad mostrada por los datos del acetileno en el ejemplo 10.1 se debe, probablemente, a la elección del modelo, porque x_1 (la temperatura) y x_3 (el tiempo de contacto) están muy correlacionados, y la ecuación original es totalmente cuadrática en los tres regresores. La selección de variables puede ser eficaz para reducir este deterioramiento.

La tabla 10.8 muestra la información resumida acerca de cinco modelos distintos de subconjuntos de variables para los datos del acetileno. El modelo A es un modelo con todos los términos lineales, que trata de reducir la multicolinealidad omitiendo los términos de segundo orden. El modelo B es totalmente cuadrático sólo en x_1 y x_2, es un modelo candidato lógico, de subconjunto, que resulta de eliminar x_3 (y todos sus términos de segundo orden), por su alta correlación con x_1. La aplicación de la regresión por segmentos y la selección en avance, con un nivel de selección para entrada de variable de $\alpha = 0.25$, también conduce al modelo B. El modelo C es totalmente cuadrático en x_2 y x_3, y es un compañero obvio del modelo B. El modelo D, que contiene ocho términos (sólo se eliminó x_1) se obtiene por eliminación en reversa, con nivel de selección $\alpha = 0.10$, para eliminar variable. El modelo E es la ecuación con mínimo C_p de todas las regresiones posibles.

El modelo lineal A, con tres términos, no es muy satisfactorio. Todavía tiene deterioramiento moderadamente fuerte, y los estadísticos R^2 y cuadrados medios residuales no se comparan en forma favorable con el ajuste con el modelo completo. El estadístico C_p indica que queda bastante sesgo en este modelo, además las gráficas de residuales muestran una distribución que indica que la ecuación predice de más, en forma sistemática cuando las conversiones porcentuales son bajas, y predice de menos cuando las conversiones porcentuales son altas. Los diagramas de dispersión de los datos también indican que son necesarios al menos algunos de los términos de segundo orden, en conclusión, el modelo A es inadecuado.

Los modelos B y C, que son ecuaciones cuadráticas completas determinadas eliminando x_1 o x_3, parecen modelos más razonables de subconjunto, aunque el estadístico C_p para el modelo C es muy grande. Ninguna ecuación presenta multicolinealidad importante. La única dificultad potencial con el modelo B (que sólo contiene a x_1 y x_2), es que el analista no lo podría usar para predecir cambios en el porcentaje de conversión, en función de cambios en el tiempo de contacto (x_3). De igual manera, el modelo C no informa el efecto de la temperatura (x_1) sobre el proceso; sin embargo, quizá se deben entregar **las dos** ecuaciones al usuario. El uso conjunto de ambas ecuaciones puede ser más informativo

http://carlos2524.jimdo.com/

que una sola ecuación; véase, no obstante, que aun con dos ecuaciones no es obvio cómo el usuario predeciría el efecto de cambios simultáneos tanto de x_1 como de x_3.

Los modelos D y E parecen menos satisfactorios que el B y el C. Ambas ecuaciones tienen factores de inflación de varianza grandes, y eso implica que la eliminación de variables no fue tan efectiva para combatir la multicolinealidad como en los modelos B y C.

También se puede investigar la eficacia predictiva de esos modelos de subconjunto. El porcentaje de conversión calculado en los 10 puntos de la figura 10.3, para las cinco ecuaciones, se ve a continuación.

	Predicción del porcentaje de conversión				
Punto	Modelo A	Modelo B	Modelo C	Modelo D	Modelo E
Interpolación					
A	19.01	16.98	16.66	16.18	16.51
B	18.52	13.48	13.61	14.91	14.84
E	32.04	29.54	29.35	28.48	28.87
F	35.16	35.41	38.94	35.29	35.56
I	49.30	47.99	46.39	46.89	47.96
J	47.25	49.40	49.04	50.31	49.50
Extrapolación					
C	21.87	22.08	48.65	– 68.86	23.35
D	37.10	38.21	44.38	– 44.71	19.36
G	33.34	31.43	49.76	19.71	26.57
H	45.61	49.40	22.47	– 157.56	46.57

Como se dijo antes, el modelo A lineal, de tres términos, tiende a sobreestimar bajas conversiones porcentuales y a subestimar altas conversiones porcentuales. Los otros cuatro modelos de subconjunto parecen funcionar en forma satisfactoria para la interpolación; sin embargo, el modelo D produce predicciones negativas de porcentaje de conversión en tres de los cuatro puntos de extrapolación, y eso es un resultado claramente no razonable. Los modelos B, C y E producen, al menos, predicciones factibles en los puntos de extrapolación, aunque la consistencia con las tres ecuaciones no es espectacular.

Llegados a este punto, es difícil recomendar una ecuación "final". Es posible que los modelos B y C se puedan usar en conjunto, o quizá sea satisfactorio el modelo E. Es necesario más análisis para emitir una recomendación final, no obstante, queda claro que la reespecificación del modelo, si se hace con cuidado, puede ayudar a reducir los efectos de la multicolinealidad.

10.5.3 Regresión ridge

Cuando se aplica el método de mínimos cuadrados a datos no ortogonales, se pueden obtener estimaciones muy malas de los coeficientes de regresión. En la sección 10.3 se dijo que la varianza de los estimados por mínimos cuadrados, de los coeficientes de regresión, puede estar muy inflada y que la longitud del vector de los estimados de los parámetros por mínimos cuadrados es excesiva, en promedio. Eso implica que el valor absoluto de los estimados por mínimos cuadrados es demasiado grande, y que esos estimados son muy inestables, indicando con esto que sus magnitudes y signos pueden cambiar mucho con una muestra distinta.

http://carlos2524.jimdo.com/

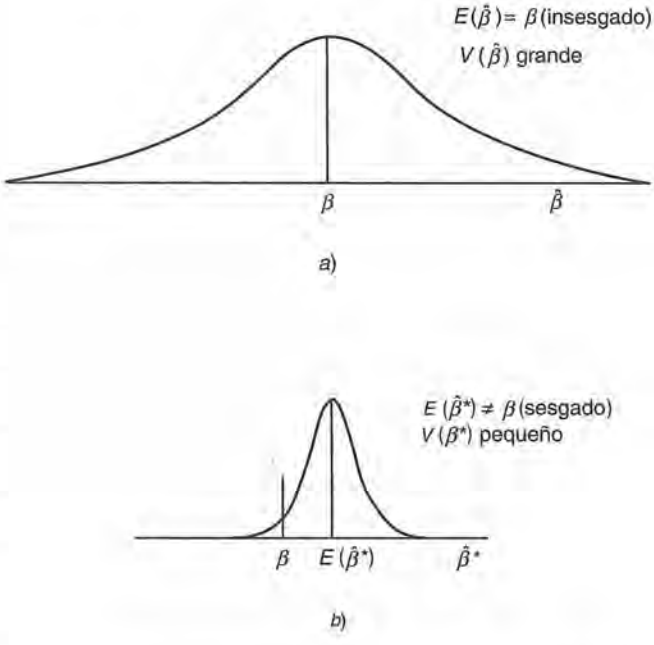


Figura 10.4 Distribución de muestreo de a) estimadores insesgados y b) sesgados de β. (Adaptado de Marquardt y Snee [1975], con autorización del editor.)

El problema con el método de los mínimos cuadrados es el requisito que $\hat{\beta}$ sea un **estimador insesgado** de β. La propiedad de Gauss-Markov, citada en la sección 3.2.3, asegura que el estimador de mínimos cuadrados tiene varianza mínima en la clase de los estimadores lineales insesgados, pero no hay garantía de que esa varianza sea pequeña. El caso se ve en la figura 10.4a, donde se ve la distribución de muestreo de $\hat{\beta}$, el estimador insesgado de β. La varianza de $\hat{\beta}$ es grande, y eso implica que los intervalos de confianza de β serán grandes, y que el estimado de punto de $\hat{\beta}$ es muy inestable.

Una forma de aliviar este problema es eliminar el requisito que el estimador de β sea insesgado. Supóngase que se puede determinar un **estimador sesgado** de β, por ejemplo $\hat{\beta}^*$, que tenga menor varianza que el estimador insesgado $\hat{\beta}$. El error cuadrático medio del estimador $\hat{\beta}^*$ se define como sigue:

$$\text{MSE}(\hat{\beta}^*) = E(\hat{\beta}^* - \beta)^2 = \text{Var}(\hat{\beta}^*) + \left[E(\hat{\beta}^*) - \beta\right]^2 \tag{10.11}$$

es decir

$$\text{MSE}(\hat{\beta}^*) = \text{Var}(\hat{\beta}^*) + (\text{sesgo en } \hat{\beta}^*)^2$$

Nótese que el MSE no es más que la distancia esperada, de $\hat{\beta}^*$ a β, elevada al cuadrado [véase la Ec. (10.7)]. Si se permite una pequeña cantidad de sesgo en $\hat{\beta}^*$, la varianza de $\hat{\beta}^*$ se puede hacer pequeña, de tal modo que su error cuadrático medio (MSE) de $\hat{\beta}^*$ sea menor que la varianza del estimador insesgado $\hat{\beta}$. La figura 10.4b ilustra el caso en el que la varianza del estimador sesgado es bastante menor que la del estimador insesgado de la figura 10.4a. En consecuencia, los intervalos de confianza de β serán mucho más angostos si se usa el estimador sesgado. La pequeña varianza del estimador sesgado implica también que $\hat{\beta}^*$ es un estimador más estable de β que el estimador insesgado $\hat{\beta}$.

http://carlos2524.jimdo.com/

Se han desarrollado varios procedimientos para obtener estimadores sesgados de coeficientes de regresión. Uno de esos procedimientos es la **regresión ridge** (o de cresta), propuesta originalmente por Hoerl y Kennard [1970a, b]. El estimador ridge se determina resolviendo una versión ligeramente modificada de las ecuaciones normales. En forma específica, el estimador de ridge de $\hat{\beta}_R$ se define como la solución de

$$(\mathbf{X}'\mathbf{X} + k\mathbf{I})\hat{\boldsymbol{\beta}}_R = \mathbf{X}'\mathbf{y} \tag{10.12}$$

que es

$$\hat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \tag{10.13}$$

en la que $k \geq 0$ es una constante que selecciona el analista. El procedimiento se llama regresión ridge porque las operaciones matemáticas se parecen al método de análisis ridge, usado antes por Hoerl [1959], para describir el comportamiento de superficies de respuesta de segundo orden; nótese que cuando $k = 0$, el estimador ridge es igual al estimador por mínimos cuadrados.

El estimador ridge es una transformación lineal del estimador de mínimos cuadrados, ya que

$$\begin{aligned}\hat{\boldsymbol{\beta}}_R &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}\\ &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}\\ &= \mathbf{Z}_k\hat{\boldsymbol{\beta}}\end{aligned}$$

Por consiguiente, como $E(\hat{\beta}_R) = E(\mathbf{Z}_k\hat{\beta}) = \mathbf{Z}_k\hat{\beta}$, entonces $\hat{\beta}_R$ es un estimador sesgado de β. Se suele llamar **parámetro de sesgo** a la constante k. La matriz de covarianza de $\hat{\beta}_R$ es

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}_R) = \sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} \tag{10.14}$$

El error cuadrático medio del estimador de cresta es

$$\begin{aligned}\operatorname{MSE}(\hat{\boldsymbol{\beta}}_R) &= \operatorname{Var}(\hat{\boldsymbol{\beta}}_R) + (\text{sesgo en } \hat{\boldsymbol{\beta}}_R)^2\\ &= \sigma^2\operatorname{Tr}\left[(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\right]\\ &\quad + k^2\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-2}\boldsymbol{\beta}\\ &= \sigma^2\sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j + k)^2} + k^2\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-2}\boldsymbol{\beta}\end{aligned} \tag{10.15}$$

en donde $\lambda_1, \lambda_2, \ldots, \lambda_p$ son los eigenvalores de $\mathbf{X}'\mathbf{X}$. El primer término del lado derecho de la ecuación (10.15) es la suma de las varianzas de los parámetros en $\hat{\beta}_R$, y el segundo término es el cuadrado del sesgo. Si $k > 0$ el sesgo en $\hat{\beta}_R$ aumenta al aumentar k. Sin embargo, la varianza disminuye al aumentar k.

Al usar la regresión ridge sería bueno escoger un valor de k, tal que la reducción en el término de varianza sea mayor que el aumento en el sesgo al cuadrado. Si se puede hacer, el error cuadrático medio del estimador ridge $\hat{\beta}_R$ será menor que la varianza del estimador

http://carlos2524.jimdo.com/

$\hat{\beta}$, por mínimos cuadrados. Hoerl y Kennard demostraron que existe un valor de k distinto de cero para el cual el MSE de $\hat{\beta}_R$ es menor que la varianza del estimador $\hat{\beta}$ por mínimos cuadrados, siempre y cuando $\beta'\beta$ sea acotado. La suma de cuadrados residuales es

$$\begin{aligned} SS_{\text{Res}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R)'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R) \\ &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}}_R - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}_R - \hat{\boldsymbol{\beta}}) \end{aligned} \tag{10.16}$$

Como el primer término del lado derecho de la ecuación (10.16) es la suma de cuadrados residuales, para los $\hat{\beta}$ estimados por mínimos cuadrados, se ve que cuando aumenta k, la suma de cuadrados residuales aumenta. En consecuencia, como la suma total de cuadrados es fija, R^2 disminuye al aumentar k, así, el estimado ridge, en general, no llegará a ser el mejor "ajuste" a los datos, pero ello no debe preocupar mucho, porque interesa más obtener un conjunto estable de estimados de los parámetros. Los estimados ridge pueden dar como resultado una ecuación que funcione mejor para predecir observaciones futuras, en comparación con los mínimos cuadrados (aunque no hay una **demostración** concluyente de que así sucederá).

Hoerl y Kennard han sugerido que un valor adecuado de k puede determinarse por inspección de la **traza ridge**. La traza ridge es una gráfica de los elementos de $\hat{\beta}_R$ en función de k, para valores de k que suelen estar en el intervalo 0 a 1. Marquardt y Snee [1975] sugieren usar hasta unos 25 valores de k, con espaciado aproximadamente logarítmico en el intervalo [0, 1]. Si la multicolinealidad es grave, la inestabilidad de los coeficientes de regresión será obvia de acuerdo con la traza ridge. Al aumentar k, algunos de los estimados ridge variarán en forma dramática. En cierto valor de k se estabilizarán los estimados ridge $\hat{\beta}_R$. El objetivo es seleccionar un valor de k razonablemente pequeño, en el cual los estimados ridge de $\hat{\beta}_R$ sean estables. Es posible que así se produzca un conjunto de estimados con MSE menor que los estimados por mínimos cuadrados.

Ejemplo 10.2 *Datos del acetileno*

Para obtener la solución ridge para los datos del acetileno se deben resolver las ecuaciones $(\mathbf{X}'\mathbf{X} + k\mathbf{I})\,\hat{\beta}_R = \mathbf{X}'\mathbf{y}$ con varios valores de $0 \le k \le 1$, estando $\mathbf{X}'\mathbf{X}$ y $\mathbf{X}'\mathbf{y}$ en forma de correlación. En la figura 10.5 se ve la traza ridge, y en la tabla 10.9 se ven los coeficientes ridge para varios valores de k. Esta tabla muestra también el cuadrado medio residual y R^2

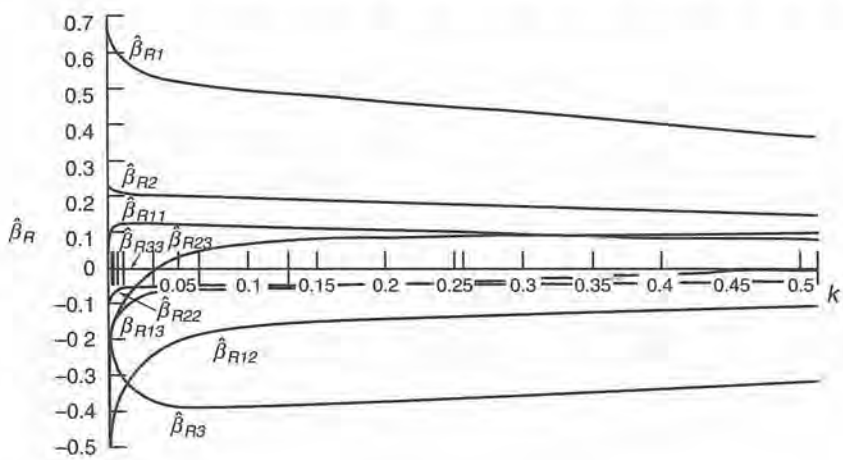


Figura 10.5 Traza ridge para los datos del acetileno, usando nueve regresores.

http://carlos2524.jimdo.com/

TABLA 10.9 Coeficientes a diversos valores de k

k	.000	.001	.002	.004	.008	.016	.032	.064	.128	.256	.512
$\beta_{R.1}$	.3377	.6770	.6653	.6362	.6003	.5672	.5392	.5122	.4806	.4379	.3784
$\beta_{R.2}$	.2337	.2242	.2222	.2199	.2173	.2148	.2117	.2066	.1971	.1807	.1554
$\beta_{R.3}$	−.6749	−.2129	−.2284	−.2671	−.3134	−.3515	−.3735	−.3800	−.3724	−.3500	−.3108
$\beta_{R.12}$	−.4799	−.4479	−.4258	−.3913	−.3437	−.2879	−.2329	−.1862	−.1508	−.1249	−.1044
$\beta_{R.13}$	−2.0344	−.2774	−.1887	−.1350	−.1017	−.0809	−.0675	−.0570	−.0454	−.0299	−.0092
$\beta_{R.23}$	−.2675	−.2173	−.1920	−.1535	−.1019	−.0433	.0123	.0562	.0849	.0985	.0991
$\beta_{R.11}$	−.8346	.0643	.1035	.1214	.1262	.1254	.1249	.1258	.1230	.1097	.0827
$\beta_{R.22}$	−.0904	−.0732	−.0682	−.0621	−.0558	−.0509	−.0481	−.0464	−.0444	−.0406	−.0341
$\beta_{R.33}$	−1.0015	−.2451	−.1853	−.1313	−.0825	−.0455	−.0267	−.0251	−.0339	−.0464	−.0586
MS_{Res}	.00038	.00047	.00049	.00054	.00062	.00074	.00094	.00127	.00206	.00425	.01002
R^2	.998	.997	.997	.997	.996	.996	.994	.992	.988	.975	.940

http://carlos2524.jimdo.com/

para cada modelo ridge; nótese que cuando aumenta k, MS_{Res} aumenta y R^2 disminuye. La traza ridge ilustra la inestabilidad de la solución por mínimos cuadrados, porque hay grandes cambios en los coeficientes de regresión para pequeños valores de k, sin embargo, los coeficientes se estabilizan con rapidez, a medida que k aumenta.

Se requiere juicio para interpretar la traza ridge y seleccionar un valor adecuado de k. Se debe elegir a k lo bastante grande como para producir coeficientes estables, pero que no sea innecesariamente grande, porque así se introduce más sesgo y aumenta el cuadrado medio residual. En la figura 10.5 se ve que se alcanza una estabilidad razonable de coeficientes en la región $0.008 < k < 0.064$, sin un grave aumento del cuadrado medio de residuales (o sin pérdida de R^2). Si se escoge $k = 0.032$, el modelo de regresión ridge es

$$\hat{y} = 0.5392x_1 + 0.2117x_2 - 0.3735x_3 - 0.2329x_1x_2 - 0.0675x_1x_3 + 0.0123x_2x_3 + 0.1249x_1^2 - 0.0481x_2^2 - 0.0267x_3^2$$

Nótese que en este modelo los estimados de β_{13}, β_{11} y β_{23} son bastante menores que los obtenidos por mínimos cuadrados, y que los estimados de β_{23} y β_{11} que antes eran negativos, ahora son positivos. El modelo ridge, expresado en términos de los regresores originales, es

$$\hat{P} = 0.7598 + 0.1392T + 0.0547H - 0.0965C - 0.0680TH - 0.0194TC + 0.0039CH + 0.0407T^2 - 0.0112H^2 - 0.0067C^2$$

La figura 10.6 muestra el funcionamiento del modelo ridge en la predicción, tanto para interpolación (puntos A, B, E, F, I y J), como en extrapolación (puntos C, D, G y H). Al

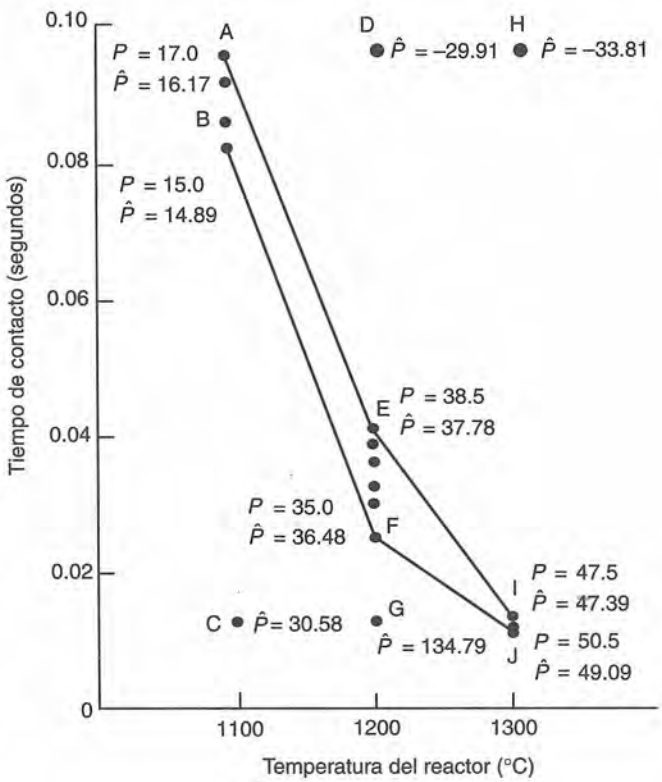


Figura 10.6 Funcionamiento del modelo ridge con $k = 0.032$, en predicción y en extrapolación, para los datos del acetileno. (Adaptado de Marquardt y Snee [1975], con autorización del editor.)

http://carlos2524.jimdo.com/

comparar las figuras 10.6 y 10.3 se nota que el modelo ridge hace predicciones tan buenas como el modelo de mínimos cuadrados con nueve términos, en la frontera de la región cubierta por los datos, sin embargo, el modelo ridge produce predicciones mucho más realistas cuando se extrapola, en comparación con los mínimos cuadrados, además, la eficiencia de extrapolación del modelo ridge es más o menos comparable a los modelos de subconjunto, por mínimos cuadrados, B $(x_1, x_2, x_1x_2, x_1^2, x_2^2)$, C$(x_2, x_3, x_2x_3, x_2^2, x_3^2)$ y E$(x_1, x_2, x_1x_2, x_2x_3, x_1^2, x_2^2, x_3^2)$ que se desarrollaron en la sección 10.5.2. Por lo anterior, se concluye que el método de regresión ridge, en este ejemplo, produjo un modelo mejor que el ajuste original por mínimos cuadrados, con nueve términos, y que es comparable con los modelos de mínimos cuadrados desarrollados por selección de variables.

Los estimados de regresión ridge se pueden calcular usando un programa ordinario de cómputo de mínimos cuadrados, y aumentando como sigue los datos estandarizados:

$$\mathbf{X}_A = \begin{bmatrix} \mathbf{X} \\ \sqrt{k}\,\mathbf{I}_p \end{bmatrix} \qquad \mathbf{y}_A = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_p \end{bmatrix}$$

en donde $\sqrt{k\,\mathbf{I}_p}$ es una matriz diagonal de $p \times p$, con elementos diagonales iguales a la raíz cuadrada del parámetro de sesgo, y $\mathbf{0}_p$ es un vector de ceros de $p \times 1$; entonces, los estimados ridge se calculan como

$$\hat{\boldsymbol{\beta}}_R = (\mathbf{X}'_A\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{y}_A = (\mathbf{X}'\mathbf{X} + k\,\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}$$

La tabla 10.10 muestra la matriz aumentada $\mathbf{X}_A$ y el vector $\mathbf{y}_A$ necesarios para producir la solución ridge para los datos del acetileno con $k = 0.032$.

Otras propiedades de la regresión ridge

La figura 10.7 ilustra la geometría de la regresión ridge en un problema con dos regresores. El punto $\hat{\boldsymbol{\beta}}$ en el centro de las elipses corresponde a la solución de mínimos cuadrados, donde la suma de cuadrados residuales asume su valor mínimo. La pequeña elipse repre-

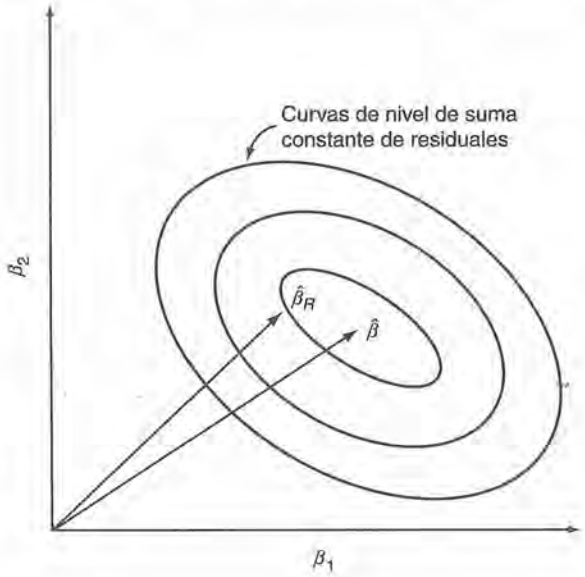


Figura 10.7 Interpretación geométrica de la regresión ridge.

http://carlos2524.jimdo.com/

TABLA 10.10 Matriz aumentada X_A y vector y_A para generar la solución de cresta con los datos del acetileno y $k = 0.032$

$$X_A = \begin{bmatrix}
.280224 & -.22544 & -.23106 & -.33766 & -.02085 & 309525 & .078278 & -.04116 & -.03452 \\
.280224 & -.15704 & -.23106 & -.25371 & -.02085 & .236588 & .078278 & -.1327 & -.03452 \\
.280224 & -.06584 & -.23514 & -.14179 & -.02579 & .140577 & .078278 & -.20378 & -.02735 \\
.280224 & .048167 & -.2229 & -.00189 & -.01098 & .0196 & .078278 & -.2107 & -.04847 \\
.280224 & .207774 & -.21882 & .193976 & -.00605 & -.14065 & .078278 & -.06745 & -.05526 \\
.280224 & .481385 & -.23106 & .529744 & -.02085 & -.44415 & .078278 & .593235 & -.03452 \\
-.04003 & -.32577 & -.00255 & -.00413 & .258949 & .073001 & -.29746 & .152387 & -.23548 \\
-.04003 & -.22544 & -.01887 & -.02171 & .261769 & .088842 & -.29746 & -.04116 & -.23418 \\
-.04003 & -.06584 & -.06784 & -.0497 & -.270231 & .089856 & -.29746 & -.20378 & -.21822 \\
-.04003 & .048167 & -.1168 & -.06968 & .278693 & .043276 & -.29746 & -.2107 & -.18419 \\
-.04003 & .207774 & -.05152 & -.09766 & .267411 & .019961 & -.29746 & -.06745 & -.22554 \\
-.04003 & .481385 & .005609 & -.14563 & .257539 & .0832021 & -.29746 & .593235 & -.23538 \\
-.36029 & -.32577 & .356528 & .452517 & -.29615 & -.46678 & .328768 & .152387 & .243742 \\
-.36029 & -.22544 & .470781 & .294227 & -.47384 & -.42042 & .328768 & -.04116 & .599999 \\
-.36029 & -.06584 & .421815 & .042401 & -.39769 & -.05859 & .328768 & -.20378 & .435271 \\
-.36029 & .207774 & .37285 & -.3893 & -.32153 & .427375 & .328768 & -.06745 & .288613 \\
.17888 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & .17888 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & .17888 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & .17888 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & .17888 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & .17888 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .17888 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .17888 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .17888
\end{bmatrix} \qquad y_A = \begin{bmatrix}
.27979 \\ .305829 \\ .312339 \\ .26894 \\ .24724 \\ .182141 \\ -.1759 \\ -.09995 \\ -.03486 \\ -.02401 \\ .041094 \\ .051944 \\ -.0458 \\ -.04146 \\ -.33865 \\ -.14335 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix}$$

http://carlos2524.jimdo.com/

senta el lugar geométrico de los puntos en el plano β_1, β_2 en donde es constante la suma de cuadrados residuales, en determinado valor mayor que el mínimo. El estimado ridge $\hat{\beta}_R$ es el vector más corto desde el origen que produce una suma de cuadrados residuales igual al valor representado por la elipse pequeña. Esto es, el estimado ridge de $\hat{\beta}_R$ produce el vector de coeficientes de regresión con la norma mínima, consistente con un aumento especificado en la suma de cuadrados residuales. Se observa que el estimador ridge contrae al estimador de mínimos cuadrados hacia el origen, en consecuencia, los estimadores ridge (y otros estimadores sesgados en general) se llaman a veces estimadores de **contracción**. Hocking [1976] observó que el estimador ridge contrae al estimador de mínimos cuadrados, con respecto a las curvas de nivel de $\mathbf{X}'\mathbf{X}$. Esto es, $\hat{\beta}_R$ es la solución de

$$\underset{\beta}{\text{Minimizar}}\,(\beta - \hat{\beta})'\mathbf{X}'\mathbf{X}(\beta - \hat{\beta})$$

$$\text{sujeta a } \beta'\beta \leq d^2 \tag{10.17}$$

en donde el radio d depende de k.

Muchas propiedades del estimador hacen suponer que el valor de k es fijo. En la práctica, ya que k se estima a partir de los datos, por inspección de la traza ridge, k es **estocástica**. Es interesante preguntar si las propiedades de optimalidad citadas por Hoerl y Kennard son válidas si k es estocástica. Algunos autores han demostrado, mediante simulaciones, que la regresión ridge ofrece, en general, mejoras en el error cuadrático medio, en comparación con los mínimos cuadrados, cuando k se estima a partir de los datos. Theobald [1974] ha generalizado las condiciones bajo las cuales la regresión ridge produce MSE menores que los mínimos cuadrados. La mejoría esperada depende de la orientación del vector β en relación con los eigenvectores de $\mathbf{X}'\mathbf{X}$. La mejora esperada es máxima cuando β coincide con el eigenvector asociado con el máximo eigenvalor de $\mathbf{X}'\mathbf{X}$. Otros resultados interesantes se pueden ver en Lowerre [1974] y en Mayer y Willke [1973].

Obenchain [1977] demostró que los estimadores ridge contraídos no estocásticamente producen los mismos estadísticos t y F, para probar hipótesis, que los mínimos cuadrados. Así, aunque la regresión ridge conduce a estimados sesgados de punto, en general no requiere una nueva teoría de distribución. Sin embargo, las propiedades de distribución todavía no se conocen cuando k se escoge estocásticamente. Se podría suponer que cuando k es pequeña, la inferencia basada en la teoría de la distribución normal acostumbrada sería aplicable en forma aproximada.

Relación con otros estimadores

La regresión ridge se relaciona mucho con la **estimación bayesiana**. En general, si se puede describir información previa acerca de β mediante una distribución normal p-variada con vector promedio β_0 y matriz de covarianza $\mathbf{V}_0$, el estimador de Bayes de β es

$$\hat{\beta}_B = \left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} + \mathbf{V}_0^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{y} + \mathbf{V}_0^{-1}\beta_0\right)$$

El uso de los métodos bayesianos en la regresión se describe en Leamer [1973, 1978] y Zellner [1971]. Dos desventajas principales de este método son que el analista de datos debe hacer una declaración explícita sobre la forma de la distribución *a priori*, y no se comprende

http://carlos2524.jimdo.com/

mucho la teoría estadística; sin embargo, si se escoge el promedio *a priori* $\beta_0 = \mathbf{0}$, y $\mathbf{V}_0 = \sigma_0^2\mathbf{I}$, entonces se obtiene

$$\hat{\beta}_B + (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \equiv \hat{\beta}_R, \quad 2qk = \frac{\sigma^2}{\sigma_0^2}$$

que es el usual estimador ridge. De hecho, el método de mínimos cuadrados se puede considerar como un estimador de Bayes que usa una distribución uniforme no acotada y *a priori*, para β. El estimador ridge es el resultado de una distribución *a priori* que impone condiciones débiles de acotamiento para β. Véase también Lindley y Smith [1972].

Theil y Goldberger [1961] y Theil [1963] introdujeron un procedimiento llamado **estimación mixta**. En esta técnica se usa información anterior o adicional, para aumentar los datos en forma directa, en lugar de a través de una distribución *a priori*. La estimación mixta comienza con el modelo acostumbrado de regresión $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ y supone que el analista puede escribir un conjunto de restricciones $r < p$ *a priori*, para β, tales que

$$\mathbf{a} = \mathbf{D}\beta + \delta$$

en donde $E(\delta) = \mathbf{0}$, $\mathrm{Var}(\delta) = \mathbf{V}$, $\mathbf{D}$ es una matriz de $r \times p$ de constantes conocidas, de rango r, y $\mathbf{a}$ es un vector de $r \times 1$, de variables aleatorias. Si se aumentan $\mathbf{y}$ y $\mathbf{X}$ para producir

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{D} \end{bmatrix} \beta + \begin{bmatrix} \varepsilon \\ \delta \end{bmatrix}$$

y al aplicar mínimos cuadrados se obtiene el estimador mixto insesgado

$$\hat{\beta}_{ME} = \left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} + \mathbf{D}'\mathbf{V}^{-1}\mathbf{D}\right)^{-1}\left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{y} + \mathbf{D}'\mathbf{V}^{-1}\mathbf{a}\right)$$

Ahora bien, si $\mathbf{D} = \mathbf{A}$ (siendo $\mathbf{A}'\mathbf{A} = \mathbf{I}$), $\mathbf{a} = \mathbf{0}$ y $\mathbf{V} = \sigma_1^2\mathbf{I}$, entonces

$$\hat{B}_{ME} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \equiv \hat{\beta}_R, \qquad k = \frac{\sigma^2}{\sigma_1^2}$$

Si bien la estimación mixta y la regresión ridge podrán ser numéricamente equivalentes, hay una diferencia de los puntos de vista que se adoptan. En la estimación mixta, $\mathbf{a}$ es una variable aleatoria, mientras que en la regresión ridge los elementos de $\mathbf{a}$ son constantes especificadas, con lo que se obtiene un estimador sesgado. La estimación mixta es menos formal que la bayesiana, porque permite introducir información anterior sin una especificación completa de una distribución *a priori* para β. En Belsley, Kuh y Welsch [1980] se describe una aplicación de la estimación mezclada para combatir la multicolinealidad.

Métodos para seleccionar k

Gran parte de la controversia acerca de la regresión de cresta se centra en torno a la elección del parámetro k de sesgo. La elección de k por inspección de la traza ridge es un procedimiento subjetivo que requiere criterio por parte del analista. Algunos autores han propuesto

http://carlos2524.jimdo.com/

procedimientos para elegir k, que son más analíticos. Hoerl, Kennard y Baldwin [1975] sugieren que una elección adecuada de k es

$$k = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \tag{10.18}$$

en donde $\hat{\beta}$ y $\hat{\sigma}^2$ se determinan con la solución por mínimos cuadrados. Demostraron, con simulaciones, que el estimador ridge resultante tuvo mejora importante en el MSE respecto al de mínimos cuadrados. En una publicación posterior, Hoerl y Kennard [1976] propusieron un procedimiento iterativo de estimación basado en la ecuación (10.18). En forma específica sugirieron la siguiente secuencia de estimaciones de β y de k:

$$\hat{\beta} : \quad k_0 = \frac{p\hat{\sigma}^2}{\hat{\beta}\hat{\beta}}$$

$$\hat{\beta}_R(k_0) : \quad k_1 = \frac{p\hat{\sigma}^2}{\hat{\beta}'_R(k_0)\hat{\beta}_R(k_0)}$$

$$\hat{\beta}_R(k_1) : \quad k_2 = \frac{p\hat{\sigma}^2}{\hat{\beta}'_R(k_1)\hat{\beta}_R(k_1)}$$

$$\vdots$$

El cambio relativo de k_j se usa para terminar el procedimiento. Si

$$\frac{k_{j+1} - k_j}{k_j} > 20T^{-1.3}$$

el algoritmo debe continuar; de lo contrario debe terminar y usarse $\hat{\beta}_R(k_j)$, siendo $T = \mathrm{Tr}(\mathbf{X}'\mathbf{X})^{-1}/p$. Este criterio de terminación se seleccionó porque T aumenta con la dispersión de los eigenvalores de $\mathbf{X}'\mathbf{X}$, permitiendo mayor contracción al aumentar el grado el deterioramiento de los datos. Los autores citan estudios de simulación en los que esta regla de terminación funcionó bien.

McDonald y Galarneau [1975] sugieren escoger k de tal modo que

$$\hat{\beta}'_R\,\hat{\beta}_R = \hat{\beta}'\hat{\beta} - \hat{\sigma}^2 \sum_{j=1}^{p} \left(\frac{1}{\lambda_j}\right) \tag{10.19}$$

Para los casos en que el lado derecho de esta ecuación (10.19) es negativo, investigaron igualar $k = 0$ (mínimos cuadrados) o $k = \infty$ ($\hat{\beta}_R = \mathbf{0}$). Ningún método, en todos los casos, fue mejor que el de mínimos cuadrados.

Los métodos que se han descrito para elegir k se enfocan en la mejora de los estimados de los coeficientes de regresión. Si el modelo se va a usar para predicción, será más adecuado tener en cuenta criterios orientados a las predicciones, para elegir k. Mallows [1973] modificó el estadístico C_p para formar uno C_k que se puede usar para determinar k. Propuso graficar C_k en función de V_k, siendo

http://carlos2524.jimdo.com/

$$C_k = \frac{SS_{\text{Res}}(k)}{\hat{\sigma}^2} - n + 2 + 2\,\text{Tr}(\mathbf{XL})$$

$$V_k = 1 + \text{Tr}(\mathbf{X'XLL'})$$

$$\mathbf{L} = (\mathbf{X'X} + k\mathbf{I})^{-1}\mathbf{X'}$$

y $SS_{\text{Res}}(k)$ es la suma de cuadrados residuales en función de k. La sugerencia es escoger la k que minimice C_k. Obsérvese que

$$\mathbf{XL} + \mathbf{X}(\mathbf{X'X} + k\mathbf{I})^{-1}\mathbf{X'} \equiv \mathbf{H}_k$$

y que $\mathbf{H}_k$ equivale a la matriz de sombrero en mínimos cuadrados ordinarios.

Otra posibilidad es un procedimiento $\text{PRESS}_{\text{Ridge}}$ donde interviene

$$\text{PRESS}_{\text{Cresta}} = \sum_{i=1}^{n} \left(\frac{e_{i,k}}{1 - h_{ii,k}} \right)^2$$

en donde $e_{i,k}$ es el i-ésimo residual para un valor determinado de k, y $h_{ii,k}$ es el i-ésimo elemento diagonal de $\mathbf{H}_k$. El valor de k se escoge de tal modo que minimice $\text{PRESS}_{\text{Ridge}}$. Nótese que este procedimiento sólo es una aproximación al valor verdadero de $\text{PRESS}_{\text{Ridge}}$ que se podría obtener en realidad eliminando una por una las observaciones (recalculando cada vez los estimados ridge), porque cuando se centran y escalan los datos, la eliminación de un punto de dato cambia las constantes de centrado y escalado, y en consecuencia las observaciones, sin embargo, si no hay grandes diagonales en la matriz de sombrero (puntos influyentes) y el tamaño de la muestra no es pequeño, $\text{PRESS}_{\text{Ridge}}$ es una buena aproximación a PRESS **exacto**:

$$\text{PRESS}(k) = \sum_{i=1}^{n} e_{(i),k}^2$$

en donde $e_{(i),k}^2$ es el residual obtenido realmente conservando la i-ésima observación para una k particular, centrando y escalando los datos, ajustando el modelo ridge y calculando $\hat{y}_{(i),k}$; por consiguiente, $e_{(i),k} = y_i - \hat{y}_{(i),k}$. Existe un programa de cómputo (PROC IML, SAS Institute [1987]) para graficar PRESS (k) en función de k.

Wahba, Golub y Health [1979] sugirieron el estadístico generalizado de validación cruzada

$$\text{GCV} = \frac{\sum_{i=1}^{n} e_{i,k}^2}{\{n - [1 + \text{Tr}(\mathbf{H}_k)]\}^2}$$

Al escoger k se selecciona de tal modo que se minimiza el estadístico GCV. Hay una relación obvia con los procedimientos análogos a PRESS, descritos arriba.

Hay muchas otras posibilidades de escoger k. Por ejemplo, Marquardt [1970] propuso usar un valor de k tal que el VIF quede entre 1 y 10, de preferencia más cercano a 1. Dempster, Shatzoff y Wermuth [1971], Goldstein y Smith [1974], Lawless y Wang [1976], Lindley y Smith [1972] y Obenchain [1975] propusieron otros métodos de selección de k.

http://carlos2524.jimdo.com/

No hay seguridad de que cualquiera de ellos produzca determinaciones semejantes de k, además, no hay garantía de que esos métodos sean mejores que la inspección directa de la traza ridge.

Ejemplo 10.3 *Elección de k para los datos del acetileno*

Se ilustrarán dos alternativas de la inspección de la traza de cresta, para seleccionar k para los datos del acetileno. De acuerdo con la ecuación (10.18), el valor de k para esos datos es

$$k = \frac{p\hat{\sigma}^2}{\hat{\beta}\hat{\beta}} = \frac{9(0.00038)}{6.77} = 0.0005$$

en la que $\hat{\beta}$ y $\hat{\sigma}^2$ se toman de la solución estandarizada de mínimos cuadrados, en la tabla 10.2. Nótese que este valor de k es bastante menor que el que se seleccionó por inspección de la traza ridge. La columna 1 de esta tabla muestra los correspondientes coeficientes de regresión ridge.

El estimado iterativo de k, sugerido por Hoerl y Kennard [1976] se puede determinar comenzando con $k_0 = 0.0005$ y se obtiene $\hat{\beta}_R(k_0)$ en la columna 1 de la tabla 10.11. Nótese que $\hat{\beta}'_R(k_0)\,\hat{\beta}_R(k_0) = 1.0999$, esto es, la longitud del vector de los coeficientes de regresión ridge, elevada al cuadrado, con $k_0 = 0.0005$, es mucho más corta que la longitud del vector de mínimos cuadrados, elevada al cuadrado. La nueva estimación de k es ahora

$$k_1 = \frac{p\hat{\sigma}^2}{\hat{\beta}'_R(k_0)\hat{\beta}_R(k_0)} = \frac{9(0.00038)}{1.0999} = 0.0031$$

El criterio para terminar es comparar el cambio relativo de k_0 con $20T^{-1.3}$, siendo

$$T = \frac{\mathrm{Tr}(\mathbf{X}'\mathbf{X})^{-1}}{p} = \frac{\sum_{j=1}^{p}(1/\lambda_j)}{p} = \frac{10\,611.6918}{9} = 1\,179.08$$

TABLA 10.11 Modelos ridge para dos opciones de k

Término	(1) $k = 0.0005$	(2) $k = 0.0005$
x_1	0.6650	0.6283
x_2	0.2258	0.2193
x_3	−0.2295	−0.2775
x_1x_2	−0.4608	−0.3816
x_1x_3	−0.4234	−0.1261
x_2x_3	−0.2329	−0.1429
x_1^2	−0.0072	0.1233
x_2^2	−0.0769	−0.0606
x_3^2	−0.3197	−0.1197
MS_{Res}	0.00045	0.00055
R^2	0.9973	0.9967

http://carlos2524.jimdo.com/

y λ_j son los eigenvalores de $\mathbf{X}'\mathbf{X}$. Así, ya que

$$\frac{k_1 - k_0}{k_0} = \frac{0.0031 - 0.0005}{0.0005} = 5.20 > 20T^{-1.3} = 20(1\,179.08)^{-1.3} = 0.0020$$

se debe hacer otra iteración. Se puede demostrar que el valor final de k determinado con este procedimiento es $k = 0.005$. La columna 2 de la tabla 10.11 presenta los coeficientes respectivos ridge.

Al comparar los coeficientes de regresión ridge para esos dos valores de k, con los correspondientes para $k = 0.032$, seleccionado de la traza ridge, se notan llamativas semejanzas. Tanto para $k = 0.0005$ como para $k = 0.005$ los grandes estimados iniciales de β_{13}, β_{11} y β_{33}, obtenidos por mínimos cuadrados, se redujeron en forma importante. Para $k = 0.005$, el estimado ridge de β_{11} sigue siendo negativo, mientras que para $k = 0.005$ el estimado se vuelve positivo. Nótese que el valor mayor de k, seleccionado con la traza ridge, produce más contracción en los estimados y se obtiene otro cambio de signo más para β_{23}. Sin embargo, los tres modelos son bastante parecidos, y aparentemente producen relaciones más razonables, para los datos del acetileno, que el modelo ordinario de mínimos cuadrados, con nueve términos.

Regresión ridge y selección de variables

Los algoritmos normales para selección de variables no funcionan bien, con frecuencia, cuando los datos son muy multicolineales. Sin embargo, la selección de variables suele funcionar bastante bien cuando los regresores son ortogonales, o casi ortogonales. Si los regresores se hicieron más ortogonales usando estimadores sesgados, la selección de variables puede ser una buena estrategia. Hoerl y Kennard [1970b] sugieren usar la traza ridge como guía para seleccionar variables. Proponen las siguientes reglas para eliminar regresores del modelo completo:

1. Eliminar los regresores que son estables, pero que tienen poco poder de predicción; esto es, los regresores con coeficientes estandarizados pequeños.
2. Eliminar los regresores con coeficientes inestables que no mantengan su poder de predicción que son, los coeficientes inestables que se corren hacia cero.
3. Eliminar uno o más de los regresores restantes, que tengan coeficientes inestables.

El subconjunto de los regresores restantes, cuya cantidad sea p, por ejemplo, se usa en el modelo "final". Se pueden examinar esos regresores, para ver si forman un subconjunto casi ortogonal. Esto se puede hacer graficando $\hat{\beta}'_R(k)\,\hat{\beta}'_R(k)$, la longitud del vector de coeficientes, elevada al cuadrado, en función de k. Si los regresores son ortogonales, la longitud al cuadrado del vector de estimados ridge debe ser $\hat{\beta}'\hat{\beta}/(1 + k)^2$, siendo $\hat{\beta}$ el estimado ordinario de β por mínimos cuadrados, en consecuencia, si el modelo de subconjunto contiene regresores casi ortogonales, las funciones $\hat{\beta}'_R(k)\,\hat{\beta}_R(k)$ y $\hat{\beta}'\beta/(1 + k)^2$, graficadas en función de k, deben ser muy parecidas.

http://carlos2524.jimdo.com/

Ejemplo 10.4 *Selección de variables con los datos del acetileno*

Para ilustrar la utilidad de la regresión ridge en la selección de variables, se examinará el modelo ridge para los datos del acetileno. La tabla 10.9 indica que cuando k aumenta, los coeficientes grandes de x_1x_3 y de x_3^2 se contraen hacia cero en forma rápida. Con $k = 0.032$, también los coeficientes de x_2x_3 y de x_2^2 son pequeños. Como esos coeficientes se calcularon con datos estandarizados, sus magnitudes reflejan la contribución relativa de los regresores correspondientes, supóngase que se decide eliminar esos cuatro términos, con base en sus pequeños coeficientes de regresión en el modelo ridge. Si se aplica la regresión ridge a los cinco regresores restantes (x_1, x_2, x_3, x_1^2, x_1x_2), se produce la traza ridge de la figura 10.8. La tabla 10.12 presenta los coeficientes de regresión, cuadrado medio de residuales y R^2 para varios valores de k. La traza ridge es mucho más estable que cuando se consideran los nueve regresores, lo que significa que la introducción de más sesgo al aumentar k no cambia mucho los coeficientes de regresión, además, hay poco cambio inmediato en MS_{Res} o en R^2. El factor máximo de inflación de varianza, para el ajuste por mínimos cuadrados a esos cinco regresores es 53.52 (todavía bastante grande, pero con una gran mejora respecto al modelo original). La eliminación de x_1x_3, x_2x_3, x_2^2 y x_3^2 mejoró mucho el adecuamiento de los datos. Se llega a la conclusión que esta ecuación de subconjunto de mínimos cuadrados es un modelo razonable para los datos. Nótese que es un modelo de subconjunto de variables distinto de los que se obtuvieron con los métodos usuales de selección de variable en la sección 10.5.2. Snee [1973] también sugirió este modelo de subconjunto partiendo del análisis gráfico de los datos.

La figura 10.9 muestra la eficiencia de predicción de este modelo de subconjunto de mínimos cuadrados. Al comparar esta figura con las 10.3 y 10.6 se observa que este modelo de subconjunto predice más o menos igual de bien que el modelo ridge en interpolación y extrapolación. Su eficiencia de predicción también es comparable, más o menos, con los modelos B, C y E de subconjunto, en la sección 10.5.2, así, parece que la regresión ridge mejoró mucho los resultados del modelo original de nueve términos, de mínimos cuadrados, sea en forma directa con el modelo ridge con nueve términos, o en forma indirecta, como ayuda en la eliminación de variables.

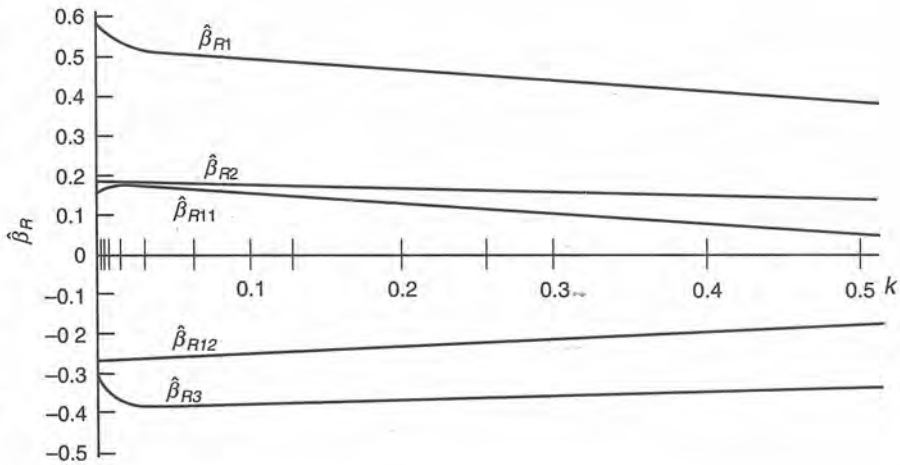


Figura 10.8 Traza ridge para los datos del acetileno, usando cinco regresores.

http://carlos2524.jimdo.com/

TABLA 10.12 Coeficientes con diversos valores de k

k	.000	.001	.002	.004	.008	.016	.032	.064	.128	.256	.512
$\beta_{R.1}$	.6036	.5946	.5878	.5771	.5622	.5449	.5268	.5070	.4815	.4437	.3882
$\beta_{R.2}$	.1940	.1935	.1934	.1931	.1925	.1913	.1891	.1849	.1773	.1645	.1442
$\beta_{R.3}$	−.3240	−.3317	−.3386	−.3493	−.3634	−.3778	−.3883	−.3916	−.3853	−.3667	−.3318
$\beta_{R.12}$	−.2738	−.2729	−.2725	−.2717	−.2703	−.2677	−.2628	−.2540	−.2384	−.2135	−.1785
$\beta_{R.11}$	.1728	.1742	.1755	.1774	.1793	.1795	.1756	.1642	.1416	.1052	.0583
MS_{Res}	.00063	.00063	.00063	.00063	.00064	.00065	.00069	.00082	.00127	.00266	.00632
R^2	.994	.994	.994	.994	.994	.994	.993	.992	.987	.973	.937

http://carlos2524.jimdo.com/

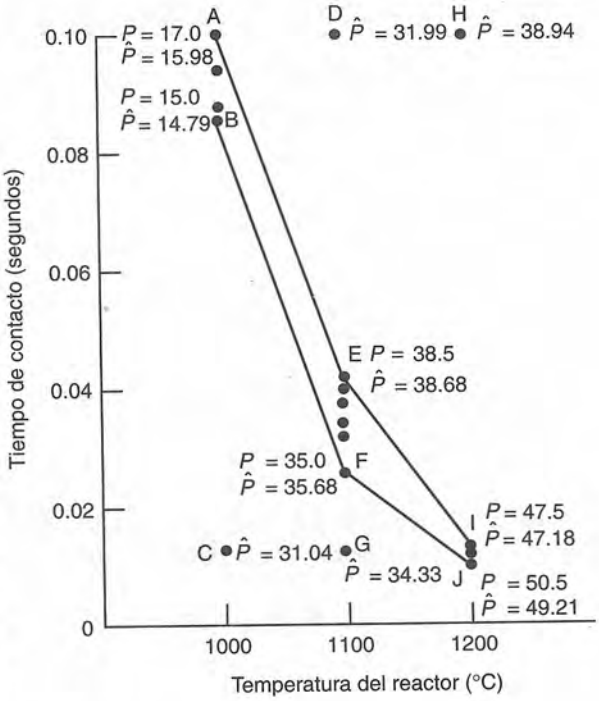


Figura 10.9 Predicción y extrapolación con el modelo de mínimos cuadrados con cinco variables, para los datos del acetileno. (Adaptado de Marquardt y Snee [1975], con autorización del editor.)

10.5.4 Otros métodos

Regresión ridge generalizada

Hoerl y Kennard [1970a] propusieron una extensión del procedimiento ordinario de regresión ridge, que permite tener parámetros separados de sesgo para cada regresor. A este procedimiento se le llama **regresión ridge generalizada**.

La descripción de este método se simplifica un poco si se transforman los datos al espacio de los **regresores ortogonales**. Para hacerlo, recuérdese que si Λ es la matriz diagonal de $p \times p$ cuyos elementos de diagonal principal son los eigenvalores $\lambda_1, \lambda_2, \ldots, \lambda_p$ de $\mathbf{X}'\mathbf{X}$, y si $\mathbf{T}$ es la matriz ortogonal correspondiente de eigenvectores, entonces

$$\mathbf{T}'\mathbf{X}'\mathbf{X}\mathbf{T} = \Lambda \tag{10.20}$$

Si se definen

$$\mathbf{Z} = \mathbf{X}\mathbf{T} \tag{10.21}$$

y

$$\alpha = \mathbf{T}'\beta \tag{10.22}$$

el modelo lineal se transforma en

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= (\mathbf{Z}\mathbf{T}')(\mathbf{T}\boldsymbol{\alpha}) + \boldsymbol{\varepsilon} \\ &= \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \end{aligned} \tag{10.23}$$

http://carlos2524.jimdo.com/

El estimador de α por mínimos cuadrados es la solución a

$$(\mathbf{Z}'\mathbf{Z})\hat{\alpha} = \mathbf{Z}'\mathbf{y} \tag{10.24}$$

que equivale a

$$\Lambda\hat{\alpha} = \mathbf{Z}'\mathbf{y} \tag{10.25}$$

o sea

$$\hat{\alpha} = \Lambda^{-1}\mathbf{Z}'\mathbf{y} \tag{10.26}$$

El vector de los estimados originales se obtiene con la ecuación (10.22) esto es

$$\hat{\beta} = \mathbf{T}\hat{\alpha} \tag{10.27}$$

Con frecuencia se dice que la ecuación (10.23) es la **forma canónica** del modelo. En términos de la forma canónica, el **estimador ridge generalizado** es la solución de

$$(\Lambda + \mathbf{K})\hat{\alpha}_{\text{GR}} = \mathbf{Z}'\mathbf{y} \tag{10.28}$$

donde $\mathbf{K}$ es una matriz diagonal con elementos $(k_1, k_2, \ldots, k_p)$. En términos del modelo original, los coeficientes ridge generalizados son

$$\hat{\beta}_{\text{GR}} = \mathbf{T}\hat{\alpha}_{\text{GR}} \tag{10.29}$$

Ahora se considerará la selección de los parámetros de sesgo en $\mathbf{K}$. El error cuadrático medio para la regresión ridge generalizada es

$$\begin{aligned}
\text{MSE}(\hat{\beta}_{\text{GR}}) &= E\left[(\hat{\beta}_{\text{GR}} - \beta)'(\hat{\beta}_{\text{GR}} - \beta)\right] \\
&= E\left[(\hat{\alpha}_{\text{GR}} - \alpha)'(\hat{\alpha}_{\text{GR}} - \alpha)\right] \\
&= \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k_j)^2} + \sum_{j=1}^{p} \frac{\alpha_j^2 k_j^2}{(\lambda_j + k_j)^2}
\end{aligned} \tag{10.30}$$

El primer término del lado derecho de la ecuación (10.30) es la suma de las varianzas de los estimados de parámetro, y el segundo término es el sesgo elevado al cuadrado. El error cuadrático medio, ecuación (10.30), se minimiza al escoger

$$k_j = \frac{\sigma^2}{\alpha_j^2}, \qquad j = 1, 2, \ldots, p \tag{10.31}$$

Desafortunadamente, la k_j óptima depende de los parámetros desconocidos σ^2 y α_j. Hoerl y Kennard [1970a] sugieren un método iterativo para determinar las k_j. A partir de la solución de mínimos cuadrados se obtiene un estimado inicial de las k_j; por ejemplo

$$k_j^0 = \frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}, \qquad j = 1, 2, \ldots, p$$

http://carlos2524.jimdo.com/

Estos estimados iniciales de las k_j se usan para calcular los estimados ridge generalizados iniciales, a partir de

$$\hat{\alpha}^0_{\mathrm{GR}} = (\Lambda + \mathbf{K}^0)^{-1}\mathbf{Z}'\mathbf{y}$$

en donde $\mathbf{K}^0 = \mathrm{diag}(k_1^0, k_2^0, \ldots, k_p^0)$. A continuación se usan los estimados iniciales $\hat{\alpha}^0_{\mathrm{GR}}$ para revisar los estimados de las k_j:

$$k_j^1 = \frac{\hat{\sigma}^2}{\left(\hat{\alpha}^0_{\mathrm{GR},j}\right)^2}, \qquad j = 1, 2, \ldots, p$$

Estos nuevos valores de k_j^1 se pueden usar para corregir los estimados del α. El proceso iterativo debe continuar hasta que se obtengan estimados estables de parámetro. Una medida de estabilidad que se usa con frecuencia es la longitud del vector $\hat{\alpha}'_{\mathrm{GR}}\hat{\alpha}_{\mathrm{GR}}$ elevada al cuadrado. En forma específica, si la longitud del vector de parámetros estimados elevada al cuadrado no cambia mucho desde la iteración $i - 1$ hasta la iteración i, entonces se debe terminar. En caso contrario, se debe continuar el proceso iterativo de estimación, nótese que no hay alguna presentación gráfica útil de los coeficientes, como la traza ridge en la regresión ridge generalizada.

Se podrá usar la ecuación (10.31) para justificar la elección del parámetro de sesgo k en la regresión ordinaria ridge. El valor de k en la ecuación (10.18) es un promedio ponderado de las k_j de la ecuación (10.31). Es claro que si se combinan las k_j para producir un solo parámetro de sesgo, no se debería usar un promedio ordinario, porque una α_j pequeña produciría un valor grande de k, induciendo demasiado sesgo en los estimados de parámetro. Sin embargo, la media armónica de las k_j es

$$k_h = \frac{p}{\sum_{j=1}^{p}(1/k_j)} = \frac{p}{\sum_{j=1}^{p}\left(\alpha_j^2/\sigma^2\right)} = \frac{p\sigma^2}{\sum_{j=1}^{p}\alpha_j^2} = \frac{p\sigma^2}{\alpha'\alpha} = \frac{p\sigma^2}{\beta'\beta} = k$$

como se ve en la ecuación (10.18).

Hemmerle [1975] demostró que el procedimiento iterativo de Hoerl y Kennard, para estimar las k_j, tiene una solución explícita en forma cerrada de modo que, en general, no es necesaria la iteración. En forma específica, sea

$$\hat{\alpha}_{\mathrm{GR}} = \mathbf{B}\hat{\alpha} \tag{10.32}$$

donde $\hat{\alpha}$ es el estimador de mínimos cuadrados, y $\mathbf{B}$ es una matriz diagonal de elementos no negativos $b_1, b_2, \ldots, b_p$. Hocking, Speed y Lynn [1976] demostraron que los resultados de Hemmerle son seleccionar

$$b_j = \begin{cases} 0 & \text{si } \tau_j^2 < 4 \\ 0.5 + \left[0.25 - \left(1/\tau_j^2\right)\right]^{1/2} & \text{si } \tau_j^2 \geq 4 \end{cases} \tag{10.33}$$

donde $\tau_j^2 = \hat{\alpha}^2\lambda_j/\hat{\sigma}^2$. Si se observa que τ_j es el estadístico t asociado con el j-ésimo regresor, se observa que si el estadístico t es "pequeño", el coeficiente ridge generalizado se

http://carlos2524.jimdo.com/

iguala a cero, mientras que si el estadístico t es "grande", ese coeficiente es una fracción b_j del coeficiente de mínimos cuadrados. En otras palabras, los coeficientes no significativos se contraen a cero, mientras que los significativos se contraen con menos severidad. A esta solución la llamaremos **solución ridge generalizada y totalmente iterada**.

Hemmerle hizo notar que la solución ridge generalizada y totalmente iterada da como resultado, con frecuencia, la introducción de demasiado sesgo (o demasiada contracción) en los estimados finales de parámetro. Propuso una técnica para evitarlo, con base en restringir a la suma de cuadrados residuales para evitar un aumento importante no deseado. Recomendó establecer un límite para la pérdida total de R^2 y que se reparta esa pérdida en forma proporcional a los regresores individuales. Su procedimiento da como resultado valores modificados de b_j, que se indican con b_j^*, definidos por

$$b_j^* = 1 - \sqrt{m}\,(1 - b_j) \tag{10.34}$$

en donde m es la relación de la perdida admisible de R^2 entre la perdida de R^2 si se usa b_j de la ecuación (10.33). Hocking *et al.* [1976] pone objeción al uso de la ecuación (10.34), porque hace que todas las b_j^* sean distintas de cero. Si se igualan a cero algunas de las b_j^*, se elimina la fuerte influencia de un eigenvalor pequeño sobre la inflación de la varianza. El uso de la ecuación (10.34) permite regresar la influencia de ese eigenvalor.

En una publicación posterior, Hemmerle y Brantle [1978] sugieren seleccionar las k_j con base en la minimización de un estimador del criterio del error cuadrático medio. Se desarrolla una solución explícita de forma cerrada para el vector de estimados de parámetro resultante. También se presenta un procedimiento para obtener estimados generalizados restringidos ridge, donde se escogen restricciones para utilizar la información anterior sobre los signos de los coeficientes de regresión, sin embargo, una simulación de Monte Carlo no pudo demostrar alguna superioridad obvia de este método.

Desafortunadamente no hay una "óptima" elección bien definida de las k_j para la regresión ridge generalizada. Nosotros concordamos con Hemmerle [1975] en que una estimación ridge generalizada y una totalmente iterada da como resultado, con frecuencia, demasiada contracción, y que es adecuado algún tipo de procedimiento restringido, en especial para datos que sean muy deteriorados. En la práctica suele funcionar bien restringir el aumento máximo de la suma de cuadrados residuales de 1 a 20%; sin embargo se necesita trabajar mucho para desarrollar mejores lineamientos de elección de los parámetros k_j y controlar la cantidad de contracción.

Regresión por componentes principales

También se pueden obtener los estimadores sesgados de coeficientes de regresión con un procedimiento llamado **regresión por componentes principales**. Examínese la forma canónica del modelo

$$\mathbf{y} = \mathbf{Z}\alpha + \varepsilon \tag{10.35}$$

en donde

$$\mathbf{Z} = \mathbf{XT}, \qquad \alpha = \mathbf{T}'\beta, \qquad \mathbf{T}'\mathbf{X}'\mathbf{XT} = \mathbf{Z}'\mathbf{Z} = \Lambda$$

Recuérdese que $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ es una matriz diagonal de $p \times p$ de los eigenvalores de $\mathbf{X}'\mathbf{X}$, y $\mathbf{T}$ es una matriz ortogonal cuyas columnas son los eigenvectores asociados

http://carlos2524.jimdo.com/

con $\lambda_1, \lambda_2, \ldots, \lambda_p$. Las columnas de $\mathbf{Z}$, que definen un nuevo conjunto de regresores ortogonales, como

$$\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_p]$$

se llaman **componentes principales**.

El estimador de $\hat{\alpha}$ por mínimos cuadrados es

$$\hat{\alpha} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \Lambda^{-1}\mathbf{Z}'\mathbf{y} \tag{10.36}$$

y la matriz de covarianza de $\hat{\alpha}$ es

$$\text{Var}(\hat{\alpha}) = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1} = \sigma^2\Lambda^{-1} \tag{10.37}$$

Así, un pequeño eigenvalor de $\mathbf{X}'\mathbf{X}$ indica que la varianza del coeficiente ortogonal de regresión correspondiente será grande. Ya que

$$\mathbf{Z}'\mathbf{Z} = \sum_{i=1}^{p}\sum_{j=1}^{p}\mathbf{Z}_i\mathbf{Z}_j' = \Lambda$$

se llama con frecuencia "varianza del j-ésimo componente principal" al eigenvalor λ_j. Si todas las λ_j son iguales a 1, los regresores **originales** son ortogonales, mientras que si una λ_j es exactamente igual a cero, esto implica una regresión perfectamente lineal entre los regresores **originales**. Si hay una o más de las λ_j cercanas a cero, quiere decir que hay multicolinealidad. Nótese también que la matriz de covarianza de los coeficientes estandarizados de regresión $\hat{\beta}$ es

$$\text{Var}(\hat{\beta}) = \text{Var}(\mathbf{T}\hat{\alpha}) = \mathbf{T}\Lambda^{-1}\mathbf{T}'\sigma^2$$

Esto implica que la varianza de $\hat{\beta}_j$ es $\sigma^2(\Sigma_{j=1}^{p}t_{ji}^2/\lambda_i)$. Por consiguiente, la varianza de $\hat{\beta}_j$ es una combinación lineal de los recíprocos de los eigenvalores. Eso demuestra cómo uno o más eigenvalores pequeños pueden destruir la precisión del estimado $\hat{\beta}_j$, por mínimos cuadrados.

Antes se hizo la observación de cómo los eigenvalores y los eigenvectores de $\mathbf{X}'\mathbf{X}$ proporcionan información específica acerca de la naturaleza de la multicolinealidad. Como $\mathbf{Z} = \mathbf{XT}$, entonces

$$\mathbf{Z}_i = \sum_{j=1}^{p} t_{ji}\mathbf{X}_j \tag{10.38}$$

siendo $\mathbf{X}_j$ la j-ésima columna de la matriz $\mathbf{X}$ y t_{ji} son los elementos de la i-ésima columna de $\mathbf{T}$ (el i-ésimo eigenvector de $\mathbf{X}'\mathbf{X}$). Si la varianza del i-ésimo componente principal (λ_i) es pequeña, quiere decir que $\mathbf{Z}_i$ es casi constante, y de acuerdo con la ecuación (10.38), hay una combinación lineal de los **regresores originales** que es casi constante. Es la definición de multicolinealidad, lo que nos dice que, las t_{ji} son las constantes de la ecuación (10.2), por consiguiente, la ecuación (10.38) explica por qué los elementos del eigenvector asociado con un pequeño eigenvalor de $\mathbf{X}'\mathbf{X}$ identifican a los regresores que intervienen en la multicolinealidad.

http://carlos2524.jimdo.com/

El método de regresión por componentes principales combate a la multicolinealidad, al usar en el modelo menos componentes que el conjunto completo de componentes principales. Para obtener el estimador de componentes principales, se supone que los regresores están ordenados por eigenvalores decrecientes, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$. Supóngase que los últimos s eigenvalores de éstos son aproximadamente iguales a cero. En la regresión por componentes principales, se eliminan los componentes principales que correspondan a eigenvalores cercanos a cero, y se aplican los mínimos cuadrados a los componentes restantes. Esto es,

$$\hat{\boldsymbol{\alpha}}_{\mathrm{PC}} = \mathbf{B}\hat{\boldsymbol{\alpha}} \tag{10.39}$$

en donde $b_1 = b_2 = \cdots = b_{p-s} = 1$, y $b_{p-s+1} = b_{p-s+2} = \cdots = b_p = 0$. Así, el estimador de componentes principales es

$$\hat{\boldsymbol{\alpha}}_{\mathrm{PC}} = \begin{array}{cl} \left[\begin{array}{c} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_{p-s} \\ \hline 0 \\ 0 \\ \vdots \\ 0 \end{array}\right] & \begin{array}{l} \\ \\ \\ p-s \text{ componentes} \\ \\ \\ \\ s \text{ componentes} \end{array} \end{array}$$

o bien, en términos de los regresores normalizados,

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\mathrm{PC}} &= \mathbf{T}\hat{\boldsymbol{\alpha}}_{\mathrm{PC}} \\ &= \sum_{j=1}^{p-s} \lambda_j^{-1} \mathbf{t}_j' \mathbf{X}' \mathbf{y} \mathbf{t}_j \end{aligned} \tag{10.40}$$

Un estudio de simulación hecho por Gunst y Mason [1977] demostró que la regresión por componentes principales permite tener una mejora considerable respecto a mínimos cuadrados, cuando los datos están determinados, también hacen notar que otra ventaja de los componentes principales es que están disponibles los procedimientos de teoría de la distribución exacta y de la selección de variables (véase Mansfield, Webster y Gunst [1977]). Algunos paquetes de cómputo resuelven la regresión por componentes principales.

Ejemplo 10.5 Regresión por componentes principales para los datos del acetileno

Se ilustrará el uso de la regresión por componentes principales con los datos del acetileno. Se comienza con la transformación lineal $\mathbf{Z} = \mathbf{XT}$ que transforma los regresores estandarizados originales en un conjunto ortogonal de variables (los componentes principales). Los eigenvalores λ_j y la matriz $\mathbf{T}$ para los datos del acetileno se muestran en la tabla 10.13. Esta matriz indica que la relación (por ejemplo), entre z_1 y los regresores estandarizados es

$$\begin{aligned} z_1 = {} & 0.3387x_1 + 0.1324x_2 - 0.4137x_3 - 0.2191x_1x_2 + 0.4493x_1x_3 \\ & + 0.2524x_2x_3 - 0.4056x_1^2 + 0.0258x_2^2 - 0.4667x_3^2 \end{aligned}$$

http://carlos2524.jimdo.com/

TABLA 10.13 Matriz T de eigenvectores y eigenvalores λ_j para los datos del acetileno

Eigenvectores									Eigenvalores λ_j
.3387	.1057	.6495	.0073	.1428	−.2488	−.2077	−.5436	.1768	4.20480
.1324	.3391	−.0068	−.7243	−.5843	.0205	−.0102	−.0295	−.0035	2.16261
−.4137	−.0978	−.4696	−.0718	−.0182	.0160	−.1468	−.7172	.2390	1.13839
−.2191	.5403	.0897	.3612	−.1661	.3733	−.5885	.0909	.0003	1.04130
.4493	.0860	−.2863	.1912	−.0943	.0333	.0575	.1543	.7969	0.38453
.2524	−.5172	−.0570	−.3447	.2007	.3232	−.6209	.1280	.0061	0.04951
−.4056	−.0742	.4404	−.2230	.1443	.5393	.3233	.0565	.4087	0.01363
.0258	.5316	−.2240	−.3417	.7342	−.0705	−.0057	.0761	.0050	0.00513
−.4667	−.0969	.1421	−.1337	−.0350	−.6299	−.3089	.3631	.3309	0.00010

Las relaciones entre los componentes principales restantes $z_2, z_3, \ldots, z_9$, y los regresores estandarizados, se determinan en forma parecida. La tabla 10.14 muestra los elementos de la matriz $\mathbf{Z}$ (a veces se llaman calificaciones de componentes principales).

El estimador de componentes principales reduce los efectos de la multicolinealidad al usar un subconjunto de los componentes principales en el modelo. Como hay cuatro eigenvalores pequeños para los datos del acetileno, eso implica que hay cuatro componentes principales que se deben eliminar. Se excluirán z_6, z_7, z_8 y z_9, y se considerarán sólo regresiones donde intervengan los primeros cinco componentes principales.

Supóngase que se considera un modelo de regresión donde sólo interviene el primer componente principal, como en

$$y = \alpha_1 z_1 + \varepsilon$$

El modelo ajustado es

$$y = -0.35225 z_1$$

o sea que $\hat{\alpha}'_{PC} = [-0.35225, 0, 0, 0, 0, 0, 0, 0, 0]$. Los coeficientes, en términos de los regresores estandarizados se determinan a partir de $\hat{\beta}_{PC} = \mathbf{T}\hat{\alpha}_{PC}$. La parte A de la tabla 10.15 muestra los coeficientes de regresión estandarizados que resultan, así como los coeficientes de regresión en términos de los regresores originales centrados. Nótese que aun cuando sólo se incluye un solo componente principal, el modelo produce estimados para los nueve coeficientes de regresión estandarizados.

En las partes B, C, D y E, respectivamente, muestran los resultados de agregar los demás componentes principales, z_2, z_3, z_4 y z_5, uno por uno al modelo. Se ve que al usar distintas cantidades de componentes principales en el modelo se producen estimados bastante diferentes de los coeficientes de regresión. Además, los estimados de componentes principales difieren mucho de los obtenidos por mínimos cuadrados (véase, por ejemplo, la Tabla 10.9). Sin embargo, el procedimiento de componentes principales con cuatro o cinco componentes incluidos da como resultado estimados de coeficientes que no difieren mucho de los que se obtienen con los demás métodos sesgados de estimación (véanse los estimados de regresión ridge ordinaria, en las tablas 10.8 y 10.11). Los componentes principales contraen los estimados grandes de β_{13} y β_{33} por mínimos cuadrados, y cambia el signo del estimado de β_{11}, originalmente negativo, obtenido por mínimos cuadrados. El modelo de cinco componentes no degrada en forma apreciable el ajuste con los datos origina-

http://carlos2524.jimdo.com/

TABLA 10.14 Matriz Z = XT de ortogonalización para datos del acetileno

Observación	Z_1	Z_2	Z_3	$Z_4(=Z_{x_1x_2})$	$Z_5(=Z_{x_1x_3})$	$Z_6(=Z_{x_2x_3})$	$Z_7(=Z_{x_1^2})$	$Z_8(=Z_{x_2^2})$	$Z_9(=Z_{x_3^2})$
1	.5415	– 1.0347	1.0487	– .1880	1.7389	– .6593	.6492	.7822	.2402
2	.4846	– .8830	1.1638	– .0468	.8909	– .3874	.5067	.2045	– .1939
3	.4046	– .6129	.2914	.0676	– .0025	– .1631	.2187	– .0898	– 1.6609
4	.3388	– .1513	3.3176	.1315	– .7526	.3579	.1269	– 1.2150	.9250
5	.2353	.6905	1.2785	– .0089	– 1.0842	.6884	– .4181	– 1.2768	1.6754
6	.0310	2.7455	.9535	– .7783	.2235	.2093	– 1.1200	1.3128	– 1.1453
7	.5940	– .0165	– 1.0885	1.1554	1.5790	.1926	– 1.3363	– .4626	.5964
8	.6385	– .2399	– .9170	1.0916	.3634	.4238	– 1.2453	– .7138	– .3611
9	.7139	– .3558	– .7151	.8354	– .9374	.3207	– .6525	.5144	– .7716
10	.7436	– .2228	– .6170	.5668	– 1.4297	– .4038	.5657	2.5203	1.4085
11	.7668	.1034	– .8626	– .0706	– 1.3472	– .3706	1.5958	– .8815	– 1.3485
12	.8726	1.1054	– 1.5272	– 1.8442	.8129	– .9285	.8411	– .8981	.7053
13	– 1.7109	.8164	– .3702	1.2052	.8885	1.9123	2.0708	.2251	– .1036
14	– 2.1618	.1860	– .1026	.5619	– .1290	– 2.5588	– .3380	– .1080	.8652
15	– 1.6050	– .6784	– .2117	– .3325	– .7456	– .0658	– .8259	– .4662	– 1.0012
16	– .8875	– 1.4521	– .6417	– 2.3461	– .0690	1.4324	– .6387	.5524	.1699

http://carlos2524.jimdo.com/

TABLA 10.15 Regresión por componentes principales, para los datos del acetileno

	Componentes principales en el modelo									
	A		B		C		D		E	
Parámetro	z_1		z_1, z_2		$z_1, z_2 \cdot z_3$		z_1, z_2, z_3, z_4		z_1, z_2, z_3, z_4, z_5	
	Estimado estandarizado	Estimado original	Estimado estandarizado	Estimado original	Estimado estandarizado	Estimado original	Estimado estandarizado	Estimado original	Estimado estandarizado	Estimado original
β_0	.0000	42.1943	.0000	42.2219	.0000	36.6275	.0000	34.6688	.0000	34.7517
β_1	.1193	1.4194	.1188	1.4141	.5087	6.0508	.5070	6.0324	.5056	6.0139
β_2	.0466	.5530	.0450	.5346	.0409	.4885	.2139	2.5438	.2195	2.6129
β_3	−.1457	−1.7327	−.1453	−1.7281	−.4272	−5.0830	−.4100	−4.8803	−.4099	4.8757
β_{12}	−.0772	−1.0369	−.0798	−1.0738	−.0260	−.3502	−.1123	−1.5115	−.1107	−1.4885
β_{13}	.1583	2.0968	.1578	2.0922	−.0143	−.1843	−.0597	−.7926	−.0588	−.7788
β_{23}	.0889	1.2627	.0914	1.2950	.0572	.8111	.1396	1.9816	.1377	1.9493
β_{11}	−.1429	−2.1429	−.1425	−2.1383	.1219	1.8295	.1751	2.6268	.1738	2.6083
β_{22}	.0091	.0968	.0065	.0691	−.1280	−1.3779	−.0460	−.4977	−.0533	−.5760
β_{33}	−.1644	−1.9033	−.1639	−1.8986	−.0786	−.9125	−.0467	−.5392	−.0463	−.5346
R^2	.5217		.5218		.9320		.9914		.9915	
MS_{Res}	.079713		.079705		.011333		.001427		.00142	

http://carlos2524.jimdo.com/

les, porque se ha perdido poco en R^2 obtenido con el modelo de mínimos cuadrados. Por todo lo anterior se llega a la conclusión de que la relación basada en los cinco primeros componentes principales proporciona un modelo más plausible para los datos del acetileno, que el que se obtuvo a través de mínimos cuadrados ordinarios.

Estimados de rango fraccionario

Marquardt [1970] sugirió una generalización de la regresión por componentes principales. Creyó que es demasiado restrictiva la suposición de un rango entero para la matriz **X**, y propuso un estimador de "rango fraccionario" que permita que el rango sea una función continua por segmentos. En forma específica, si el rango de la matriz **X** está en el intervalo $[r, r+1]$, el estimador de rango fraccionario de Marquardt es

$$\hat{\alpha}_{\mathrm{FR}} = (1-c)\hat{\alpha}_r + c\hat{\alpha}_{r+1} \tag{10.41}$$

para $0 \le c \le 1$, donde $\hat{\alpha}_r$ y $\hat{\alpha}_{r+1}$ son los estimadores de componentes principales de α, para los rangos supuestos r y $r+1$, esto significa que los últimos $r-1$ elementos de $\hat{\alpha}_{\mathrm{FR}}$ son cero, el $(p-r+1)$-ésimo elemento es $c\hat{\alpha}_{p-r+1}$ y los primeros $p-r$ elementos son los estimadores de mínimos cuadrados $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{p-r}$. Los criterios para seleccionar r y c se describen por Hocking *et al.* [1976].

Análisis de regresión de raíz latente

El procedimiento de regresión de (o *por*) raíz latente fue desarrollado por Hawkins [1973] y por Webster *et al.* [1974], de acuerdo con la misma filosofía que en la de componentes principales. El procedimiento forma estimadores, a partir de los eigenvalores (o raíces latentes) de la matriz de correlación de regresores y variables de respuesta

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} 1 & \mathbf{y}'\mathbf{X} \\ \mathbf{X}'\mathbf{y} & \mathbf{X}'\mathbf{X} \end{bmatrix}$$

Sean $0 \le l_0 \le l_1 \le \cdots \le l_p$, y sean $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_p$ los eigenvalores y eigenvectores de $\mathbf{A}'\mathbf{A}$, y los últimos p elementos de $\mathbf{v}_j$ se representan por δ_j, de modo que $\mathbf{v}_j' = [v_{0j}, \delta'_j]$. El estimador de raíz latente es

$$\hat{\beta}_{LR} = \sum_{j=s}^{p} l_j^{-1} \phi_j \delta_j \tag{10.42}$$

$$\phi_j = -SS_{\mathrm{T}}^{1/2} v_{0j} \sum_{q=s}^{p} v_{0q}^2 l_q^{-1} \tag{10.43}$$

Los s términos que corresponden a $j = 0, 1, \ldots, s-1$, eliminados de la ecuación (10.42), corresponden a aquellos eigenvectores para los cuales tanto $|v_{0j}|$ como l_j son casi cero. Así, al igual que en la regresión por componentes principales, la de raíz latente trata de identificar y eliminar las multicolinealidades que no contribuyen a la predicción. La regresión de raíz latente se reduce a mínimos cuadrados cuando no se eliminan términos ($s = 0$).

http://carlos2524.jimdo.com/

Gunst, Webster y Mason [1976] y Gunst y Mason [1977] indican que la regresión de raíz latente puede producir una notable mejoría del error cuadrático medio, en comparación con los mínimos cuadrados. Gunst [1979] señala que la regresión de raíz latente puede producir coeficientes de regresión que se parecen mucho a los que se determinan por componentes principales, en especial cuando sólo hay una o dos multicolinealidades fuertes en **X**. En White y Gunst [1979] se encuentran varias propiedades de la regresión de raíz latente, para muestras grandes.

10.5.5 Comparación y evaluación de los estimadores sesgados

Se han hecho varios estudios de simulación de Monte Carlo para examinar la eficacia de los estimadores sesgados, y para tratar de determinar cuáles procedimientos funcionan mejor. Por ejemplo, véanse McDonald y Galarneau [1975], Hoerl y Kennard [1976], Hoerl, Kennard y Baldwin [1975] (quienes comparan los mínimos cuadrados con ridge), Gunst *et al.* [1976] (raíz latente y mínimos cuadrados), Lawless [1978], Hemmerle y Brantle [1978] (ridge, ridge generalizada y mínimos cuadrados), Lawless y Wang [1976] (mínimos cuadrados, ridge y componentes principales), Wichern y Churchill [1978], Gibbons [1979] (diversas formas de ridge), Gunst y Mason [1977] (ridge, componentes principales, raíz latente y otros) y Dempster *et al.* [1977]. En el estudio de Dempster *et al.* [1977] se comparan 57 estimadores diferentes, para 160 configuraciones distintas de modelo. Si bien no hay un solo procedimiento que de acuerdo con esos estudios sea el mejor en general, hay bastante evidencia que indica la superioridad de la estimación sesgada respecto a los mínimos cuadrados si hay multicolinealidad. Nuestras preferencias particulares, en la práctica, son para la regresión ridge ordinaria seleccionando k por inspección de la traza ridge. El procedimiento es directo, fácil de implementar en un programa usual de cómputo de mínimos cuadrados, y el analista puede aprender con mucha rapidez a interpretar la traza ridge; también a veces es útil para determinar el valor "óptimo" de k, sugerido por Hoerl, Kennard y Baldwin [1975], y la k "óptima" estimada iterativamente de Hoerl y Kennard [1976], y comparar los modelos resultantes con el obtenido mediante la traza ridge.

Como se dijo antes, si el error cuadrático medio se considera función de β, entonces se minimiza cuando β se alinea con el eigenvector estandarizado que corresponde al máximo eigenvalor de $\mathbf{X}'\mathbf{X}$. De igual modo, el error cuadrático medio se maximiza cuando β se alinea con el eigenvector estandarizado que corresponde al mínimo eigenvalor de $\mathbf{X}'\mathbf{X}$. Esto implica que las mejoras potenciales del error cuadrático medio obtenidas con la estimación sesgada dependen de la orientación del vector de parámetros. En su estudio de simulación, Gibbons [1979] informa que si β tiene una alineación favorable, los estimadores del tipo ridge siempre son mejores que los obtenidos por mínimos cuadrados, mientras que si β tiene alineación desfavorable, los estimadores del tipo ridge no siempre son mejores. Así, si el analista tiene alguna información anterior acerca de la alineación de β con los eigenvectores de $\mathbf{X}'\mathbf{X}$, podría tomar una decisión acerca de la utilidad potencial de la estimación sesgada para ese problema en particular, sin embargo, parece no haber una forma sencilla de obtener esa información.

El uso de los estimadores sesgados en la regresión no carece de controversias. Algunos autores critican la regresión ridge y otras técnicas sesgadas de estimación afines. Conniffe y Stone [1973, 1975] criticaron el uso de la traza ridge para seleccionar el parámetro de sesgo, ya que $\hat{\beta}_R$ cambia con lentitud, y al final se estabiliza al aumentar k, aun para los regresores ortogonales. También dicen que si los datos no son adecuados para respaldar un análisis adecuado por mínimos cuadrados, no es probable que ayude algo la regresión ridge, porque los estimados de parámetro no tendrán sentido. Marquardt y Snee [1975] y Smith

http://carlos2524.jimdo.com/

y Goldstein [1975] no aceptan estas conclusiones, y creen que los estimadores sesgados son una valiosa herramienta para el analista que esté confrontado con datos deteriorados. Varios autores notaron que si bien se puede demostrar que existe una k tal que el error cuadrático medio del estimador ridge siempre es menor que el error cuadrático medio del estimador por mínimos cuadrados, no hay seguridad que la traza ridge (o cualquier otro método que seleccione en forma estocástica al parámetro de sesgo, por análisis de los datos) produzca una k óptima.

Draper y Van Nostrand [1977a, b, 1979] también critican los estimadores sesgados. Citan fallas con varios de los detalles técnicos de los estudios de simulación que se usaron como base de las afirmaciones de mejoras en el error cuadrático medio para la estimación sesgada, y sugieren que las simulaciones se diseñaron para favorecer a los estimadores sesgados. Hacen notar que la regresión ridge en realidad sólo es adecuada en casos en que se agrega información externa a un problema de mínimos cuadrados. Eso puede estar en forma de la formulación e interpretación bayesiana del procedimiento, o un problema de mínimos cuadrados restringidos, en el que se escogen las restricciones de β de tal modo que reflejen el conocimiento que tiene el analista de los coeficientes de regresión, para "mejorar el adecuamiento" de los datos.

Smith y Campbell [1980] sugieren el uso del análisis bayesiano explícito, o la estimación mixta, para resolver problemas de multicolinealidad. Rechazan los métodos ridge por débiles e imprecisos, porque sólo incorporan muy holgadamente ideas en información anteriores en el análisis. Cuando se conoce información previa explícita, desde luego se debe aplicar la estimación bayesiana o mixta, sin embargo, con frecuencia la información previa no se reduce con facilidad a una distribución específica *a priori*, y los métodos de regresión de cresta permiten contar con un método para incorporar este conocimiento, al menos en forma aproximada.

Ha habido también algo de controversia acerca de si los regresores y la respuesta se deben centrar y escalar para que $\mathbf{X}'\mathbf{X}$ y $\mathbf{X}'\mathbf{y}$ estén en forma de correlación. Eso da como resultado una eliminación artificial de la ordenada al origen en el modelo. De hecho, la ordenada al origen, en el modelo ridge, se estima con $\bar{y}$. Hoerl y Kennard [1970a, b] usan este método, como Marquardt y Snee [1975], quienes hacen notar que el centrado tiende a minimizar cualquier deterioramiento no esencial al ajustar polinomios. Por otra parte, Brown [1977] cree que no se deben centrar las variables, porque el centrado sólo afecta el estimado de la ordenada al origen y no las pendientes. Belsley, Kuh y Welsch [1980] sugieren no centrar los regresores, para poder diagnosticar el papel de la ordenada al origen en cualquier dependencia casi lineal. El centrado y el escalamiento permiten al analista considerar a los estimados de los parámetros como coeficientes de regresión estandarizados, lo que con frecuencia es atractivo, intuitivamente, además, al centrar los regresores se puede eliminar el deterioramiento no esencial, reduciendo así la inflación de la varianza en los estimados de parámetros. En consecuencia, nosotros recomendamos tanto el centrado como el escalamiento de los datos.

A pesar de las objeciones citadas, creemos que los métodos sesgados de estimación son técnicas útiles que debe tener en cuenta el analista al manejar la multicolinealidad. Los métodos sesgados de estimación sí se comparan muy favorablemente con otros métodos de manejo de multicolinealidad, como la eliminación de variables. Como dicen Marquardt y Snee [1975], con frecuencia es mejor usar algo de la información de todos los regresores, como lo hace la regresión ridge, que usar toda la información de algunos de los regresores, y nada de información de otros, como hace la eliminación de variables. Además, se puede concebir la eliminación de variables como una forma de estimación sesgada porque, como se dijo en el capítulo 9, con frecuencia los modelos de regresión de subconjunto de variables producen estimados sesgados de los coeficientes de regresión. De hecho, la eliminación de

http://carlos2524.jimdo.com/

variables contrae con frecuencia el vector de estimados de los parámetros, como la regresión ridge. No recomendamos el uso automático o mecánico de la regresión ridge sin un estudio concienzudo de los datos, y un análisis cuidadoso de la adecuación del modelo final. Cuando se usan en forma adecuada, los métodos de estimación sesgada son herramientas valiosas entre los recursos del analista de datos.

PROBLEMAS

10.1 Para los datos de tiempo de entrega de cajas de gaseosas, del ejemplo 3.1.
- **a.** Determinar la correlación simple entre cajas (x_1), y distancia (x_2).
- **b.** Calcular los factores de inflación de varianza.
- **c.** Calcular el número de condición de $\mathbf{X}'\mathbf{X}$. ¿Hay evidencia de multicolinealidad en esos datos?

10.2 Con los datos del cemento de Hald, del ejemplo 9.1:
- **a.** A partir de la matriz de correlaciones entre los regresores, ¿cabe sospechar que hay multicolinealidad?
- **b.** Calcular los factores de inflación de varianza.
- **c.** Determinar los eigenvalores de $\mathbf{X}'\mathbf{X}$.
- **d.** Calcular el número de condición de $\mathbf{X}'\mathbf{X}$.

10.3 Usar los datos del cemento de Hald (Ej. 9.1) para determinar el eigenvector asociado con el eigenvalor mínimo de $\mathbf{X}'\mathbf{X}$. Interpretar los elementos de ese vector. ¿Qué se puede decir acerca de la fuente de la multicolinealidad en estos datos?

10.4 Determine los índices de condición y las proporciones de descomposición de varianza para los datos de cemento de Hald (Ej. 9.1), suponiendo que los regresores son centrados. ¿Qué se puede decir de la multicolinealidad en esos datos?

10.5 Repetir el problema 10.4 sin centrar los regresores, y comparar los resultados. ¿Cuál procedimiento cree el lector que es mejor?

10.6 Usar los regresores x_2 (yardas por aire), x_7 (porcentaje de jugadas por tierra) y x_8 (yardas del contrario por aire) de los datos de la NFL (National Futbol League), en la tabla B.1 del apéndice.
- **a.** La matriz de correlación, ¿muestra algún indicio de multicolinealidad?
- **b.** Calcular los factores de inflación de variable y el número de condición de $\mathbf{X}'\mathbf{X}$. ¿Hay alguna evidencia de multicolinealidad?

10.7 Para los datos de rendimiento de gasolina, en la tabla B.3 del apéndice.
- **a.** La matriz de correlación ¿muestra alguna indicación de multicolinealidad?
- **b.** Calcular los factores de inflación de varianza y el número de condición de $\mathbf{X}'\mathbf{X}$. ¿Hay alguna evidencia de multicolinealidad?

10.8 Usar los datos de rendimiento de gasolina en la tabla B.3 del apéndice para calcular los eigenvectores asociados con los eigenvalores mínimos de $\mathbf{X}'\mathbf{X}$. Interpretar los elementos de esos vectores. ¿Qué se puede decir acerca de la fuente de multicolinealidad en estos datos?

10.9 Usar los datos de rendimiento de gasolina en la tabla B.3 del apéndice, y calcular los índices de condición y las proporciones de descomposición de varianza, con los regresores centrados. ¿Qué se puede afirmar acerca de la multicolinealidad en esos datos?

http://carlos2524.jimdo.com/

10.10 Analizar los datos de precios de viviendas en la tabla B.4 del apéndice, para ver si hay multicolinealidad. Usar los factores de inflación de varianza y el número de condición de $\mathbf{X}'\mathbf{X}$.

10.11 Analizar los datos del proceso químico en la tabla B.5 del apéndice, para ver si hay evidencia de multicolinealidad. Usar los factores de inflación de varianza y el número de condición de $\mathbf{X}'\mathbf{X}$.

10.12 La tabla siguiente muestra los índices de condición y las proporciones de descomposición de varianza para los datos del acetileno, usando regresores centrados. Con esta información, diagnosticar la multicolinealidad en los datos y sacar las conclusiones adecuadas.

			Proporciones de descomposición de varianza								
Número	Eigenvalor	Índices de condición	T	H	C	TH	TC	HC	T2	H2	C2
1	4.204797	1.000000	0.0001	0.0024	0.0001	0.0004	0.0000	0.0004	0.0000	0.0001	0.0000
2	2.162611	1.394387	0.0000	0.0305	0.0000	0.0044	0.0000	0.0035	0.0000	0.0412	0.0000
3	1.138392	1.921882	0.0010	0.0000	0.0003	0.0002	0.0000	0.0001	0.0001	0.0139	0.0000
4	1.041305	2.009480	0.0000	0.2888	0.0000	0.0040	0.0000	0.0032	0.0000	0.0354	0.0000
5	0.384532	3.306788	0.0001	0.5090	0.0000	0.0023	0.0000	0.0029	0.0000	0.4425	0.0000
6	0.049510	9.215620	0.0034	0.0049	0.0000	0.0874	0.0000	0.0565	0.0033	0.0319	0.0071
7	0.013633	17.562062	0.0096	0.0051	0.0031	0.8218	0.0000	0.7922	0.0042	0.0001	0.0053
8	0.0051232	28.648589	0.1514	0.0936	0.1461	0.0773	0.0007	0.1210	0.0002	0.3526	0.0229
9	0.0000969	208.285	0.8343	0.0657	0.8504	0.0022	0.9993	0.0201	0.9920	0.0822	0.9646

10.13 Aplicar la regresión ridge a los datos del cemento de Hald, en el ejemplo 9.1.

a. Usar la traza ridge para seleccionar un valor adecuado de k. El modelo final ¿es bueno?

b. ¿Cuánta inflación ha resultado en la suma residual de cuadrados por el uso de la regresión ridge?

c. Comparar el modelo de regresión ridge con el de dos regresores, donde intervengan x_1 y x_2, desarrollado por el método de todas las regresiones posibles en el ejemplo 9.1.

10.14 Aplicar la regresión ridge a los datos del cemento de Hald (Ej. 9.1) usando el valor de k de la ecuación (10.18). Comparar este valor de k con el que se seleccionó con la traza ridge, en el problema 10.13. ¿Difiere mucho el modelo final del que hay en el problema 10.13?

10.15 Usar el procedimiento iterativo de estimación que se describe en la sección 10.5.3 para estimar k en la regresión ridge para los datos del cemento de Hald (Ej. 9.1). Comparar el valor de k obtenido con el procedimiento con los obtenidos con la traza ridge (Prob.10.13) y la ecuación (10.18) (Prob. 10.14). ¿Difiere mucho el modelo final de los que se determinaron en los problemas 10.13 y 10.14?

10.16 Estimar los parámetros, usando regresión ridge, en un modelo para los datos de rendimiento de gasolina en la tabla B.3 del apéndice.

a. Usar la traza ridge para seleccionar un valor adecuado de k. ¿Es adecuado el modelo que resulta?

b. ¿Cuánta inflación ha causado el uso de la regresión ridge en la suma de cuadrados residuales?

c. ¿Cuánta reducción en R^2 ha resultado por el uso de la regresión ridge?

10.17 Estimar los parámetros de un modelo para los datos de rendimiento de gasolina en la tabla B.3 del apéndice, con regresión ridge y el valor de k determinado con la ecuación (10.18). ¿Difiere mucho este modelo del desarrollado en el problema 10.16?

http://carlos2524.jimdo.com/

10.18 Estimar los parámetros en un modelo de los datos de rendimiento de gasolina, en la tabla B.3 del apéndice, aplicando la regresión ridge y k determinada con el procedimiento iterativo que se describió en la sección 10.5.3. Comparar el modelo obtenido con el que se obtuvo en el problema 10.16.

10.19 Estimar los parámetros del modelo para los datos del cemento de Hald (Ej. 9.1) usando regresión por componentes principales.

a. ¿Cuál es la baja de R^2 para este modelo, en comparación con mínimos cuadrados?

b. ¿Cuánta contracción del vector de coeficientes se obtuvo?

c. Comparar el modelo de componentes principales con el ridge ordinaria, obtenido en el problema 10.13. Comentar todas las diferencias aparentes en los modelos.

10.20 Estimar los parámetros de modelo para los datos de rendimiento de gasolina, aplicando la regresión por componentes principales.

a. ¿Cuánto aumentó la suma de cuadrados residuales en comparación con mínimos cuadrados?

b. ¿Cuánta contracción se produjo en el vector de coeficientes?

c. Comparar los modelos por componentes principales y por ridge ordinaria (Prob. 10.16). ¿Cuál modelo prefiere el lector?

10.21 Demostrar que el estimador ridge es la solución del problema

$$\underset{\boldsymbol{\beta}}{\text{Minimizar}} \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)' \mathbf{X}'\mathbf{X}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)$$

$$\text{sujeta a} \quad \boldsymbol{\beta}'\boldsymbol{\beta} \leq d^2$$

10.22 **Estimadores de contracción pura (Stein [1960]).** El estimador de contracción pura se define como $\hat{\beta}_s = c\,\hat{\beta}$, siendo $0 \leq c \leq 1$ una constante elegida por el analista. Describir la clase de contracción que introduce este estimador, y compararla con la contracción debida a la regresión ridge. En forma intuitiva, ¿cuál estimador parece preferible?

10.23 Demostrar que el estimador de contracción pura (Prob. 10.22) es la solución de

$$\underset{\boldsymbol{\beta}}{\text{Minimizar}} \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)' \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)$$

$$\text{sujeta a} \quad \boldsymbol{\beta}'\boldsymbol{\beta} \leq d^2$$

10.24 El criterio del error cuadrático medio para la regresión ridge es

$$E\left(L_1^2\right) = \sum_{j=1}^{p} \frac{\lambda_j}{\left(\lambda_j + k\right)^2} + \sum_{j=1}^{p} \frac{\alpha_j^2 k^2}{\left(\lambda_j + k\right)^2}$$

Tratar de determinar el valor de k que minimiza a $E(L_1^2)$. ¿Qué dificultades se encuentran?

10.25 Acerca del criterio del error cuadrático medio para la regresión ridge generalizada, ecuación (10.30). Demostrar que el error cuadrático medio se minimiza si $k_j = \sigma^2/\alpha_j^2$, $j = 1, 2, \ldots, p$.

10.26 **Regresión ridge dirigida**. Supóngase que en lugar de contraer todos los elementos del vector de parámetros, sólo se contraen los coeficientes que corresponden a eigenvalores pequeños. Si se hace la contracción usando ridge generalizada, eso se llama procedimiento

http://carlos2524.jimdo.com/

ridge dirigido, según Guilkey y Murphy [1975]. Sclove [1968] sugirió una técnica semejante, pero contrae el subconjunto adecuado uniformemente hacia el origen. ¿Qué ventajas tendrían esos métodos en comparación con la regresión ridge ordinaria? ¿Qué tipo de estrategia de contracción recomienda el lector?

10.27 Demostrar que si $\mathbf{X}'\mathbf{X}$ está en forma de matriz de correlación, si Λ es la matriz diagonal de los eigenvalores de $\mathbf{X}'\mathbf{X}$, y si $\mathbf{T}$ es la matriz correspondiente de eigenvectores, entonces los factores de inflación de varianza son los elementos de la diagonal principal de $\mathbf{T}\Lambda^{-1}\mathbf{T}'$.

http://carlos2524.jimdo.com/

CAPÍTULO 11

REGRESIÓN ROBUSTA

11.1 LA NECESIDAD DE UNA REGRESIÓN ROBUSTA

Cuando las observaciones **y** en el modelo de regresión lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ están normalmente distribuidas, el método de mínimos cuadrados es un buen procedimiento de estimación de parámetros, porque produce un estimador del vector β de parámetros que tiene buenas propiedades estadísticas, sin embargo, hay muchos casos en los que hay evidencias de que la distribución de la variable de respuesta tiene una distribución (considerablemente) no normal y/o hay valores atípicos que afectan al modelo de regresión. Un caso de mucho interés práctico es aquél en el que las observaciones tienen una distribución que tiene colas más largas o gruesas que la distribución normal. Esas distribuciones tienden a generar valores atípicos, que pueden tener una gran influencia sobre el método de los mínimos cuadrados, al "jalar" demasiado la ecuación de regresión en su dirección.

Por ejemplo, examínense las 10 observaciones de la figura 11.1. El punto identificado con A en esa figura está en el extremo derecho del espacio de x, pero su valor de respuesta queda cerca del promedio de las otras 9 respuestas. Si se tienen en cuenta todas las observaciones, el modelo de regresión que resulta es $\hat{y} = 2.12 + 0.971x$, y $R^2 = 0.526$, no obstante, si se ajusta el modelo de regresión lineal con todas las observaciones **excepto** A, se obtiene $\hat{y} = 0.715 + 1.45x$, para el cual $R^2 = 0.894$. Ambas rectas se ven en la figura 11.1. Es claro que el punto A tuvo un efecto radical sobre el modelo de regresión, y sobre el valor obtenido de R^2.

Una forma de manejar esta situación es eliminar la observación A, así se obtiene una recta que pasa muy bien por el resto de los datos, más agradable desde un punto de vista estadístico, sin embargo, lo que se está haciendo ahora es descartar observaciones tan solo porque es agradable desde un punto de vista de modelado estadístico, y por lo general esa práctica no es buena. A veces, los datos se pueden eliminar (o modificar) con base en el conocimiento **de la materia**, pero cuando se hace eso con una base puramente estadística, en general se mete uno en problemas; también se observa que en casos más complicados, donde intervienen muchos regresores y la muestra es mayor, puede dificultarse identificar que el modelo de regresión se ha distorsionado por observaciones como la A.

Un **procedimiento de regresión robusta** es aquel que amortigua el efecto de las observaciones que serían muy influyentes si se usaran los mínimos cuadrados, lo que nos indica que un procedimiento robusto tiende a dejar grandes los residuales asociados con valores atípicos, facilitando así la identificación de puntos influyentes. Además de la insensibilidad a los valores atípicos, un procedimiento de estimación robusta debería producir, en esencia, los mismos resultados que los mínimos cuadrados cuando la distribución básica es normal, y cuando no hay valores atípicos. Otro objetivo deseable de la regresión robusta es que los procedimientos de estimación y los de referencia sean relativamente fáciles de llevar a cabo.

http://carlos2524.jimdo.com/

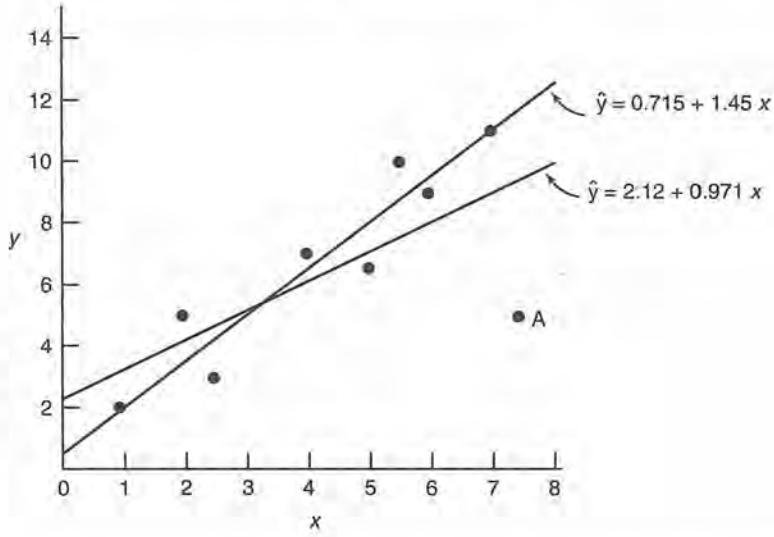


Figura 11.1 Diagrama de dispersión de una muestra que contiene una observación influyente.

El promotor de muchos de los trabajos sobre regresión robusta fue el estudio de robustez en Princeton (véase Andrews *et al.* [1972]). Después se han propuesto varios tipos de estimadores robustos. Entre las referencias importantes y básicas están Andrews [1974], Carroll y Ruppert [1988], Hogg [1974, 1979a,b], Huber [1972, 1973, 1981], Krasker y Welsch [1982], Rousseeuw [1984, 1998] y Rousseeuw y Leroy [1987].

Los métodos de regresión robusta son técnicas que en potencia se pueden usar cuando hay valores atípicos. Hay varias clases de valores atípicos que se presentan en el contexto de modelos de regresión. Una clasificación frecuente es la siguiente:

1. **Valor atípico de regresión.** Es un punto que se desvía de la regresión lineal que se determina con las $n - 1$ observaciones restantes.
2. **Valor atípico residual.** Es un punto que tiene un residual estandarizado o estudentizado grande, cuando se usa en la muestra de n observaciones con que se ajusta un modelo, nótese que un punto puede ser un valor atípico residual sin que haya fuerte indicación que sea un valor atípico de regresión, o que un punto sea atípico de regresión sin que sea atípico residual (esto sucede con frecuencia cuando el punto es muy influyente). El punto A de la figura 1.1 es valor atípico de regresión y atípico residual al mismo tiempo, porque está fuera de la línea de ajuste de las otras 9 observaciones, y con un residual estandarizado mayor que 2.
3. **Valor atípico en el espacio *X*.** Es una observación remota en una o más coordenadas x. Un valor atípico en el espacio x también puede ser un atípico de regresión y/o atípico residual. Algunas técnicas de regresión robusta son perjudicadas por valores atípicos en el espacio x, como se describirá más adelante.
4. **Valor atípico en el espacio *Y*.** Es una observación con coordenada y inusual. El efecto que tiene la observación sobre el modelo de regresión depende de su coordenada x, y de la disposición general de las demás observaciones en la muestra. Un valor atípico en el espacio y puede ser atípico residual y/o posiblemente uno atípico de regresión.
5. **Valores atípicos en los espacios *X* y *Y*.** Es una observación que es atípica en sus coordenadas x y y al mismo tiempo. Ese punto podrá no tener efecto sobre la línea

http://carlos2524.jimdo.com/

de regresión, o podrá ser un valor atípico residual, atípico de regresión o posiblemente ambas cosas. El efecto de esos puntos depende por completo de la disposición de las demás observaciones de la muestra.

La observación identificada con A en la figura 11.2 ilustra un caso en donde en la muestra hay un valor atípico en el espacio x y en y. Este punto no tiene de hecho impacto sobre la línea de regresión; tan sólo es un punto de **gran influencia**. Algunas personas podrían pensar que ese punto es de gran influencia "buena", porque está aproximadamente sobre la recta que pasa por las demás observaciones, y tiene el efecto de aumentar el valor de R^2 y bajar el error estándar de la pendiente, por otra parte, podría no ser aconsejable pasar por alto la presencia de esas observaciones aparentemente atípicas en la muestra. Nótese que si se cambia la coordenada x de este punto, para que quede cerca del recorrido de los demás valores de x, como se ve en la figura 11.2, la observación sería un valor atípico de espacio y, atípico residual y atípico de regresión. Si sólo se cambia la coordenada y para hacerla más o menos igual al promedio de las demás respuestas, como se ve en la figura 11.2, se obtendría un valor atípico de espacio x, atípico residual y atípico de regresión.

Para motivar algo de la siguiente discusión, y para seguir demostrando por qué puede ser bueno usar una alternativa de los mínimos cuadrados cuando las observaciones no tienen distribución normal, considérese el modelo de regresión lineal simple

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{11.1}$$

en donde los errores son variables aleatorias independientes que tienen la **distribución doble exponencial**

$$f(\varepsilon_i) = \frac{1}{2\sigma} e^{-|\varepsilon_i|/\sigma}, \qquad -\infty < \varepsilon_i < \infty \tag{11.2}$$

La distribución doble exponencial se ve en la figura 11.3. Es más "picuda" en el centro que la normal, y las colas van a cero cuando $|\varepsilon_i|$ tiende a infinito, sin embargo, como la función de densidad tiende a cero cuando $e^{-|\varepsilon_i|}$ tiende a cero, y la función de densidad normal tiende

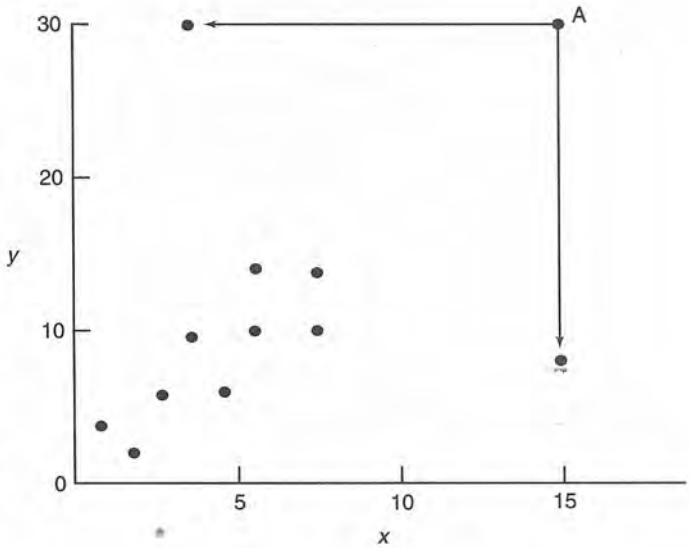


Figura 11.2 Diagrama de dispersión que contiene un valor atípico en el espacio x y uno en el espacio y.

http://carlos2524.jimdo.com/

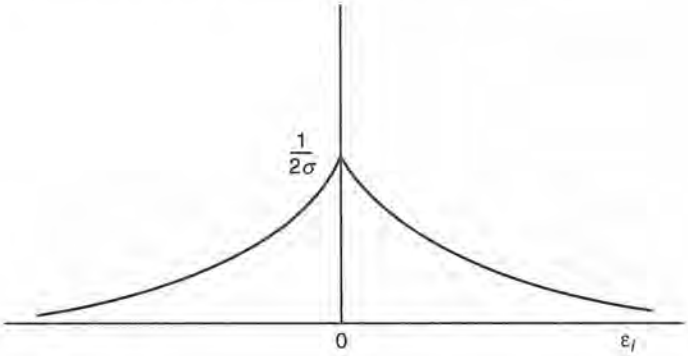


Figura 11.3 La distribución doble exponencial.

a cero cuando $e^{-\varepsilon_i^2}$ tiende a cero, se ve que la distribución doble exponencial tiene **colas más pesadas** que la normal.

Para estimar β_0 y β_1 se usará el método de **máxima verosimilitud**. Esta función es

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} \frac{1}{2\sigma} e^{-|\varepsilon_i|/\sigma} = \frac{1}{(2\sigma)^n} \exp\left(- \sum_{i=1}^{n} |\varepsilon_i|/\sigma \right) \tag{11.3}$$

Por consiguiente, al maximizar la función de verosimilitud se minimiza $\Sigma_{i=1}^{n}|\varepsilon_i|$, la suma de los errores absolutos. Recuérdese que el método de máxima verosimilitud aplicado al modelo de regresión, con errores de distribución normal, conduce al criterio de los mínimos cuadrados, así, la hipótesis de una distribución de error con colas más gruesas que las de la normal implica que el método de los mínimos cuadrados ya no es una técnica óptima de estimación. Nótese que el criterio del error absoluto ponderaría los valores atípicos con mucho menos severidad que los mínimos cuadrados. La minimización de la suma de errores absolutos se llama con frecuencia problema de regresión L_1-norm (mínimos cuadrados es el problema de regresión L_2-norm). Este criterio lo sugirió primero F. Y. Edgeworth en 1887, quien dijo que los mínimos cuadrados están demasiado influidos por grandes valores atípicos. Una forma de resolver el problema es a través de un método de programación lineal. Para conocer más detalles de la regresión L_1-norm véase Sielken y Hartley [1973], Book *et al.* [1980], Gentle, Kennedy y Sposito [1977], Bloomfield y Steiger [1983] y Dodge [1987].

El problema de regresión L_1-norm es un caso especial de la regresión L_p-norm, en el que los parámetros del modelo se escogen para minimizar $\Sigma_{i=1}^{n}|\varepsilon_i|^p$, donde $1 \le p \le 2$. Cuando $1 < p < 2$, se puede formular y resolver el modelo con técnicas de programación no lineal. Forsythe [1972] estudió en forma extensa este procedimiento, para el modelo de regresión lineal simple.

11.2 ESTIMADORES *M*

El problema de regresión L_1-norm surge en forma natural a partir del método de máxima verosimilitud, con errores doblemente exponenciales. En general, se puede definir una **clase de estimadores robustos** que minimizan una función ρ de los residuales, por ejemplo

$$\underset{\beta}{\text{Minimizar}} \sum_{i=1}^{n} p(e_i) = \underset{\beta}{\text{Minimizar}} \sum_{i=1}^{n} p(y_i - \mathbf{x}_i'\beta) \tag{11.4}$$

http://carlos2524.jimdo.com/

en donde $\mathbf{x}_i'$ representa el i-ésimo renglón de $\mathbf{X}$. Un estimador de este tipo se llama **estimador M**, donde M quiere decir **máxima verosimilitud**. Esto es, la función ρ se relaciona con la función de verosimilitud para una elección adecuada de la distribución del error. Por ejemplo, si se usa el método de mínimos cuadrados (que implica que la distribución de errores es normal), entonces $\rho(z) = \frac{1}{2}z^2 - \infty < z < \infty$.

El estimador M no necesariamente es invariante de escala [es decir, si se multiplicaran los errores $y_i - \mathbf{x}_i'\boldsymbol{\beta}$ por una constante, la nueva solución de la ecuación (11.4) podría no ser igual que la anterior]. Para obtener una versión invariante de escala para este estimador, se resuelve

$$\underset{\boldsymbol{\beta}}{\text{Minimizar}} \sum_{i=1}^{n} \rho\left(\frac{\bar{e}_i}{s}\right) = \underset{\boldsymbol{\beta}}{\text{Minimizar}} \sum_{i=1}^{n} \rho\left(\frac{\bar{y}_i - \mathbf{x}_i'\boldsymbol{\beta}}{s}\right) \tag{11.5}$$

en donde s es un estimado robusto de escala. Una elección muy frecuente de s es la mediana de la desviación absoluta

$$s = \text{mediana}|e_i - \text{mediana}(e_i)|/0.6745 \tag{11.6}$$

La constante de ajuste 0.6745 hace que s sea aproximadamente un estimador insesgado de σ si n es grande y la distribución de errores es normal.

Para minimizar la ecuación (11.5), se igualan a cero las primeras derivadas parciales de ρ con respecto a $\beta_j (j = 0, 1, \ldots, k)$, y se obtiene una condición necesaria para un mínimo. De este modo se obtiene el sistema de $p = k + 1$ ecuaciones

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{s}\right) = 0, \qquad j = 0, 1, \ldots, k \tag{11.7}$$

en donde $\psi = \rho'$ y x_{ij} es la i-ésima observación del j-ésimo regresor, y $x_{i0} = 1$. En general, la función ψ es no lineal, y se debe resolver la ecuación (11.7) por métodos iterativos. Si bien se podrían emplear varias técnicas no lineales de optimización, lo que más se usa son los **mínimos cuadrados reponderados iterativamente**. Este método se suele atribuir a Beaton y Tukey [1974].

Para aplicar los mínimos cuadrados reponderados iterativamente se supone un estimado inicial $\hat{\beta}_0$ y que s es un estimado de escala. A continuación se escriben las $p = k + 1$ ecuaciones (11.7):

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{s}\right) = \sum_{i=1}^{n} \frac{x_{ij}\{\psi[(y_i - \mathbf{x}_i'\boldsymbol{\beta})/s]/(y_i - \mathbf{x}_i'\boldsymbol{\beta})/s\}(y_i - \mathbf{x}_i'\boldsymbol{\beta})}{s} = 0, \qquad j = 0, 1, \ldots, k \tag{11.8}$$

en la forma

$$\sum_{i=1}^{n} x_{ij}w_{i0}(y_i - \mathbf{x}_i'\boldsymbol{\beta}) = 0, \qquad j = 0, 1, \ldots, k \tag{11.9}$$

http://carlos2524.jimdo.com/

TABLA 11.1 Funciones de criterio robusto

Criterio	$\rho(z)$	$\psi(z)$	$w(z)$	Intervalo
Mínimos cuadrados	$\frac{1}{2}z^2$	z	1.0	$\lvert z\rvert < \infty$
Función t de Huber	$\frac{1}{2}z^2$	z	1.0	$\lvert z\rvert \le t$
$t = 2$	$\lvert z\rvert t - \frac{1}{2}t^2$	$t\,\text{signo}(z)$	$\dfrac{t}{\lvert z\rvert}$	$\lvert z\rvert > t$
Función E_a de Ramsay	$a^{-2}[1 - \exp(-a\lvert z\rvert)\cdot(1 + a\lvert z\rvert)]$	$z\exp(-a\lvert z\rvert)$	$\exp(-a\lvert z\rvert)$	$\lvert z\rvert < \infty$
$a = 0.3$				
Función de onda de Andrews	$a[1 - \cos(z/a)]$	$\text{sen}(z/a)$	$\dfrac{\text{sen}(z/a)}{z/a}$	$\lvert z\rvert \le a\pi$
$a = 1.339$	$2a$	0	0	$\lvert z\rvert > a\pi$
Función 17A de Hampel	$\frac{1}{2}z^2$	z	1.0	$\lvert z\rvert \le a$
$a = 1.7$				
$b = 3.4$	$a\lvert z\rvert - \frac{1}{2}a^2$	$a\,\text{signo}(z)$	$a/\lvert z\rvert$	$a < \lvert z\rvert \le b$
$c = 8.5$				
	$\dfrac{a\left(c\lvert z\rvert - \frac{1}{2}z^2\right)}{c-b} - (7/6)a^2$	$\dfrac{a\,\text{signo}(z)(c - \lvert z\rvert)}{c-b}$	$\dfrac{a(c - \lvert z\rvert)}{\lvert z\rvert(c-b)}$	$b < \lvert z\rvert \le c$
	$a(b + c - a)$	0	0	$\lvert z\rvert > c$

en donde

$$w_{i0} = \begin{cases} \dfrac{\psi\left[\left(y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}}_0\right)/s\right]}{\left(y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}}_0\right)/s} & \text{si } y_i \ne \mathbf{x}'_i\hat{\boldsymbol{\beta}}_0 \\ 1 & \text{si } y_i = \mathbf{x}'_i\hat{\boldsymbol{\beta}}_0 \end{cases} \tag{11.10}$$

En notación matricial, la ecuación (11.9) es

$$\mathbf{X}'\mathbf{W}_0\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{W}_0\mathbf{y} \tag{11.11}$$

en donde $\mathbf{W}_0$ es una matriz diagonal de $n \times n$ de "pesos" con elementos diagonales w_{10}, $w_{20}, \ldots, w_{n0}$ obtenidos por la ecuación (11.10). Se reconoce que la ecuación (11.11) representa las ecuaciones normales de mínimos cuadrados ponderados. En consecuencia, el **estimador de una etapa** es

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}'\mathbf{W}_0\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_0\mathbf{y} \tag{11.12}$$

En el siguiente paso se recalculan los factores de ponderación o pesos con la ecuación (11.10), pero con $\hat{\boldsymbol{\beta}}_1$ en lugar de $\hat{\boldsymbol{\beta}}_0$. Por lo general sólo se requieren unas pocas iteraciones para alcanzar la convergencia. El procedimiento de mínimos cuadrados iterativamente reponderados se podría implementar con un programa computacional estándar de mínimos cuadrados ponderados.

En la tabla 11.1 se presentan varias funciones de criterio robusto de uso frecuente. El comportamiento de esas funciones ρ y de sus funciones ψ correspondientes se ilustra en

http://carlos2524.jimdo.com/

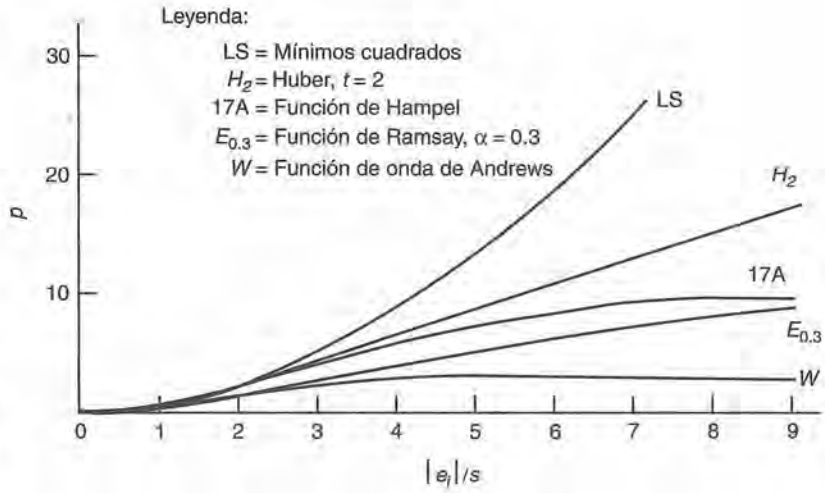


Figura 11.4 Funciones de criterio robusto.

las figuras 11.4 y 11.5, respectivamente. Los procedimientos de regresión robusta se pueden clasificar de acuerdo con el comportamiento de su función ψ. Esta función ψ controla el factor de ponderación que se asigna a cada residual y (además de una constante de proporcionalidad), a veces se llama **función de influencia**. Por ejemplo, la función ψ para mínimos cuadrados no es acotada, por lo que los mínimos cuadrados tienden a ser no robustos cuando se usan con datos procedentes de una distribución con colas gruesas. La función t de Huber (Huber [1964]) tiene una función ψ **monótona** y no pondera residuales grandes con tanta intensidad como los mínimos cuadrados. Las tres últimas funciones de influencia en realidad **redescienden** a medida que el residual se hace más grande. La función E_a de Ramsay (véase Ramsay [1977]) es **redescendente suave**, esto es, la función ψ es asintótica a cero para $|z|$ grande. La función de onda de Andrews y la función 17A de Hampel (véase Andrews *et al.* [1972] y Andrews [1974]) son **redescendente dura**, que quiere decir que la función ψ es igual a cero cuando $|z|$ es suficientemente grande. Se debe hacer notar que las funciones ρ asociadas con las funciones ψ redescendentes son no convexas, y en teoría eso puede causar problemas de convergencia en el procedimiento iterativo de estimación, sin embargo, eso no sucede con frecuencia, además, cada una de las funciones de criterio robusto requiere que el analista especifique ciertas "constantes de ajuste" para las funciones ψ. En la tabla 11.1 se han mostrado los valores característicos de esas constantes de ajuste.

El valor de arranque $\hat{\beta}_0$ que se usa en la estimación robusta puede ser una consideración importante. El uso de la solución por mínimos cuadrados puede disfrazar los puntos de gran influencia. Los estimados L_1-norm podrían ser una elección posible de valores de arranque, también, Andrews [1974] y Dutter [1977] sugieren procedimientos para elegir los valores de arranque.

Es importante conocer algo acerca de la estructura de los errores en los estimados $\hat{\beta}$ de la regresión robusta final. Es importante determinar la matriz de covarianza de $\hat{\beta}$ si se deben establecer intervalos de confianza, o si se van a hacer otras inferencias, Huber [1973] demostró que, asintóticamente, $\hat{\beta}$ tiene una distribución aproximadamente normal, y su matriz de covarianza es

$$\sigma^2 \frac{E\left[\psi^2(\varepsilon/\sigma)\right]}{\left\{E\{\psi'(\varepsilon/\sigma)\}\right\}^2}(\mathbf{X}'\mathbf{X})^{-1}$$

http://carlos2524.jimdo.com/

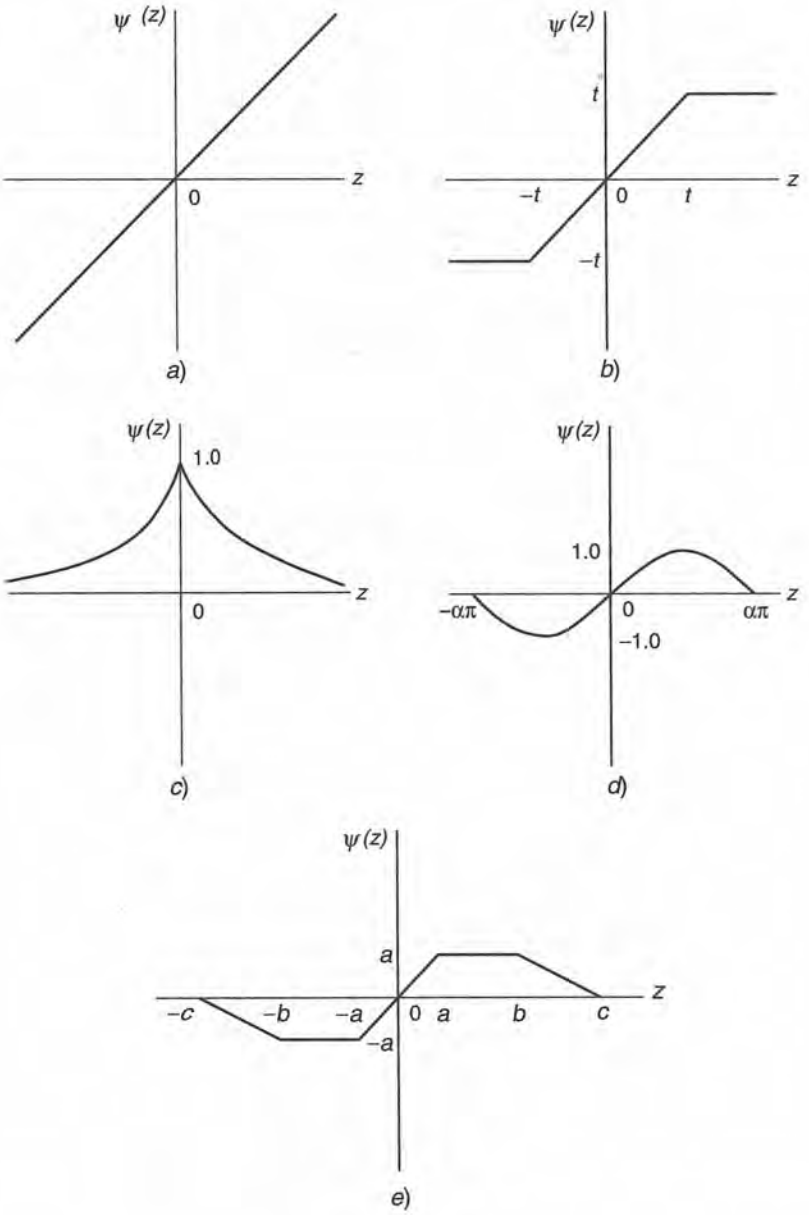


Figura 11.5 Funciones robustas de influencia: a) mínimos cuadrados; b) funciones t de Huber; c) función E_a de Ramsay; d) función de onda de Andrews; e) función 17A de Hampel.

Por lo anterior, una aproximación razonable para la matriz de covarianza de $\hat{\beta}$ es

$$\frac{(ns^2)}{n-p}\,\frac{\sum_{i=1}^{n}\psi^2[(y_i-\mathbf{x}_i'\boldsymbol{\beta})/s]}{\left\{\sum_{i=1}^{n}\psi'[(y_i-\mathbf{x}_i'\boldsymbol{\beta})/s]\right\}^2}(\mathbf{X}'\mathbf{X})^{-1}$$

http://carlos2524.jimdo.com/

También, el programa de cómputo para mínimos cuadrados ponderados obtiene un estimado de la matriz de covarianza

$$\frac{\sum_{i=1}^{n} w_i\left(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)^2}{n-p}\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}$$

En Welsch [1975] y en Hill [1979] se presentan otras sugerencias. No hay acuerdo generalizado acerca de cuál aproximación a la matriz de covarianza de $\hat{\boldsymbol{\beta}}$ es la mejor. Tanto Welsch como Hill hacen notar que esos estimados de matriz de covarianza funcionan mal con matrices $\mathbf{X}$ que tienen puntos atípicos. También el deterioramiento (multicolinealidad) distorsiona los estimados de regresión robusta, sin embargo, hay indicaciones de que en muchos casos se pueden hacer inferencias aproximadas acerca de $\hat{\boldsymbol{\beta}}$ mediante procedimientos parecidos a los de la teoría normal acostumbrada.

Ejemplo 11.1 *Datos del tiempo de entrega*

Se examinará el ejemplo 3.1 en el que se desarrolló un modelo por mínimos cuadrados, que relaciona al tiempo de entrega con dos regresores: cajas (x_1) y distancia (x_2). Los estadísticos de resumen del modelo no revelan algo desacostumbrado en este ajuste, sin embargo, una gráfica de probabilidad normal de los residuales (Fig. 4.2) indicó que es dudosa la suposición de normalidad, y que es posible que los errores pertenezcan a una distribución con colas gruesas. Al profundizar el análisis se vio que hay dos observaciones relativamente influyentes, los puntos 9 y 22.

Para ilustrar los métodos de regresión robusta se obtuvieron ajustes robustos para estos datos, usando la función t de Huber, la exponencial de Ramsay, la de onda de Andrews y la 17A de Hampel. Estos ajustes, junto con el análisis por mínimos cuadrados, se resumen en las tablas 11.2 a 11.6. Todos los ajustes robustos se obtuvieron usando mínimos cuadrados reponderados iterativamente, comenzando con la solución de mínimos cuadrados ordinarios. Como la mediana de los residuales por mínimos cuadrados es 0.436360, el estimado robusto de la escala empleada es

$$s = \text{mediana}|e_i| - 0.436360|/0.6745 = 1.627$$

Este parámetro se mantuvo constante en cada iteración. El criterio de convergencia para los estimados robustos fue detener el proceso de iteración cuando el cambio máximo en cualquiera de los coeficientes fuera menor que 0.1 por ciento.

Las tablas 11.2 a 11.6 muestran los valores reales y ajustados, los residuales, los pesos asignados a cada residual y los estimados de parámetro para cada método de estimación. Obsérvese que en los mínimos cuadrados ponderan por igual a todos los residuales (factor de ponderación = 1.0), aun los valores extremos. El procedimiento t de Huber, resumido en la tabla 11.3, subestima seis de los residuales (puntos 1, 4, 9, 20, 23 y 24). El peso asignado a la observación 9 sólo es 0.327, su efecto es reducir los estimados de β_1 y β_2 y aumentar el de β_0, en comparación con los mínimos cuadrados. La función E_a de Ramsay (Tabla 11.4), redescendiente suave, rebaja en forma más enérgica la ponderación de los residuales. El peso asignado a la observación 9 es ahora 0.132, y los otros 5 puntos de residuales (1, 4, 20, 23 y 24) tienen factores de ponderación menores que 0.5; en consecuencia, se

http://carlos2524.jimdo.com/

TABLA 11.2 Ajuste por mínimos cuadrados para los datos del tiempo de entrega

Observación i	y_i	$\hat{y}_i$	e_i	Peso
1	.166800E + 02	.217081E + 02	− .502808E + 01	.100000E + 01
2	.115000E + 02	.103536E + 02	.114639E + 01	.100000E + 01
3	.120300E + 02	.120798E + 02	− .497937E + 01	.100000E + 01
4	.148800E + 02	.995565E + 01	.492435E + 01	.100000E + 01
5	.137500E + 02	.141944E + 02	− .444398E + 00	.100000E + 01
6	.181100E + 02	.183996E + 02	− .289574E + 00	.100000E + 01
7	.800000E + 01	.715538E + 01	.844624E + 00	.100000E + 01
8	.178300E + 02	.166734E + 02	.115660E + 02	.100000E + 01
9	.792400E + 02	.718203E + 02	.741971E + 01	.100000E + 01
10	.215000E + 02	.191236E + 02	.237641E + 01	.100000E + 01
11	.403300E + 02	.380925E + 02	.223749E + 01	.100000E + 01
12	.2100000E + 02	.215930E + 02	− .593041E + 00	.100000E + 01
13	.135000E + 02	.124730E + 02	.102701E + 01	.100000E + 01
14	.197500E + 02	.186825E + 02	.106754E + 01	.100000E + 01
15	.240000E + 02	.233288E + 02	.671202E + 00	.100000E + 01
16	.290000E + 02	.296629E + 02	− .662928E + 00	.100000E + 01
17	.153500E + 02	.149136E + 02	.436360E + 00	.100000E + 01
18	.190000E + 02	.155514E + 02	.344862E + 01	.100000E + 01
19	.950000E + 01	.770681E + 01	.179319E + 01	.100000E + 01
20	.351000E + 02	.408880E + 02	− .578797E + 01	.100000E + 01
21	.179000E + 02	.205142E + 02	− .261418E + 01	.100000E + 01
22	.523200E + 02	.560065E + 02	− .368653E + 01	.100000E + 01
23	.187500E + 02	.233576E + 02	− .460757E + 01	.100000E + 01
24	.198300E + 02	.244029E + 02	− .457285E + 01	.100000E + 01
25	.107500E + 02	.109626E + 02	− .212584E + 00	.100000E + 01

$\hat{\beta}_0 = 2.3412$
$\hat{\beta}_1 = 1.6159$
$\hat{\beta}_2 = 0.014385$

reducen más los estimados de β_1 y β_2, y se agranda todavía más el estimado de β_0. Los redescendentes duros: la función de onda de Andrews y la 17A de Hampel (tablas 11.5 y 11.6) tratan todavía con más severidad a la observación 9. La función de onda elimina el punto 9 del conjunto de base de datos (peso = 0), mientras que la 17A casi lo hace (peso = 0.019), en ambos casos, los estimados de β_1 y β_2 disminuyen todavía más, mientras que el estimado de β_0 continúa aumentando.

Es interesante comparar los estimados robustos de parámetro con los de los mínimos cuadrados, cuando se elimina el punto 9. En el ejemplo 6.1 se vio que al eliminar la observación 9 se obtuvieron estimados por mínimos cuadrados de $\hat{\beta}_0 = 4.477$, $\hat{\beta}_1 = 1.498$ y $\hat{\beta}_2 = 0.010$. Los procedimientos robustos tienen, aproximadamente, el mismo efecto sobre los estimados de los parámetros, aunque en general no son tan severos con respecto a β_2, siendo esto que si se elimina el punto 9 y se usan los mínimos cuadrados ordinarios, o si se usa un procedimiento de regresión robusta, aumentará el estimado de la ordenada al origen y disminuirán los estimados de β_1 y β_2. Así, en este ejemplo, el uso de los procedimientos de

http://carlos2524.jimdo.com/

TABLA 11.3 Ajuste robusto (Huber $t = 2$) para los datos del tiempo de entrega

Observación i	y_i	$\hat{y}_i$	e_i	Peso
1	.166800E + 02	.217651E + 02	− .508511E + 01	.639744E + 00
2	.115000E + 02	.109809E + 02	.519115E + 00	.100000E + 01
3	.120300E + 02	.126296E + 02	− .599594E + 00	.100000E + 01
4	.148800E + 02	.105856E + 02	.429439E + 01	.757165E + 00
5	.137500E + 02	.146038E + 02	− .853800E + 00	.100000E + 01
6	.181100E + 02	.186051E + 02	− .495085E + 00	.100000E + 01
7	.800000E + 01	.794135E + 01	.586521E − 01	.100000E + 01
8	.178300E + 02	.169564E + 02	.873625E + 00	.100000E + 01
9	.792400E + 02	.692795E + 02	.996050E + 01	.327017E + 00
10	.215000E + 02	.193269E + 02	.217307E + 01	.100000E + 01
11	.403300E + 02	.372777E + 02	.305228E + 01	.100000E + 01
12	.210000E + 02	.216097E + 02	− .609734E + 00	.100000E + 01
13	.135000E + 02	.129900E + 02	.510021E + 00	.100000E + 01
14	.197500E + 02	.188904E + 02	.859556E + 00	.100000E + 01
15	.240000E + 02	.232828E + 02	.717244E + 00	.100000E + 01
16	.290000E + 02	.293174E + 02	− .317449E + 00	.100000E + 01
17	.153500E + 02	.152908E + 02	.592377E − 01	.100000E + 01
18	.190000E + 02	.158847E + 02	.311529E + 01	.100000E + 01
19	.950000E + 01	.845286E + 01	.104714E + 01	.100000E + 01
20	.351000E + 02	.399326E + 02	− .483256E + 01	.672828E + 00
21	.179000E + 02	.205793E + 02	− .267929E + 01	.100000E + 01
22	.523200E + 02	.542361E + 02	− .191611E + 01	.100000E + 01
23	.187500E + 02	.233102E + 02	− .456023E + 01	.713481E + 00
24	.198300E + 02	.243238E + 02	.449377E + 01	.723794E + 00
25	.107500E + 02	.115474E + 02	− .797359E + 00	.100000E + 01

$\hat{\beta}_0 = 3.3736$
$\hat{\beta}_1 = 1.5282$
$\hat{\beta}_2 = 0.013739$

regresión robusta es un compromiso entre mantener el punto 9 de los datos o eliminarlo y usar los mínimos cuadrados; también, nótese que los procedimientos robustos identifican con facilidad y en forma automática, las observaciones influyentes en esos datos.

¿Cuál conjunto de estimados robustos se debe usar en el modelo "final"? En este ejemplo particular no hay mucha diferencia, porque todos los procedimientos robustos producen estimados de coeficientes que son bastante parecidos. Con más análisis, como por ejemplo, con un estudio de los residuales de los diversos modelos, o una investigación de sus eficacias relativas como ecuaciones de predicción, se podrá contar con una base para la selección del modelo final.

Ejemplo 11.2 *Datos de pérdidas por la chimenea*

Andrews [1974] usó datos de pérdidas por la chimenea, analizados por Daniel y Wood [1980], para ilustrar la regresión robusta. Los datos, tomados de una planta de ácido nítrico

http://carlos2524.jimdo.com/

TABLA 11.4 Ajuste robusto (función de Ramsay, $E_{0.3}$) para los datos del tiempo de entrega

Observación i	y_i	$\hat{y}_i$	e_i	Peso
1	.166800E + 02	.218009E + 02	−.512091E + 01	.388862E + 00
2	.115000E + 02	.112454E + 02	.254584E + 00	.953571E + 00
3	.120300E + 02	.128682E + 02	−.838215E + 00	.857082E + 00
4	.148800E + 02	.108415E + 02	.403846E + 01	.474482E + 00
5	.137500E + 02	.147670E + 02	−.101696E + 01	.829548E + 00
6	.181100E + 02	.186905E + 02	−.580544E + 00	.898687E + 00
7	.800000E + 01	.826846E + 01	−.268460E + 00	.952527E + 00
8	.178300E + 02	.170677E + 02	.762255E + 00	.868454E + 00
9	.792400E + 02	.682279E + 02	.110121E + 02	.131649E + 00
10	.215000E + 02	.194307E + 02	.206932E + 01	.682987E + 00
11	.403300E + 02	.369364E + 02	.339359E + 01	.535303E + 00
12	.210000E + 02	.216035E + 02	−.603534E + 00	.895015E + 00
13	.135000E + 02	.132081E + 02	.291877E + 00	.947139E + 00
14	.197500E + 02	.189862E + 02	.763769E + 00	.868651E + 00
15	.240000E + 02	.232651E + 02	.734923E + 00	.873345E + 00
16	.290000E + 02	.291901E + 02	−.190116E + 00	.964755E + 00
17	.153500E + 02	.154431E + 02	−.931215E − 01	.983506E + 00
18	.190000E + 02	.160129E + 02	.298707E + 01	.576113E + 00
19	.950000E + 01	.875713E + 01	.742875E + 00	.871123E + 00
20	.351000E + 02	.395347E + 02	−.443471E + 01	.440968E + 00
21	.179000E + 02	.205893E + 02	−.268928E + 01	.609354E + 00
22	.523200E + 02	.534802E + 02	−.116016E + 01	.806251E + 00
23	.187500E + 02	.232921E + 02	−.454212E + 01	.432732E + 00
24	.198300E + 02	.243045E + 02	−.447455E + 01	.437999E + 00
25	.107500E + 02	.117882E + 02	−.103817E + 01	.826364E + 00

$\hat{\beta}_0 = 3.8021$
$\hat{\beta}_1 = 1.4894$
$\hat{\beta}_2 = 0.013523$

por oxidación de amoníaco, se ven en la tabla 11.7. Un ajuste por mínimos cuadrados ordinarios (MCO) para estos datos da como resultado

$$\hat{y} = -39.9 + 0.72x_1 + 1.30x_2 - 0.15x_3$$

Los residuales de este modelo se ven en la columna 1 de la tabla 11.8, y en la figura 11.6a se ve una gráfica de probabilidad normal. Daniel y Wood hacen notar que el residual del punto 21 es anormalmente grande y tiene una influencia considerable sobre los coeficientes de regresión. Después de un análisis detallado, omitieron de los datos los puntos 1, 3, 4 y 21. El ajuste MCO[1] para los datos restantes es

$$\hat{y} = -37.6 + 0.80x_1 + 0.58x_2 - 0.07x_3$$

[1] Daniel y Wood ajustan un modelo donde intervienen x_1, x_2 y x_1^2. Andrews eligió trabajar con los tres regresores originales. Observa que si se omite x_3 y se agrega x_1^2, se obtienen residuales más pequeños, pero los resultados generales son iguales.

http://carlos2524.jimdo.com/

TABLA 11.5 Ajuste robusto (función de onda de Andrews con $a = 1.48$) para los datos del tiempo de entrega.

Observación i	y_i	$\hat{y}_i$	e_i	Peso
1	.166800E + 02	.216430E + 02	− .496300E + 01	.427594E + 00
2	.115000E + 02	.116923E + 02	− .192338E + 00	.998944E + 00
3	.120300E + 02	.131457E + 02	− .111570E + 01	.964551E + 00
4	.148800E + 02	.114549E + 02	.342506E + 01	.694894E + 00
5	.137500E + 02	.152191E + 02	− .146914E + 01	.939284E + 00
6	.181100E + 01	.188574E + 02	− .747381E + 00	.984039E + 00
7	.800000E + 01	.890189E + 01	− .901888E + 00	.976864E + 00
8	.178300E + 02	.174040E + 02	.425984E + 00	.994747E + 00
9	.792400E + 02	.660818E + 02	.131582E + 02	.0
10	.215000E + 02	.192716E + 02	.222839E + 01	.863633E + 00
11	.403300E + 02	.363170E + 02	.401296E + 01	.597491E + 00
12	.210000E + 02	.218392E + 02	− .839167E + 00	.980003E + 00
13	.135000E + 02	.135744E + 02	− .744338E − 01	.999843E + 00
14	.197500E + 02	.189979E + 02	.752115E + 00	.983877E + 00
15	.240000E + 02	.232029E + 02	.797080E + 00	.981854E + 00
16	.290000E + 02	.296336E + 02	.366350E + 00	.996228E + 00
17	.153500E + 02	.158247E + 02	− .474704E + 00	.993580E + 00
18	.190000E + 02	.164593E + 02	.254067E + 01	.824146E + 00
19	.950000E + 01	.946384E + 01	.361558E − 01	.999956E + 00
20	.351000E + 02	.387684E + 02	− .366837E + 01	.655336E + 00
21	.179000E + 02	.209308E + 02	− .303081E + 01	.756603E + 00
22	.523200E + 02	.523766E + 02	− .566063E − 01	.999908E + 00
23	.187500E + 02	.232271E + 02	− .447714E + 01	.515506E + 00
24	.198300E + 02	.240095E + 02	− .417955E + 01	.567792E + 00
25	.107500E + 02	.123027E + 02	− 1.55274E + 01	.932266E + 00

$\hat{\beta}_0 = 4.6532$

$\hat{\beta}_1 = 1.4582$

$\hat{\beta}_2 = 0.012111$

En la columna 2 de la tabla 11.8 aparecen los residuales con este modelo, y en la figura 11.6b está la gráfica de probabilidad normal correspondiente. Esta gráfica no indica algún comportamiento anormal en los residuales.

Andrews [1974] observa que la mayoría de los usuarios de regresión no tienen la destreza de Daniel y Wood, y emplea métodos de regresión robusta para llegar a resultados equivalentes. Un ajuste robusto a los datos de pérdidas en la chimenea, con la función de onda con $a = 1.5$, da como resultado

$$\hat{y} = -37.2 + 0.82x_1 + 0.52x_2 - 0.07x_3$$

Es virtualmente la misma ecuación que la determinada por Daniel y Wood mediante MCO y un análisis muy cuidadoso. Los residuales de este modelo se ven en la columna 3 de la tabla 11.8, y la gráfica de probabilidad normal está en la figura 11.7a. Los cuatro puntos anormales se identifican con claridad en esta gráfica. Por último, Andrews obtuvo un ajuste robusto a los datos, eliminando los puntos 1, 3, 4 y 21. La ecuación resultante es idéntica a la que se determinó con todos los 21 puntos. Los residuales de este ajuste, y la gráfica

http://carlos2524.jimdo.com/

TABLA 11.6 Ajuste robusto (función 17A de Hampel) para los datos del tiempo de entrega

Observación i	y_i	$\hat{y}_i$	e_i	Peso
1	.166800E + 02	.216086E + 02	− .492859E + 01	.561047E + 00
2	.115000E + 02	.116605E + 02	− .160516E + 00	.100000E + 01
3	.120300E + 02	.130997E + 02	− .106968E + 01	.100000E + 01
4	.148800E + 02	.114491E + 02	.343090E + 01	.805460E + 00
5	.137500E + 02	.152238E + 02	− .147384E + 01	.100000E + 01
6	.181100E + 02	.188502E + 02	− .740196E + 00	.100000E + 01
7	.800000E + 01	.887367E + 01	− .873671E + 00	.100000E + 01
8	.178300E + 02	.174110E + 02	.418966E + 00	.100000E + 01
9	.792400E + 02	.661574E + 02	.130826E + 02	.192790E − 01
10	.215000E + 02	.192131E + 02	.228695E + 01	.100000E + 01
11	.403300E + 02	.363522E + 02	.397779E + 01	.695775E + 00
12	.210000E + 02	.218738E + 02	− .873835E + 00	.100000E + 01
13	.135000E + 02	.135479E + 02	− .478834E − 01	.100000E + 01
14	.197500E + 02	.189657E + 02	.784337E + 00	.100000E + 01
15	.240000E + 02	.232006E + 02	.799403E + 00	.100000E + 01
16	.290000E + 02	.286019E + 02	.398080E + 00	.100000E + 01
17	.153500E + 02	.158235E + 02	− .473491E + 00	.100000E + 01
18	.190000E + 02	.164756E + 02	.252442E + 01	.100000E + 01
19	.950000E + 01	.945380E + 02	.462003E − 01	.100000E + 01
20	.351000E + 02	.388032E + 02	− .370325E + 01	.746183E + 00
21	.179000E + 02	.209744E + 02	− .307436E + 01	.900263E + 00
22	.523200E + 02	.524915E + 02	− .171475E + 00	.100000E + 01
23	.187500E + 02	.232246E + 02	− .447458E + 01	.618068E + 00
24	.198300E + 02	.239757E + 02	− .414568E + 01	.666865E + 00
25	.107500E + 02	.122886E + 02	− .153862E + 01	.100000E + 01

$\hat{\beta}_0 = 4.6192$

$\hat{\beta}_1 = 1.4676$

$\hat{\beta}_2 = 0.011993$

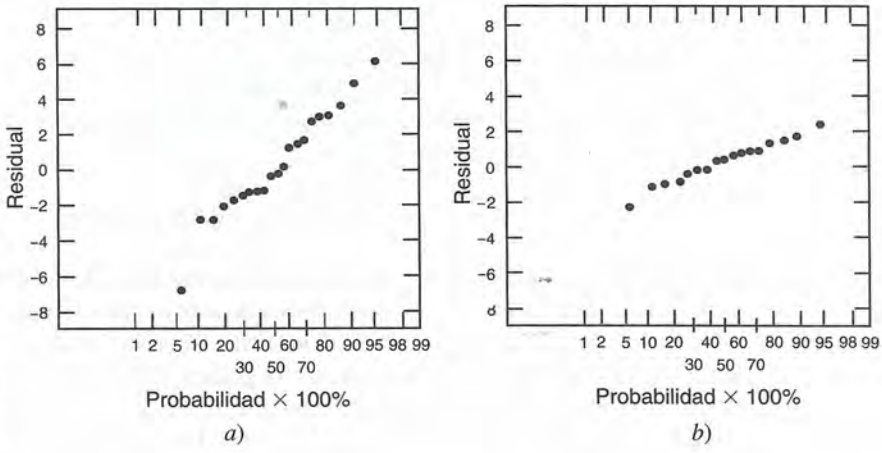


Figura 11.6 Gráficas de probabilidad normal para ajustes por mínimos cuadrados; a) mínimos cuadrados con los 21 puntos; b) mínimos cuadrados omitiendo los puntos 1, 3, 4 y 21. (De Andrews [1974], con autorización del editor.)

http://carlos2524.jimdo.com/

TABLA 11.7 Datos de pérdidas por la chimenea, tomados de Daniel y Wood [1980].

Observación número	Pérdidas por la chimenea y	Flujo de aire x_1	Temperatura de entrada de agua de enfriamiento x_2	Concentración de ácido x_3
1	42	80	27	89
2	37	80	27	88
3	37	75	25	90
4	28	62	24	87
5	18	62	22	87
6	18	62	23	87
7	19	62	24	93
8	20	62	24	93
9	15	58	23	87
10	14	58	18	80
11	14	58	18	89
12	13	58	17	88
13	11	58	18	82
14	12	58	19	93
15	8	50	18	89
16	7	50	18	86
17	8	50	19	72
18	8	50	19	79
19	9	50	20	80
20	15	56	20	82
21	15	70	20	91

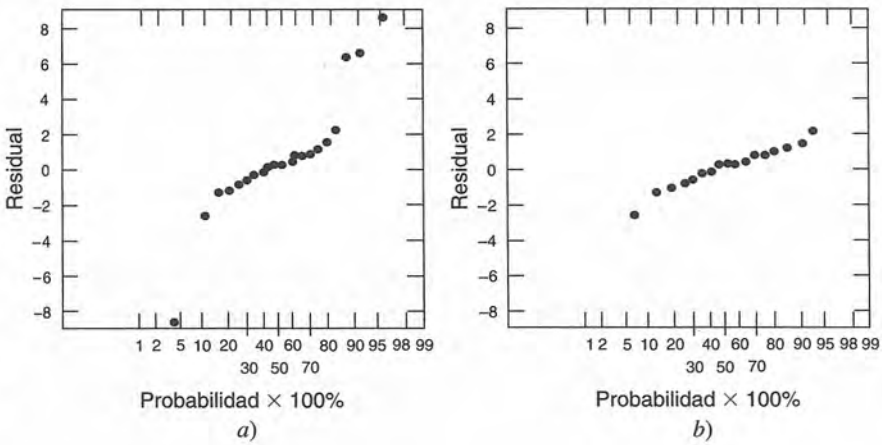


Figura 11.7 Gráficas de probabilidad normal para ajustes robustos: a) ajuste robusto de los 21 puntos; b) ajuste robusto sin los puntos 1, 3, 4 y 21. (De Andrews [1974], con autorización del editor.)

de probabilidad normal correspondiente, están en la columna 4 de la tabla 11.8, y en la figura 11.7b, respectivamente. Esta gráfica de probabilidad normal es casi idéntica a la obtenida con el análisis de MCO eliminando los puntos 1, 3, 4 y 21 (Fig. 11.6b).

De nuevo se ve que la aplicación rutinaria de la regresión robusta ha conducido a la identificación automática de los puntos anormales. También produce un ajuste que no depende en forma importante de esos puntos, así, se puede considerar que los métodos de regresión robusta son procedimientos para aislar puntos anormalmente influyentes, para poder estudiarlos más.

http://carlos2524.jimdo.com/

TABLA 11.8 Residuales con diversos ajustes a los datos de pérdidas por la chimenea[a]

	Residuales			
	Mínimos cuadrados		Ajuste robusto de Andrews	
Observación	(1) Los 21 puntos	(2) Sin 1, 3, 4 y 21	(3) Los 21 puntos	(4) Sin 1, 3, 4 y 21
1	3.24	6.08[b]	6.11	6.11
2	−1.92	1.15	1.04	1.04
3	4.56	6.44	6.31	6.31
4	5.70	8.18	8.24	8.24
5	−1.71	−0.67	−1.24	−1.24
6	−3.01	−1.25	−0.71	−0.71
7	−2.39	−0.42	−0.33	−0.33
8	−1.39	0.58	0.67	0.67
9	−3.14	−1.06	−0.97	−0.97
10	1.27	0.35	0.14	0.14
11	2.64	0.96	0.79	0.79
12	2.78	0.47	0.24	0.24
13	−1.43	−2.51	−2.71	−2.71
14	−0.05	−1.34	−1.44	−1.44
15	2.36	1.34	1.33	1.33
16	0.91	0.14	0.11	0.11
17	−1.52	−0.37	−0.42	−0.42
18	−0.46	0.10	0.08	0.08
19	−0.60	0.59	0.63	0.63
20	1.41	1.93	1.87	1.87
21	−7.24	−8.63	−8.91	−8.91

[a]Adaptado de la tabla 5, en Andrews [1974], con autorización del editor.
[b]Los residuales subrayados corresponden a puntos no incluidos en el ajuste.

Cálculo de los estimados M

Actualmente, la regresión robusta no es una opción en la mayor parte de los programas estadísticos, sin embargo, como se hizo notar, se pueden obtener los estimados M, mediante los mínimos cuadrados reponderados iterativamente por lo que, en principio, es bastante fácil obtener estimados M. Se puede usar la subrutina SAS PROC NLIN de SAS para implementar un procedimiento iterativo. También en S-PLUS hay disponibles procedimientos de estimación robusta. En este capítulo se describirán más adelante (Sec. 11.4.4).

Los métodos de regresión robusta tienen mucho que ofrecer al analista de datos. Pueden ser de extrema ayuda para localizar valores atípicos y observaciones muy influyentes. Siempre que se hace un análisis por mínimos cuadrados, sería conveniente también hacer un ajuste robusto. Si concuerdan, en forma sustancial, los resultados de los dos procedimientos, se deben usar los resultados de mínimos cuadrados, sin embargo, si son distintos, se deben identificar las razones de las diferencias. Las observaciones que se ponderan bajo en el ajuste robusto se deben examinar con cuidado.

http://carlos2524.jimdo.com/

11.3 PROPIEDADES DE LOS ESTIMADORES ROBUSTOS

En esta sección se presentarán dos propiedades importantes de los estimadores robustos: el **quiebre** y la **eficiencia**. Se observará que el punto de quiebre de un estimador es un asunto práctico que se debe tener en cuenta al seleccionar un procedimiento robusto de estimación, generalmente, los estimadores M funcionan mal con respecto al punto de quiebre. Esto ha impulsado el desarrollo de muchos otros métodos alternativos. En la sección 11.4 se resumirán varios de esos procedimientos.

11.3.1 Punto de quiebre

El **punto de quiebre** con muestra finita es la mínima fracción de datos anómalos que puede hacer inútil al estimador. El mínimo punto de quiebre posible es $1/n$, lo que equivale a que una sola observación puede distorsionar tanto al estimador como para hacer que su utilidad práctica sea nula, para quien forma el modelo. El punto de quiebre de los MCO es $1/n$. Para visualizarlo, examínense los diagramas de dispersión en las figuras 11.1 y 11.2. Es claro que el punto A de la figura 11.1 ha causado que el ajuste por MCO para todas las observaciones sea inútil, en esencia. De igual manera si el punto A de la figura 11.2 se moviera a cualquiera de los lugares alternos que se muestran, se afectarían los MCO en forma dramática.

Los estimadores M se pueden alterar debido a valores atípicos en el espacio de x, de manera idéntica que los MCO, consecuentemente, el punto de quiebre de la clase de estimadores M es $1/n$. Esto tiene un impacto potencialmente grave sobre su uso práctico, porque se puede dificultar la determinación del grado de contaminación de la muestra por datos anómalos. La mayoría de los analistas de datos con experiencia creen que la fracción de los datos que contaminan está comúnmente entre 1 y 10%. En consecuencia, en general se desea que el punto de quiebre de un estimador sea mayor que el 10%. Esto ha conducido al desarrollo de **estimadores de punto de quiebre alto**. En la sección 11.4.1 se describirán algunos de esos estimadores.

11.3.2 Eficiencia

Supóngase que un conjunto de datos no tiene errores grandes, que no hay observaciones influyentes, y que las observaciones provienen de una distribución normal. Si se usa un estimador robusto para ese conjunto de datos, se trata de que los resultados sean prácticamente idénticos a los de MCO, porque los mínimos cuadrados ordinarios son la técnica correcta para esos datos. Se puede concebir la eficiencia de un estimador robusto como el cuadrado medio residual obtenido con los MCO, dividido entre el cuadrado medio residual obtenido con el procedimiento robusto. Es obvio que se desea que esa medida de eficiencia se aproxime a 1.

Hay un gran énfasis, en las publicaciones sobre regresión robusta, hacia la **eficiencia asintótica**, es decir, la eficiencia de un estimador cuando el tamaño de muestra n se hace infinito. Es un concepto útil para comparar los estimadores robustos, pero muchos problemas prácticos de regresión manejan tamaños de muestra de pequeños a moderados (por ejemplo, $n < 50$), y se sabe que las eficiencias con muestra pequeña son muy distintas a sus valores asintóticos; en consecuencia, quien forma el modelo debe interesarse en el

http://carlos2524.jimdo.com/

comportamiento asintótico de cualquier estimador, que se pudiera usar en determinada situación, aunque tampoco debe darle excesiva importancia. Lo más importante, desde un punto de vista práctico, es la **eficiencia con muestra finita**, es decir, lo bien que funciona determinado estimador respecto a los MCO, con datos "limpios", para tamaños de muestra parecidos a los del problema que se tiene a la mano. La eficiencia de un estimador robusto con muestra finita se define como la relación del cuadrado medio residual por MCO entre el cuadrado medio residual del estimador robusto, estando aplicados los MCO sólo a los datos limpios. Los métodos de simulación de Monte Carlo se usan con frecuencia para evaluar la eficiencia con muestra finita.

11.4 RESEÑA DE OTROS ESTIMADORES ROBUSTOS DE REGRESIÓN

En esta sección se presentará una breve perspectiva de algunos de los estimadores de regresión robusta, que se han propuesto como alternativas de los estimadores M en la sección 11.2. Se describirán las ventajas y desventajas de cada método, así como algunos aspectos de cómputo. Ésta es un área en la que en la actualidad se investiga mucho, por lo que esta descripción sólo es una introducción a este importante aspecto del modelado por regresión.

11.4.1 Estimadores de punto de quiebre alto

Como los estimadores por MCO y los M tienen el inconveniente de un bajo punto de quiebre ($1/n$), se ha tratado mucho de determinar estimadores que funcionen mejor con respecto a esta propiedad. Con frecuencia, se considera deseable tener un punto de quiebre más o menos de 50%. En esta sección se describirán algunos de los más conocidos estimadores de regresión, de punto de quiebre alto.

Mínima mediana de cuadrados

El estimador de mínima mediana de cuadrados (MMC) se obtiene determinando el modelo de regresión que minimice el cuadrado del residual de orden m, siendo $m = [n/2] + [(p + 1)/2]$, donde [.] representa la parte entera del argumento; este estimador fue introducido por Rousseeuw [1984] (véanse también Rousseeuw y Leroy [1987] y Rousseeuw [1998]). La base de este criterio es directa. Los mínimos cuadrados ordinarios minimizan la suma de los cuadrados de los residuales, por lo que también minimiza el promedio de los residuales al cuadrado. Ahora bien, el promedio no suele ser un buen estimador de la ubicación, cuando hay valores atípicos, y con frecuencia es preferible la mediana, por consiguiente, la MMC minimiza la mediana de los residuales al cuadrado. El criterio MMC también tiene una buena interpretación geométrica: es la línea (o el plano) que pasa por el centro de la banda más angosta que abarca la mitad de las observaciones.

El criterio MMC está implementado en el programa de cómputo PROGRESS (programa para regresión robusta), descrito en Rousseeuw y Leroy [1987], que usa un procedimiento combinatorio para ajustar el modelo de regresión. Para el caso de un solo regresor eso implica evaluar los modelos para todos los $\binom{n}{2}$ subconjuntos y utilizar el modelo que minimiza la mediana de los residuales al cuadrado. Los programas más recientes usan muestreo aleatorio (no exhaustivo) o algoritmos genéticos. El estimador MMC también se encuentra en S-PLUS y en SAS/IML, llamando a la subrutina LMS (*least median of squares*).

http://carlos2524.jimdo.com/

Ejemplo 11.3 *Datos de pérdidas por la chimenea*

Se usarán los datos de pérdidas por la chimenea que se presentaron en el ejemplo 11.2. Se observó que al menos hay cuatro puntos anómalos de datos: los números 1, 3, 4 y 21. En el ejemplo 11.2 se determinó un modelo adecuado de regresión, usando la función de onda de Andrews para ilustrar las técnicas de estimación de M. El modelo de regresión MMC para los datos de pérdidas por la chimenea, determinado con S-PLUS, es

$$\hat{y} = -39.25 + 0.75x_1 + 0.500x_2 + 0.00x_3$$

Regresando al ejemplo 11.2, obsérvese lo parecido que es el ajuste MMC al modelo obtenido con la función de onda de Andrews. Ambos ajustes robustos son muy distintos del ajuste por mínimos cuadrados ordinarios, en especial con respecto a los coeficientes de x_2 y de x_3, en realidad, el modelo MMC elimina al regresor x_3. Este ajuste MMC también dejaría grandes los residuales estandarizadas de las observaciones 1, 3, 4 y 21, indicando que se deberían examinar con más cuidado.

Es obvio que el ajuste MMC trabajó bien en el ejemplo 11.3. En general, la ventaja de la MMC es que tiene un punto de quiebre alto, en teoría del 50% que nos dice que podría ser anómala hasta la mitad de los datos, sin que el modelo de regresión se tornara inútil. Ahora se presentará un ejemplo (adaptado de Ryan [1997]) para revisar algunas de sus desventajas.

Ejemplo 11.4

La tabla 11.9 y la figura 11.8 presentan datos de Ryan [1997], quien modificó un ejemplo de Stefanski [1991]. Nótese que hay dos valores atípicos, las observaciones 6 y 7. Un ajuste por mínimos cuadrados ordinarios a las siete observaciones buenas produciría un modelo con un estimado de pendiente casi de cero. De hecho, los valores atípicos tienen poco efecto sobre el ajuste por mínimos cuadrados ordinarios (obsérvese que son valores atípicos en el espacio y). Se supondrá que es el modelo que preferiría el usuario.

En esta muestra hay tres observaciones buenas, aproximadamente alineadas con los dos puntos atípicos. Esto es desafortunado en especial en este caso, porque la MMC minimizará el quinto residual al cuadrado en ese ordenamiento (porque $[n/2] + [(p + 1)/2] = [9/2] + [(2 + 1)/2] = 5$), y como cinco puntos están en una recta, el quinto residual al cuadrado en ese ordenamiento debe ser cero. El modelo de regresión MMC es la recta que pasa por

TABLA 11.9 Datos para el ejemplo 11.4

Observación	x	y
1	2.5	5.0
2	7.5	2.6
3	14.0	2.6
4	15.0	3.9
5	16.0	5.1
6	22.0	13.0
7	23.0	14.2
8	23.0	5.1
9	29.0	2.7

http://carlos2524.jimdo.com/

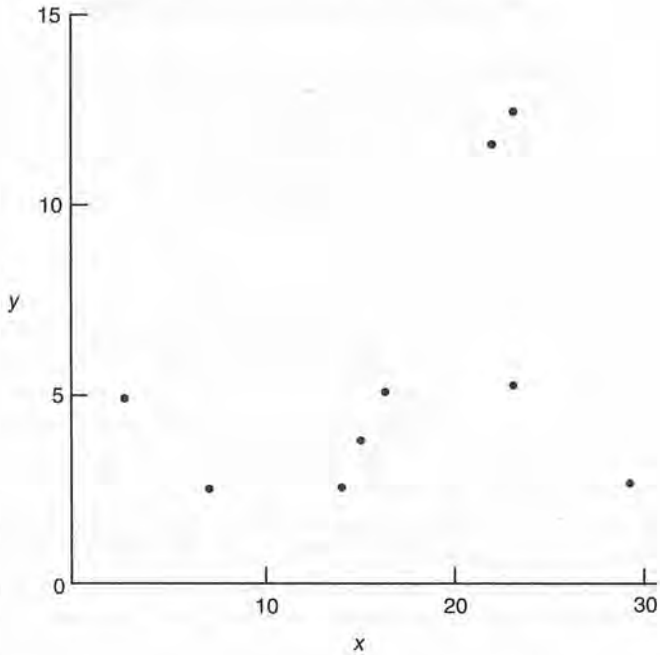


Figura 11.8 Diagrama de dispersión para los datos de la tabla 11.9 (ejemplo 11.4).

esos cinco puntos. Es claro que esta línea es muy distinta de la de mínimos cuadrados para los siete valores buenos de los datos.

Este ejemplo ilustra un grave problema con la MMC, la llamada **propiedad de ajuste exacto**. Básicamente, la MMC ajusta sólo la mitad de los datos, por lo que es posible que los valores atípicos que estén alineados con valores de datos buenos alejen al modelo de regresión de la ecuación deseada para los datos buenos. Consecuentemente, la MMC puede funcionar mal en comparación con los mínimos cuadrados ordinarios, cuando éstos son el criterio adecuado. La **eficiencia asintótica** de la MMC en realidad es cero, porque requiere minimizar al residual al cuadrado en determinado punto, ignorando de hecho el ajuste en las demás $n-1$ observaciones. Cuando n se vuelve grande, cabría esperar que el ajuste en esos $n-1$ puntos sea malo, en comparación con el de los mínimos cuadrados. La eficiencia de la MMC con muestra finita también puede ser muy baja. En consecuencia, es difícil recomendar la mínima mediana de cuadrados (MMC) como procedimiento independiente de regresión, además siempre se deben comparar los ajustes por MMC y por cuadrados mínimos ordinarios.

Mínimos cuadrados recortados

El **estimador de suma mínima recortada de cuadrados**, o de **mínimos cuadrados recortados** (MCR) se calcula determinando los parámetros del modelo que satisfaga

$$\underset{\beta}{\text{Minimizar}} \sum_{i=1}^{h} e_{(i)}^2 \tag{11.13}$$

en donde $e_{(1)}^2 < e_{(2)}^2 < \cdots < e_{(n)}^2$ son los residuales elevados al cuadrado y ordenados, y se debe determinar h. Las mejores propiedades de robustez se obtienen cuando $h = n/2$, aproximadamente, en cuyo caso se alcanza un punto de quiebre de 50%, sin embargo, como ese

http://carlos2524.jimdo.com/

punto de 50% puede producir a veces malos resultados, será mejor usar un valor mayor de h para aumentar la eficiencia. Si se define un **porcentaje de recorte** α, una elección razonable de h sería $h = [n(1-\alpha)] + 1$. Una buena referencia sobre MCR se encuentra en Rousseeuw [1984].

Si bien los mínimos cuadrados recortados se pueden configurar para tener buenas propiedades de quiebre, su eficiencia asintótica es mala en comparación con los mínimos cuadrados ordinarios. Su eficiencia con muestra finita también puede ser muy mala, además, también puede ser difícil el cálculo de los estimados por MCR, en especial con problemas donde las muestras son de moderadas a grandes. El ajuste MCR se puede aproximar con PROGRESS y con un algoritmo desarrollado por Hawkins [1994], también, los mínimos cuadrados recortados tienen la propiedad de ajuste exacto de la MMC. Para obtener un ajuste MCR también se puede usar S-PLUS.

Estimadores S

Los estimadores S fueron desarrollados por Rousseeuw y Yohai [1984], también los describen Rousseeuw y Leroy [1987]. El estimador se obtiene como sigue:

$$\underset{\beta}{\text{Minimizar}}\ S[e_1(\beta), e_2(\beta), \ldots, e_n(\beta)] \tag{11.14}$$

en la que las $e_i(\beta)$ son los residuales de una solución β probable, o candidata, y $S[e_1(\beta), e_2(\beta), \ldots, e_n(\beta)]$ se determina por la solución de

$$\frac{1}{n}\sum_{i=1}^{n} \rho(e_i/s) = k \tag{11.15}$$

Rousseeuw y Yohai [1984] sugieren que la función $\rho(\cdot)$ debería ser

$$\rho(z) = \begin{cases} \dfrac{z^2}{2} - \dfrac{z^4}{2c^2} + \dfrac{z^6}{6c^4}, & |z| \le c \\ \dfrac{c^2}{6}, & |z| > c \end{cases} \tag{11.16}$$

La elección de la constante c determina k, e implica un balance entre el punto de quiebre y la eficiencia. Rousseeuw y Leroy [1987] dan una guía para la selección de esos parámetros, y presentan algunas eficiencias asintóticas. En general, se pueden configurar los estimadores S para tener un punto de quiebre alto (50%), pero la eficiencia (tanto asintótica como para muestras finitas) puede ser baja. En el punto de quiebre de 50% los estimadores S tienen mayor eficiencia asintótica que la MMC o los MCR.

De acuerdo con nuestra experiencia, la estimación de S no se usa mucho en la práctica. Esto se debe, en parte, a la falta de programas de cómputo. Con la Biblioteca ROBETH (Marazzi [1993]) es posible calcular estimados S para modelos de regresión en S-PLUS.

Estimadores MM

El estimador *MM* fue propuesto originalmente por Yohai [1987]; el objetivo fue producir un estimador de punto de quiebre alto, que mantuviera una buena eficiencia, el estimador *MM* tiene tres etapas. El estimado inicial es un estimado S, por lo que tiene punto de quiebre

http://carlos2524.jimdo.com/

alto. En la segunda etapa se calcula un estimado M de la desviación estándar del error, con los residuales del estimado S inicial. El último paso es un estimado M de los parámetros de regresión mediante una función ψ dura redescendente, esto es, se asigna un peso igual a cero a los residuales suficientemente grandes. Los estimados MM se pueden calcular con la versión 4.5 de S-PLUS.

En una extensa evaluación del desempeño de varios estimadores de regresión robusta, Simpson y Montgomery (1998c) informan que los estimadores MM tienen alta eficiencia y trabajan bien en la mayor parte de los escenarios de valores atípicos. Su único punto débil se descubre en casos en donde hay un gran porcentaje de valores atípicos en el espacio x que tengan residuales de tamaño moderado. La estimación MM maneja mucho mejor valores atípicos grandes, aun cuando sean puntos de gran influencia.

11.4.2 Estimadores de influencia limitada

Estimadores GM

Se hizo notar antes que el punto de quiebre de un estimador M es $1/n$, esto es, los residuales se ponderan de acuerdo con su magnitud, sin tener en cuenta la influencia asociada con la observación. Es posible que una observación tenga un residual muy pequeño y, sin embargo, la diagonal del sombrero h_{ii} correspondiente puede ser muy grande. Así, un punto de gran influencia podría tener asignado un peso total (o casi total) con un estimador M, en consecuencia, una sola observación anómala podría destruir la utilidad del estimador.

Esta dificultad con los estimadores M ha llevado a desarrollar los **estimadores de regresión de influencia limitada**. Un ejemplo de ellos sería hacer una regresión robusta, como se describió para el estimador M, pero reemplazar los residuales ordinarios por residuales estudentizados, entre otras alternativas posibles están los residuales $DFFITS_i$ o los PRESS $e_{(i)}$. La base en este caso es que si h_{ii} en la diagonal de la matriz sombrero es grande (cercana a la unidad), el residual estudentizado tendrá mayor magnitud que el residual ordinario (estandarizado), por lo que la observación tendrá menos peso asignado. Para aumentar la complejidad de esos estimadores, la medida de la diagonal de la matriz sombrero tiene también un punto de quiebre de $1/n$. Así, un estimado robusto de la influencia se usa con frecuencia, basado en los estimados de vector medio y de matriz de covarianza, a partir del elipsoide de volumen mínimo (véase Rousseeuw y Leroy [1987]), de los estimados M de covarianza (Maronna [1976]) o el método híbrido de Rocke y Woodruff [1996]. Los estimadores de regresión de influencia limitada también se llaman **estimadores *M* generalizados** o **estimadores *GM***. Hay buenas discusiones acerca de regresión de influencia limitada en Krasker y Welsch [1982], Staudte y Sheather [1990], Simpson y Montgomery [1998a] y Ryan [1997].

La ventaja de estos estimadores es que no son tan sensibles a valores atípicos en el espacio x como los estimadores M. Su desventaja es que su punto de quiebre no puede ser mayor que $1/p$, donde p es la cantidad de parámetros, resultado que cuando hay muchos parámetros, el punto de quiebre de un estimador de influencia acotada no puede ser mucho mejor que el peor de los casos, $1/n$; sin embargo, cuando p es pequeño, con frecuencia los estimadores de influencia acotada pueden producir muy buenos resultados. Como se puede usar el método de los mínimos cuadrados iterativamente reponderados para obtener estimadores de regresión de influencia limitada, su implementación en computadora es relativamente directa.

http://carlos2524.jimdo.com/

Estimadores compuestos

Como es difícil obtener todas las propiedades deseables en un estimador, como su influencia limitada, punto de quiebre alto y alta eficiencia, con un solo procedimiento de estimación, se ha dado mucha atención, en los años recientes, a procedimientos de combinación de dos o más procedimientos que formen un **estimador compuesto**. Un método para desarrollar un estimador compuesto sería usar un estimador de punto de quiebre alto en la primera etapa, seguido de un estimador de influencia limitada en la segunda, lo cual nos indica que el procedimiento inicial de estimación de punto de quiebre alto va seguido sólo de una iteración del procedimiento de influencia acotada, para mantener la propiedad de quiebre alta de los estimados iniciales. Simpson, Ruppert y Carroll [1992], Coakley y Hettmansperger [1993] y Simpson y Montgomery [1998a, b, c] propusieron esta clase de estimadores. El estimador de Coakley-Hettmansperger usa los mínimos cuadrados recortados como estimado inicial, y sigue con un proceso *GM* de una etapa, con una función ψ del tipo de Huber. El procedimiento de Simpson-Montgomery usa un estimador *S* en la primera etapa, con un procedimiento *GM* de un paso, con una función ψ basada en el bipeso de Tukey (véase el Prob. 11.11).

Los estudios preliminares indican que los estimadores compuestos prometen mucho, pero en la actualidad su aplicación es limitada por la falta de programas de cómputo y de extensos análisis de eficiencia que guíen al profesionista en la selección de los componentes de cada etapa de un procedimiento. Simpson y Montgomery (1998b, c) citan los resultados de dos estudios de eficiencia en donde tanto el procedimiento de Coakley-Hettmansperger y el de ellos (Simpson y Montgomery) funcionaron muy bien respecto al punto de quiebre y menos razonablemente bien con respecto a la eficiencia. El programa S-PLUS del procedimiento de Coakley-Hettmansperger se consigue en Wilcox [1997] y el programa S-PLUS para el de Simpson-Montgomery se consigue directamente de los autores.

11.4.3 Otros procedimientos

Estimación R y L

Además de los estimadores descritos en las secciones anteriores, hay varios otros métodos de regresión robusta. La **estimación *R*** se basa en **rangos** o **jerarquía**. Para ilustrar el procedimiento general, considérese el reemplazo de un factor en la función objetivo de mínimos cuadrados $S(\beta) = \Sigma_{i=1}^{n}(y_i - \mathbf{x}_i'\beta)^2$ por su rango.

Así, si R_i es el rango de $y_i - \mathbf{x}_i'\beta$, se desea minimizar $\Sigma_{i=1}^{n}(y_i - \mathbf{x}_i'\beta)R_i$. En forma más general, se podrían reemplazar los rangos (que son los enteros 1, 2, . . . , *n*) por la **función de puntuación** $a(i)$, $i = 1, 2, \ldots, n$, de modo que la función objetivo es

$$\underset{\beta}{\text{Minimizar}} \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\boldsymbol{\beta})a(R_i)$$

Si se iguala la función de puntuación con los rangos, esto es, si $a(i) = i$, los resultados se llaman **puntaje de Wilcoxon**, o **calificaciones de Wilcoxon**. Otra posibilidad es usar los **puntajes medianos**, que equivaldría a, $a(i) = -1$ si $i < (n + 1)/2$, y $a_i = 1$ si $i > (n + 1)/2$. Entre las referencias importantes sobre la estimación *R* en la regresión están Adichie [1967], Hogg y Randles [1975], Jaeckel [1972] y Jurečková [1977]. Los estimadores *R* no

http://carlos2524.jimdo.com/

tienen punto de quiebre alto, pero bajo ciertas condiciones (Jurečková [1977]) son asintóticamente equivalentes a los estimadores *M*. Hettmansperger y McKean [1998] presentan una descripción detallada de la teoría, inferencia, diagnóstico e implementación de modelos lineales basados en rangos, también incluyen un capítulo sobre los estimadores de influencia limitada basados en rangos (*R* generalizada) y métodos de rango de quiebre alto (HBR, de *high-breakdown rank*).

La estimación *L* se basa en **estadísticos de orden**. Por ejemplo, supóngase que se desea estimar el parámetro de localización de una distribución a partir de una muestra aleatoria $x_1, x_2, \ldots, x_n$. Los estadísticos de orden para esta muestra son $x_{[1]} \leq x_{[2]} \leq \cdots \leq x_{[n-1]} \leq x_{[n]}$. La mediana de la muestra sería un estimador *L*, porque es una medida de localización basada en esos estadísticos de orden. En Andrews *et al.* [1972] se describen varios otros estimadores *L* para el problema de la localización. El uso de estimación *L* en el contexto de regresión no es tan sencillo como la estimación *M* y *R*. Denby y Larson [1977] describen la **regresión por rebanadas**, que para un regresor divide los datos en grupos, y ajusta la línea recta usando los centroides de los grupos. Se podría usar una extensión de esta técnica del tipo regresión por etapas para la regresión múltiple. Moussa-Hamouda y Leone [1974, 1977a, b] proponen procedimientos de regresión lineal simple con observaciones repetidas en *y* a cada *x*, que implican recortar o descartar valores remotos de *y*. En general, los estimadores *L* no tienen alto quiebre, y en algunos casos no es obvia una generalización clara de un procedimiento que pase de la regresión lineal simple a la múltiple. DeJong, DeWet y Welsh [1988] proponen un estimador *L* de influencia limitada para regresión múltiple.

Regresión ridge robusta

Dos de los problemas más frecuentes con los que se encuentra el analista de regresión son la no normalidad de las observaciones, y la multicolinealidad. Aunque uno se imagina esos dos problemas por separado, en una gran cantidad de casos prácticos los valores atípicos resultantes de distribuciones de error con colas gruesas y con multicolinealidad se presentan en forma simultánea. Algunos autores han sugerido que podrían ser suficientes los métodos de estimación robustos o los sesgados, por separado, para manejar el problema combinado, sin embargo, como los estimados por regresión robusta son inestables con frecuencia, cuando la matriz **X** está deteriorada, sería preferible contar con una técnica para manejar, en forma directa, ambos problemas. Hogg [1979b] ha sugerido una forma robusta de regresión ridge; recuérdese que el estimador ridge $\hat{\beta}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$ se puede calcular aumentando los datos **X, y** con *p* seudoobservaciones, como sigue:

$$\mathbf{X}_A = \left[\frac{\mathbf{X}}{\sqrt{k}\,\mathbf{I}}\right], \qquad \mathbf{y}_A = \left[\frac{\mathbf{y}}{\mathbf{0}}\right]$$

para aplicar entonces los mínimos cuadrados ordinarios a $\mathbf{X}_A$, $\mathbf{y}_A$, obteniendo

$$\hat{\beta}_R = (\mathbf{X}'_A\mathbf{X}_A)^{-1}\mathbf{X}'_A\mathbf{y}_A = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$$

Una versión robusta de la regresión ridge reemplazaría los mínimos cuadrados para $\mathbf{X}_A$, $\mathbf{y}_A$, por la función objetivo robusta, que fuera adecuada.

Askin y Montgomery [1980, 1984] investigaron este método e hicieron notar que los estimadores ridge generalizados, los de componentes principales y los de rango fraccionario, se pueden ajustar mediante mínimos cuadrados con datos aumentados en forma adecua-

http://carlos2524.jimdo.com/

da, para que se puedan calcular con facilidad las versiones robustas de estos estimadores sesgados. Dicen que el procedimiento combinado de estimación produce estimados estables de coeficientes, y al mismo tiempo localizan e identifican valores atípicos. Los procedimientos computacionales del procedimiento también son buenos, porque el algoritmo de mínimos cuadrados iterativamente reponderados suele converger hacia los estimados finales, en menos iteraciones que las que se necesitarían si sólo se usara el criterio robusto. Para conocer otros métodos de abordar el problema de la distribución de los errores no normales cuando existe simultáneamente multicolinealidad y valores atípicos, véanse Walker [1984], Pfaffenberger y Dielman [1985], y Simpson y Montgomery [1996], también el trabajo de Walker y Birch [1988] es una referencia útil; demuestran que existe una relación entre influencia y multicolinealidad, y que la influencia de cada observación depende mucho del valor del parámetro de sesgo que se use en la regresión ridge.

Regresión robusta no lineal

Ha habido relativamente escasa aplicación de los métodos de regresión robusta a la regresión no lineal. A diferencia de muchos modelos de regresión lineal, con frecuencia se necesitan valores extremos para identificar qué modelo no lineal ajustar. En consecuencia, los métodos de ajuste robusto serían muy útiles en casos en los que se conoce el tipo de modelo no lineal que se debe usar. Por fortuna, es lo que suele suceder en la práctica. Algunas referencias útiles sobre regresión robusta no lineal son Dutter y Huber [1981], Lawrence y Arthur [1990], Rousseeuw y Leroy [1987], Stromberg y Ruppert [1989] y Stromberg [1993].

11.4.4 Cálculo de estimadores de regresión robusta

La mayor parte de los programas estadísticos más usados no incluyen la regresión robusta, o bien su capacidad es muy limitada en ese aspecto. La mayor parte de ellos sí tienen opción para los mínimos cuadrados ponderados, por lo que es bastante fácil obtener estimados *M* o de otra clase, empleando el esquema IRLS descrito en la sección 11.2. El programa S-PLUS (versión 4.5) actualmente contiene una opción de menús para estimados *M*, *LMS (MMC)*, *LTS* (*MMR*) y *MM*. La potencia de S-PLUS para regresiones robustas aumenta mucho con la inclusión de la biblioteca ROBETH (Marazzi [1993]) que permite acceso a los estimadores *S, R, L, GM* y a los estimadores compuestos. Wilcox [1997] cita numerosas funciones de S-PLUS para *GM* y para estimadores compuestos, además de las rutinas de inferencia por regresión robusta.

La tabla 11.10 muestra los resultados de aplicar varios de los procedimientos de regresión robusta que se describieron en este capítulo, a los datos de pérdidas por la chimenea, del ejemplo 11.2 (Tabla 11.7) usando S-PLUS. Para cada estimador se usaron los parámetros de afinación predeterminados en S-PLUS. También en la tabla 11.10 se ve el ajuste por MCO, así como resultados para los estimadores compuestos de Coakley-Hettmansperger y Simpson-Montgomery que se describieron en la sección 11.4.2. El programa de Wilcox [1997] en S-PLUS se usó para obtener sus estimados. Como se dijo antes, los ajustes robustos difieren de los de MCO porque hay varias observaciones atípicas en este conjunto de datos.

http://carlos2524.jimdo.com/

TABLA 11.10 Estimados por regresión robusta, para los datos de pérdidas en la chimenea, para S-PLUS usando los parámetros predeterminados

Estimador	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
MCO	−39.920	0.716	1.295	−0.152
M de Huber	−41.029	0.829	0.928	−0.128
M de Andrews	−42.293	0.928	0.649	−0.112
MM	−37.652	0.798	0.577	−0.067
LMS	−39.250	0.750	0.500	0.000
LTS	−36.292	0.736	0.369	0.008
S	−36.917	0.850	0.431	−0.074
Coakley–Hettmansperger	−36.230	0.709	0.560	−0.020
Simpson–Montgomery	−38.770	0.807	0.647	−0.075

PROBLEMAS

11.1 Explicar la diferencia entre un valor atípico de regresión y un valor atípico residual.

11.2 Formar un conjunto de datos con un solo regresor y $n = 10$ observaciones, en el que haya un valor atípico residual que no sea atípico de regresión, un valor atípico de espacio x o uno atípico de espacio y.

11.3 Describir los casos en los que no es probable que un estimador M produzca resultados satisfactorios. Formar un conjunto de datos con un solo regresor y $n = 10$ observaciones, para ilustrar lo anterior.

11.4 **Continuación del problema 11.3** Formar un conjunto de datos con dos regresores y $n = 10$ observaciones, para ilustrar un caso en el que no sea probable que un estimador M produzca resultados satisfactorios.

11.5 Explique por qué un estimador con un punto de quiebre de 50% puede no dar resultados satisfactorios al ajustar un modelo de regresión.

11.6 Se tiene la distribución continua de probabilidades $f(x)$. Suponer que θ es un parámetro de ubicación desconocido, y que se puede expresar la función de densidad en la forma $f(x - \theta)$ para $-\infty < \theta < \infty$. Sea $x_1, x_2, \ldots, x_n$ una muestra aleatoria de tamaño n, tomada de la función de densidad.

a. Demostrar que el estimador de máxima verosimilitud para θ es la solución de

$$\sum_{i=1}^{n} \psi(x_i - \theta) = 0$$

que maximiza el logaritmo de la función verosimilitud $\ln L(\mu) = \Sigma_{i=1}^{n} \ln f(x_i - \theta)$, siendo $\psi(x) = \rho'(x)$ y $\rho(x) = -\ln f(x)$.

b. Si $f(x)$ es una distribución normal, determinar $\rho(x)$, $\psi(x)$ y el estimador correspondiente de máxima verosimilitud de θ.

c. Si $f(x) = (2\sigma)^{-1}e^{-|x|/\sigma}$ (la distribución doble exponencial), determinar $\rho(x)$ y $\psi(x)$. Demostrar que el estimador de máxima verosimilitud de θ es la mediana de la muestra. Comparar este estimador con el que se determinó en la parte b. ¿Parece que la mediana muestral es un estimador razonable en este caso?

http://carlos2524.jimdo.com/

d. Si $f(x) = [\pi(1 + x^2)]^{-1}$ (la distribución de Cauchy), determinar $\rho(x)$ y $\psi(x)$. ¿Cómo se resolvería $\sum_{i=1}^{n} \psi(x_i - \theta)$ en este caso?

11.7 **Regresión Winsorizada (Yale y Forsythe [1976]).** La winsorización es una técnica de estimación robusta que ordena los valores de la muestra por orden creciente, esto es, $x_{(1)} \leq x_{(2)} \cdots x_{(n)}$, para luego redefinir los valores más extremos (atípicos posibles) como los valores siguientes a los más extremos. Así, si sólo se winsorizan las observaciones máxima y mínima, se igualarían $x'_{(1)} = x_{(2)}$ y $x'_{(n)} = x_{(n-1)}$. Describa cómo se podría aplicar la winsorización al problema de regresión.

11.8 Examinar los datos de pérdidas por la chimenea de la tabla 11.7. Determinar los estimados L_1-norm para los coeficientes de regresión β. Comparar los estimados de los parámetros con los citados en el ejemplo 11.2.

11.9 Determinar estimados robustos de los coeficientes de regresión para los datos de pérdidas por la chimenea, en la tabla 11.7, usando la función $E_{0.3}$ definida en la tabla 11.1. Compare los estimados de los coeficientes con los del ejemplo 11.2. ¿Qué tan efectiva fue la función $E_{0.3}$ para este problema?

11.10 Determinar estimados robustos de los coeficientes de regresión para los datos de pérdidas por la chimenea, en la tabla 11.7, usando la función 17A de Hampel, definida en la tabla 11.1. Comparar los coeficientes de regresión con los del ejemplo 11.2, obtenidos con la función de onda de Andrews. ¿Qué tan efectiva fue la función 17A de Hampel en este problema?

11.11 **Bipeso de Tukey**. Una función ψ muy usada para regresiones robustas es la de bipeso de Tukey, siendo

$$\psi(z) = \begin{cases} z\left[1 - (z/a)^2\right]^2, & |z| \leq a \\ 0, & |z| > a \end{cases}$$

con $a = 5$ o 6, por lo general. Trazar la función ψ para $a = 5$, y describir su comportamiento. ¿Cree el lector que la función bipeso de Tukey produzca resultados parecidos a los de la función de onda de Andrews?

11.12 La Fuerza Aérea de Estados Unidos usa modelos de regresión para estimar costos, aplicación en que casi siempre se presentan valores atípicos. Simpson y Montgomery [1998a] presentan 19 observaciones sobre datos de costo de primeras unidades de satélite (y) y el peso del sistema electrónico (x). Los datos se ven en la tabla siguiente.

http://carlos2524.jimdo.com/

Observación	Costo, $/1 000	Peso (lb)
1	2449	90.6
2	2248	87.8
3	3545	38.6
4	794	28.6
5	1619	28.9
6	2079	23.3
7	918	21.1
8	1231	17.5
9	3641	27.6
10	4314	39.2
11	2628	34.9
12	3989	46.6
13	2308	80.9
14	376	14.6
15	5428	48.1
16	2786	38.1
17	2497	73.2
18	5551	40.8
19	5208	44.6

a. Trazar un diagrama de dispersión de los datos. Describir qué clases de valores atípicos pueden estar presentes.

b. Ajustar una recta a esos datos, mediante mínimos cuadrados ordinarios. ¿Parece satisfactorio este ajuste?

c. Ajustar una recta a estos datos con un estimador M que elija el lector. ¿Es satisfactorio el ajuste? Describa por qué, para este problema, el estimador M es mala opción.

d. Ajustar una recta a esos datos, mediante mínima mediana de cuadrados (MMC). ¿Es satisfactorio el ajuste? Describir la diferencia entre los resultados con MMC y los obtenidos en las partes b y c de este problema.

e. Describir las clases de estimadores que es posible sean adecuados para este conjunto de datos.

11.13 Examinar los datos del tiempo de entrega de bebidas gaseosas, presentados en el ejemplo 3.1. Determine el ajuste para estos datos mediante mínimos cuadrados recortados. Comparar los resultados con los estimados de parámetro obtenidos usando diversos estimadores M en el ejemplo 11.1.

11.14 Examinar los tiempos de entrega de bebidas gaseosas que se presentaron en el ejemplo 3.1. Determinar el ajuste por mínima mediana de cuadrados (MMC) para estos datos. Comparar los resultados con los estimados de parámetro obtenidos con diversos estimadores M en el ejemplo 11.1.

11.15 Examinar los datos de regresión de la tabla 11.9.

a. Determinar el ajuste por mínimos cuadrados ordinarios y trazar la línea en el diagrama de dispersión.

b. Determinar el ajuste por mínima mediana de cuadrados y trazar la línea en el diagrama de dispersión.

c. Determinar el ajuste por mínimos cuadrados recortados y trazar la recta en el diagrama de dispersión.

http://carlos2524.jimdo.com/

d. Comentar los resultados obtenidos. ¿Confirma este ejercicio los argumentos del texto, acerca de estimadores de punto de quiebre alto (Sec. 11.4.1)?

11.16 Explicar por qué un estimador de punto de quiebre alto podría ser de utilidad en la primera etapa de un procedimiento de regresión robusta compuesta. Explicar por qué la segunda etapa debe ser un estimador de influencia limitada.

11.17 La tabla B.14 del apéndice muestra datos sobre los puntos transitorios de un inversor electrónico. Ajustar un modelo a esos datos, usando un estimador M. ¿Hay indicios de que las observaciones se hayan anotado en forma incorrecta?

11.18 **Continuación del problema 11.17.** Ajustar un modelo a los datos del inversor, usando mínima mediana de cuadrados. Comparar los resultados con los obtenidos con el estimador M, en el problema 11.17.

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

CAPÍTULO 12

INTRODUCCIÓN A LA REGRESIÓN NO LINEAL

Los modelos de regresión lineal proporcionan un marco rico y flexible que se adapta a las necesidades de muchos analistas, sin embargo, los modelos de regresión lineal no son adecuados para todas las situaciones. Hay muchos problemas en ingeniería y en ciencias donde la variable de respuesta y las variables predictoras se relacionan a través de una función **no lineal** conocida. Esto conduce a un **modelo de regresión no lineal**. Cuando se aplica el método de mínimos cuadrados a esos modelos, las ecuaciones normales que resultan son no lineales y, en general, son difíciles de resolver. El método acostumbrado es minimizar en forma directa la suma de cuadrados de residuales mediante un procedimiento iterativo. En este capítulo se describirá la estimación de los parámetros de un modelo de regresión no lineal y se indicará cómo hacer las inferencias adecuadas acerca de los parámetros del modelo, también ilustraremos programas de cómputo para regresión no lineal.

12.1 MODELOS DE REGRESIÓN LINEALES Y NO LINEALES

12.1.1 Modelos de regresión lineales

En los capítulos anteriores se ha enfocado la atención al **modelo de regresión lineal**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{12.1}$$

Entre estos modelos se incluyen no sólo las relaciones de primer orden, como la ecuación 12.1, sino también los modelos polinomiales y otras relaciones más complejas. De hecho, el modelo de regresión lineal se podría escribir como sigue:

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \cdots + \beta_r z_r + \varepsilon \tag{12.2}$$

en donde z_i representa cualquier **función** de los regresores originales $x_1, x_2, \ldots, x_k$, incluyendo transformaciones como $\exp(x_i)$, $\sqrt{x_i}$ y $\operatorname{sen}(x_i)$. Esos modelos se llaman modelos **lineales** de regresión, porque son **lineales en los parámetros desconocidos**, las β_j, $j = 1, 2, \ldots, k$.

El modelo de regresión lineal (12.1) se puede escribir en su forma general siguiente:

$$\begin{aligned} y &= \mathbf{x}'\boldsymbol{\beta} + \varepsilon \\ &= f(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon \end{aligned} \tag{12.3}$$

http://carlos2524.jimdo.com/

donde $\mathbf{x}' = [1, x_1, x_2, \ldots, x_k]$. Como el valor esperado de los errores del modelo es cero, el valor esperado de la variable de respuesta es

$$\begin{aligned} E(y) &= E[f(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon] \\ &= f(\mathbf{x}, \boldsymbol{\beta}) \end{aligned}$$

Se acostumbra llamar **función del valor esperado** a $f(\mathbf{x}, \boldsymbol{\beta})$ para el modelo. Es obvio que en este caso la función del valor esperado no es más que una función lineal de los parámetros desconocidos.

12.1.2 Modelos de regresión no lineales

Hay muchas situaciones en las que puede no ser lo adecuado un modelo lineal de regresión, por ejemplo, cuando se tiene el conocimiento directo de la forma de la relación entre la variable de respuesta y los regresores, quizá con base en la teoría que gobierna los fenómenos. La relación real entre la respuesta y los regresores puede ser una ecuación diferencial, o su solución; con frecuencia, esa ecuación conducirá a un modelo de forma no lineal.

Cualquier modelo que no sea lineal en los parámetros desconocidos es un **modelo de regresión no lineal**. Por ejemplo, el modelo

$$y = \theta_1 e^{\theta_2 x} + \varepsilon \tag{12.4}$$

es no lineal en los parámetros desconocidos θ_1 y θ_2. Aquí se usará el símbolo θ para representar un parámetro en un modelo no lineal, para subrayar la diferencia entre el caso lineal y el no lineal.

En general, se escribirá el modelo de regresión no lineal en la forma

$$y = f(\mathbf{x}, \theta) + \varepsilon \tag{12.5}$$

en donde θ es un vector de $p \times 1$, de parámetros desconocidos, y ε es un error aleatorio no correlacionado cuyo $E(\varepsilon) = 0$, y cuya $\mathrm{Var}(\varepsilon) = \sigma^2$. También, se supondrá habitualmente que los errores tienen distribución normal, como en la regresión lineal. Como

$$\begin{aligned} E(y) &= E[f(\mathbf{x}, \boldsymbol{\theta}) + \varepsilon] \\ &= f(\mathbf{x}, \boldsymbol{\theta}) \end{aligned} \tag{12.6}$$

a la función $f(\mathbf{x}, \theta)$ se le llama **función de valor esperado** para el modelo de regresión no lineal. Esto se parece mucho al caso de la regresión lineal, excepto que ahora la función de valor esperado es una función **no lineal** de los parámetros.

En un modelo de regresión no lineal, al menos una de las derivadas de la función de valor esperado con respecto a los parámetros depende de cuando menos uno de los parámetros. En la regresión lineal, esas derivadas **no** son funciones de los parámetros desconocidos. Para ilustrar lo anterior, considérese un modelo de regresión lineal

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

http://carlos2524.jimdo.com/

cuya función de valor esperado $f(\mathbf{x}, \beta) = \beta_0 + \Sigma_{j=1}^{k} \beta_j x_j$. Ahora bien,

$$\frac{\partial f(\mathbf{x}, \boldsymbol{\beta})}{\partial \beta_j} = x_j, \qquad j = 0, 1, \ldots, k$$

siendo $x_0 \equiv 1$. Obsérvese que en el caso lineal las derivadas **no** son funciones de las β.

Ahora considérese el modelo no lineal

$$\begin{aligned} y &= f(x, \boldsymbol{\theta}) + \varepsilon \\ &= \theta_1 e^{\theta_2 x} + \varepsilon \end{aligned}$$

Las derivadas de la función de valor esperado con respecto a θ_1 y θ_2 son

$$\frac{\partial f(x, \boldsymbol{\theta})}{\partial \theta_1} = e^{\theta_2 x}$$

y

$$\frac{\partial f(x, \boldsymbol{\theta})}{\partial \theta_2} = \theta_1 x e^{\theta_2 x}$$

Como las derivadas son función de los parámetros desconocidos θ_1 y θ_2, el modelo es no lineal.

12.2 MÍNIMOS CUADRADOS NO LINEALES

Supóngase que una muestra de n observaciones de la respuesta y los regresores es y_i, x_{i1}, $x_{i2}, \ldots, x_{ik}$ para $i = 1, 2, \ldots, n$. Se ha observado antes que el método de mínimos cuadrados en la regresión lineal implica minimizar la función de mínimos cuadrados

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i - \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \right) \right]^2$$

Como éste es un modelo lineal de regresión, al diferenciar $S(\beta)$ con respecto a los parámetros desconocidos e igualar a cero las derivadas, las ecuaciones normales que resultan son ecuaciones **lineales** y, en consecuencia, son fáciles de resolver.

Ahora véase el caso de la regresión no lineal. El modelo es

$$y_i = f(\mathbf{x}_i, \theta) + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

en donde ahora $\mathbf{x}_i' = [1, x_{i1}, x_{i2}, \ldots, x_{ik}]$ para $i = 1, 2, \ldots, n$. La función de mínimos cuadrados es

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]^2 \tag{12.7}$$

http://carlos2524.jimdo.com/

Para determinar los estimados por mínimos cuadrados se debe diferenciar la ecuación (12.7) con respecto a cada elemento de θ. Con ello se obtiene un conjunto de p ecuaciones normales para el caso de regresión no lineal. Las ecuaciones normales son

$$\sum_{i=1}^{n} [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})] \left[\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_j} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \qquad \text{para } j = 1, 2, \ldots, p \tag{12.8}$$

En un modelo de regresión no lineal, las derivadas entre corchetes grandes serán funciones de los parámetros desconocidos. Además, la función del valor esperado también es no lineal, por lo que las ecuaciones normales pueden ser muy difíciles de resolver.

Ejemplo 12.1

Para el modelo no lineal de regresión, en la ecuación (12.4):

$$y = \theta_1 e^{\theta_2 x} + \varepsilon$$

Las ecuaciones normales de mínimos cuadrados para este modelo son

$$\sum_{i=1}^{n} \left[y_i - \hat{\theta}_1 e^{\hat{\theta}_2 x_i} \right] e^{\hat{\theta}_2 x_i} = 0$$

$$\sum_{i=1}^{n} \left[y_i - \hat{\theta}_1 e^{\hat{\theta}_2 x_i} \right] \hat{\theta}_1 x_i e^{\hat{\theta}_2 x_i} = 0 \tag{12.9}$$

Después de simplificar, las ecuaciones normales son

$$\sum_{i=1}^{n} y_i e^{\hat{\theta}_2 x_i} - \hat{\theta}_1 \sum_{i=1}^{n} e^{2\hat{\theta}_2 x_i} = 0$$

$$\sum_{i=1}^{n} y_i x_i e^{\hat{\theta}_2 x_i} - \hat{\theta}_1 \sum_{i=1}^{n} x_i e^{2\hat{\theta}_2 x_i} = 0 \tag{12.10}$$

Estas ecuaciones son no lineales en $\hat{\theta}_1$ y $\hat{\theta}_2$, y no existe solución sencilla de forma cerrada. En general, se deben usar **métodos iterativos** para determinar los valores de $\hat{\theta}_1$ yθ_2. Para complicar más el problema, a veces hay varias soluciones de las ecuaciones normales, lo que nos lleva a que hay diferentes valores estacionarios para la función suma de cuadrados de residuales, $S(\theta)$.

La geometría de los mínimos cuadrados lineales y no lineales

Es bueno examinar la geometría del problema de mínimos cuadrados para comprender las complicaciones que introduce un modelo no lineal. Para determinada muestra, la función suma de cuadrados de residuales $S(\theta)$ sólo depende de los parámetros del modelo θ. Así, en el espacio de parámetros, que es el espacio definido por $\theta_1, \theta_2, \ldots, \theta_p$, se puede representar la función $S(\theta)$ con una gráfica de curvas de nivel, en la que cada curva de nivel en la superficie es una línea de suma constante de residuales al cuadrado.

http://carlos2524.jimdo.com/

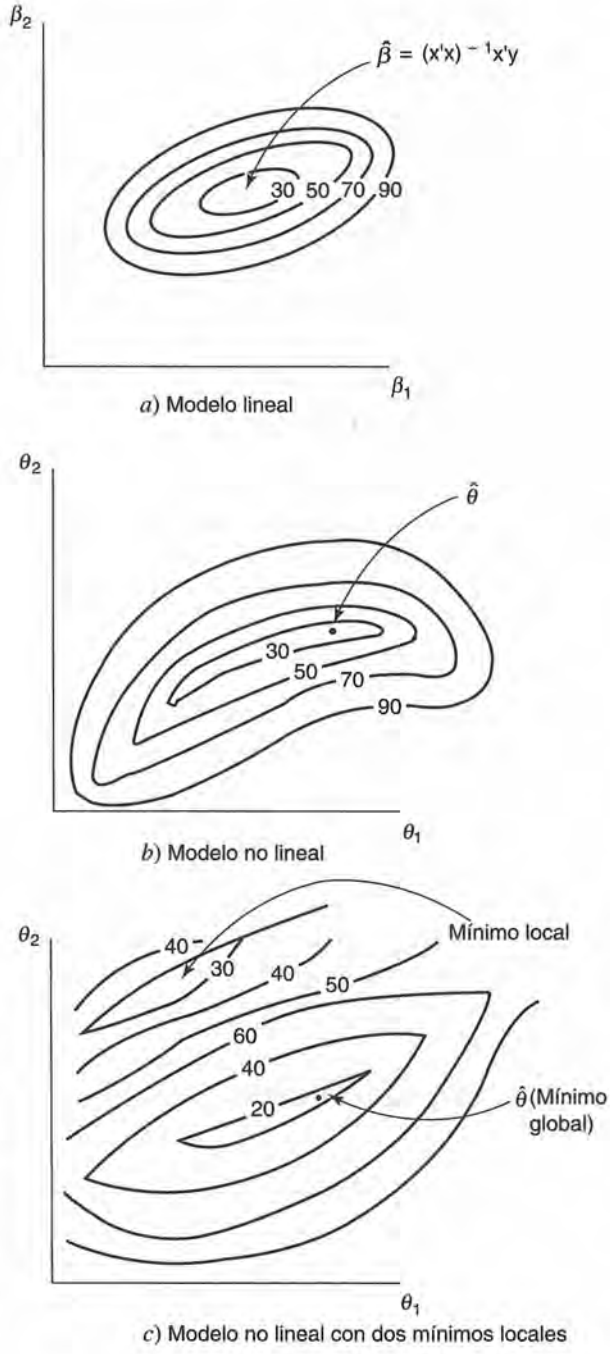


Figura 12.1 Curvas de nivel de la función suma de cuadrados de residuales: a) modelo lineal; b) modelo no lineal; c) modelo no lineal con mínimos locales y globales.

Supóngase que el modelo de regresión es lineal, esto es, que los parámetros son $\theta = \beta$, y que la suma de cuadrados de residuales es $S(\beta)$. La figura 12.1a muestra la gráfica de curvas de nivel para este caso. Si el modelo es lineal en los parámetros desconocidos, las curvas de nivel son elipsoidales, y tienen un mínimo global único en el estimador $\hat{\beta}$ por mínimos cuadrados.

http://carlos2524.jimdo.com/

Cuando el modelo es no lineal, las curvas de nivel suelen aparecer como en la figura 12.1b; nótese que esas curvas no son elípticas, y de hecho son muy alargadas y de forma irregular; es muy común una apariencia "aplatanada". La forma y la orientación específica de las curvas de nivel de suma de cuadrados de residuales dependen de la forma del modelo no lineal, y de la muestra de datos que se haya obtenido. Frecuentemente la superficie será muy alargada cerca del valor óptimo, por lo que muchas soluciones de θ producirán una suma de cuadrados de residuales cercana al óptimo global. Esto causa un problema, que es **deteriorado**, y en tales problemas con frecuencia es difícil determinar el máximo global de θ. En algunos casos, las curvas de nivel son tan irregulares que hay varios mínimos locales, y quizá más de un mínimo global. La figura 12.1c muestra un caso en el que hay un mínimo local y un mínimo global.

Estimación de máxima verosimilitud

Nos hemos concentrado en los mínimos cuadrados para el caso no lineal. Si los términos de error en el modelo están normal e independientemente distribuidos con varianza constante, la aplicación de la máxima verosimilitud al problema de estimación conducirá a los mínimos cuadrados. Por ejemplo, considérese el modelo de la ecuación (12.4):

$$y_i = \theta_1 e^{\theta_2 x_i} + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{12.11}$$

Si los errores están normal e independientemente distribuidos, con media cero y varianza σ^2, entonces la función de verosimilitud es

$$L(\theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[y_i - \theta_1 e^{\theta_2 x_i}\right]^2\right] \tag{12.12}$$

Es claro que maximizar esta función de verosimilitud equivale a minimizar la suma de cuadrados de residuales. En consecuencia, en el caso de la teoría normal, los estimados por mínimos cuadrados son iguales a los estimados por máxima verosimilitud.

12.3 TRANSFORMACIÓN A UN MODELO LINEAL

A veces es útil considerar una **transformación** que induzca la linealidad en la función de valor esperado del modelo. Por ejemplo, se tiene el modelo

$$\begin{aligned} y &= f(x, \theta) + \varepsilon \\ &= \theta_1 e^{\theta_2 x} + \varepsilon \end{aligned} \tag{12.13}$$

Ahora bien, como $E(y) = f(x, \theta) = \theta_1 e^{\theta_2 x}$ se puede linealizar la función del valor esperado sacando logaritmos

$$\ln E(y) = \ln \theta_1 + \theta_2 x$$

http://carlos2524.jimdo.com/

Consecuentemente es muy atractivo reformular el modelo como sigue:

$$\begin{aligned} \ln y &= \ln \theta_1 + \theta_2 x + \varepsilon \\ &= \beta_0 + \beta_1 x + \varepsilon \end{aligned} \tag{12.14}$$

y usar la regresión **lineal** simple para estimar β_0 y β_1, sin embargo, los estimados por mínimos cuadrados de los parámetros de la ecuación (12.14) no serán, en general, equivalentes a los estimados no lineales de parámetros en el modelo original, ecuación (12.13). La razón es que en el **modelo no lineal original**, los mínimos cuadrados implican la minimización de la suma de los residuales al cuadrado respecto a y, mientras que en el **modelo transformado**, ecuación (12.14), se está minimizando la suma de los residuales al cuadrado respecto a $\ln y$.

Nótese que en la ecuación (12.13) la estructura del error es **aditiva**, por lo que el sacar logaritmos **no puede** producir el modelo de la ecuación (12.14). Si la estructura del error es **multiplicativa**, por ejemplo

$$y = \theta_1 e^{\theta_2 x} \varepsilon \tag{12.15}$$

entonces sí será correcto sacar logaritmos, porque

$$\begin{aligned} \ln y &= \ln \theta_1 + \theta_2 x + \text{en } \varepsilon \\ &= \beta_0 + \beta_1 x + \varepsilon^* \end{aligned} \tag{12.16}$$

y si ε^* sigue una distribución normal, se aplicarán todas las propiedades del modelo normal de regresión lineal.

Un modelo no lineal que se puede transformar en una forma lineal equivalente se llama **intrínsecamente lineal**, sin embargo, lo importante suele implicar la estructura de los errores: ¿se aplican las suposiciones estándar de los errores al modelo no lineal original o al linealizado? Esta pregunta a veces no es fácil de contestar.

Ejemplo 12.2 *Datos de puromicina*

Bates y Watts [1988] usan el modelo de **Michaelis-Menten** de cinética química, para relacionar la velocidad inicial de una reacción enzimática con la concentración x del sustrato. El modelo es

$$y = \frac{\theta_1 x}{x + \theta_2} + \varepsilon \tag{12.17}$$

Los datos de velocidad inicial de reacción para una enzima tratada con puromicina se ven en la tabla 12.1 y se grafican en la figura 12.2

Se ve que la función se puede linealizar con facilidad, ya que

$$\begin{aligned} \frac{1}{f(x, \theta)} &= \frac{x + \theta_2}{\theta_1 x} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1} \frac{1}{x} \\ &= \beta_0 + \beta_1 u \end{aligned}$$

http://carlos2524.jimdo.com/

TABLA 12.1 Velocidad de reacción y concentración de sustrato en el experimento de puromicina

Concentración del sustrato (ppm)	Velocidad [(cuentas/min)/min]	
0.02	47	76
0.06	97	107
0.11	123	139
0.22	152	159
0.56	191	201
1.10	200	207

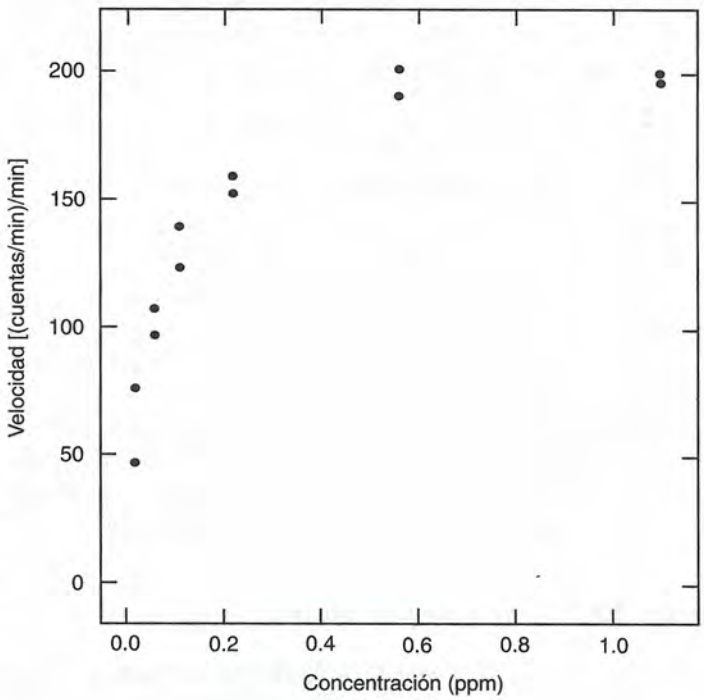


Figura 12.2 Gráfica de velocidad de reacción en función de concentración de sustrato, para el experimento de puromicina. (Adaptado de Bates y Watts [1988], con autorización del editor.)

por lo que se puede tratar de ajustar el modelo **lineal**

$$y^* = \beta_0 + \beta_1 u + \varepsilon$$

en donde $y^* = 1/y$ y $u = 1/x$. El ajuste resultante por mínimos cuadrados es

$$\hat{y}^* = 0.005107 + 0.0002472u$$

La figura 12.3a muestra un diagrama de dispersión de los datos transformados y^* y u, con la recta superpuesta. Como hay réplicas en los datos, es fácil ver, en la figura 12.2, que la varianza de los datos originales es aproximadamente constante, mientras que según la figura 12.3a, la suposición de varianza constante en la escala transformada es irrazonable.

http://carlos2524.jimdo.com/

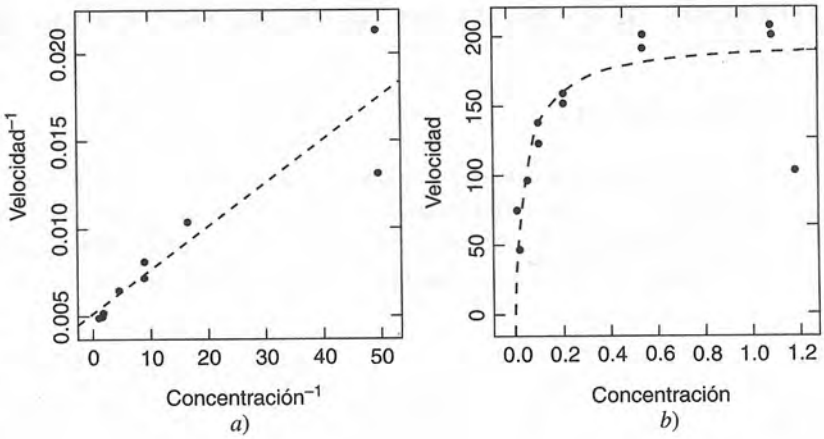


Figura 12.3 a) Gráfica del inverso de la velocidad en función del inverso de la concentración, para los datos de puromicina. b) Curva ajustada en la escala original. (Adaptado de Bates y Watts [1988], con autorización del editor.)

Ahora bien, como

$$\beta_0 = \frac{1}{\theta_1}$$

y

$$\beta_1 = \frac{\theta_2}{\theta_1}$$

entonces

$$0.005107 = \frac{1}{\hat{\theta}_1}$$

$$0.0002472 = \frac{\hat{\theta}_2}{\hat{\theta}_1}$$

por lo que se pueden estimar θ_1 y θ_2 en el modelo original como sigue:

$$\hat{\theta}_1 = 195.81$$

y

$$\hat{\theta}_2 = 0.04841$$

La figura 12.3b muestra la curva ajustada en la escala original, junto con los datos. Obsérvese en la figura que la asíntota ajustada es demasiado baja. La varianza en los puntos replicados ha sido distorsionada por la transformación, por lo que las corridas con baja concentración (alta concentración recíproca) dominan el ajuste por mínimos cuadrados, y en consecuencia el modelo no se ajusta bien a los datos cuando las concentraciones son altas.

http://carlos2524.jimdo.com/

12.4 ESTIMACIÓN DE PARÁMETROS EN UN SISTEMA NO LINEAL

12.4.1 Linealización

Un método muy usado en los algoritmos de cómputo para regresión no lineal es la **linealización** de la función no lineal, seguida por el método iterativo de Gauss-Newton para estimar parámetros. La linealización se hace con un **desarrollo en serie de Taylor** de $f(\mathbf{x}_i, \theta)$ respecto al punto $\theta'_0 = [\theta_{10}, \theta_{20}, \ldots, \theta_{p0}]$, reteniendo sólo los términos lineales. Así se obtiene

$$f(\mathbf{x}_i, \boldsymbol{\theta}) = f(\mathbf{x}_i, \boldsymbol{\theta}_0) + \sum_{j=1}^{p} \left[\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_j} \right]_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} (\theta_j - \theta_{j0}) \tag{12.18}$$

Si se definen

$$f_i^0 = f(\mathbf{x}_i, \boldsymbol{\theta}_0)$$

$$\beta_j^0 = \theta_j - \theta_{j0}$$

$$Z_{ij}^0 = \left[\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_j} \right]_{\boldsymbol{\theta} = \boldsymbol{\theta}_0}$$

se nota que el modelo de regresión no lineal se puede escribir en la forma

$$y_i - f_i^0 = \sum_{j=1}^{p} \beta_j^0 Z_{ij}^0 + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{12.19}$$

Esto es, ahora se cuenta con un modelo de regresión lineal. Se acostumbra llamar θ_0 a los valores iniciales de los parámetros.

La ecuación (12.19) se puede escribir en la forma

$$\mathbf{y}_0 = \mathbf{Z}_0 \beta_0 + \varepsilon \tag{12.20}$$

por lo que el estimado de β_0 es

$$\begin{aligned} \hat{\beta}_0 &= (\mathbf{Z}'_0 \mathbf{Z}_0)^{-1} \mathbf{Z}'_0 \mathbf{y}_0 \\ &= (\mathbf{Z}'_0 \mathbf{Z}_0)^{-1} \mathbf{Z}'_0 (\mathbf{y} - \mathbf{f}_0) \end{aligned} \tag{12.21}$$

Ahora bien, como $\beta_0 = \theta - \theta_0$, se puede definir

$$\hat{\theta}_1 = \hat{\beta}_0 + \theta_0 \tag{12.22}$$

como los estimados corregidos de θ. A veces a $\hat{\beta}_0$ se le llama **vector de incrementos**. Ahora se pueden reemplazar los estimados corregidos θ_1 en la ecuación (12.18) (en los mismos papeles que desempeñaban los estimados iniciales θ_0), para a continuación producir otro conjunto de estimados corregidos, por ejemplo, $\hat{\theta}_2$, y así sucesivamente.

http://carlos2524.jimdo.com/

En general, en la k-ésima iteración se tiene

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{k+1} &= \hat{\boldsymbol{\theta}}_k + \hat{\boldsymbol{\beta}}_k \\ &= \hat{\boldsymbol{\theta}}_k + (\mathbf{Z}'_k\mathbf{Z}_k)^{-1}\mathbf{Z}'_k(\mathbf{y} - \mathbf{f}_k)\end{aligned} \tag{12.23}$$

siendo

$$\begin{aligned}\mathbf{Z}_k &= \left[Z_{ij}^k\right] \\ \mathbf{f}_k &= \left[f_1^k, f_2^k, \ldots, f_n^k\right]' \\ \hat{\boldsymbol{\theta}}_k &= \left[\theta_{1k}, \theta_{2k}, \ldots, \theta_{pk}\right]'\end{aligned}$$

Este procedimiento de iteración continúa hasta la convergencia, esto es, hasta que

$$\left|\left(\hat{\theta}_{j,\,k+1} - \hat{\theta}_{jk}\right)/\hat{\theta}_{jk}\right| < \delta, \qquad j = 1, 2, \ldots, p$$

en donde δ es un número pequeño, por ejemplo 1.0×10^{-6}. En cada iteración se debe evaluar la suma de cuadrados de residuales, $S(\hat{\theta}_k)$, para asegurar que se haya obtenido una reducción en su valor.

Ejemplo 12.3 *Datos de puromicina*

Bates y Watts [1988] aplican el método de Gauss-Newton para ajustar el modelo de Michaelis-Menten a los datos de la puromicina, en la tabla 12.1, con los valores iniciales $\theta_{10} = 205$ y $\theta_{20} = 0.08$. Después se describirá cómo se obtuvieron esos valores iniciales. En este punto de partida, la suma de cuadrados de residuales es $S(\theta_0) = 3\ 155$; los datos, los valores ajustados, los residuales y las derivadas, evaluadas en cada observación, se ven en la tabla 12.2. Para ilustrar cómo se calcularon las cantidades necesarias, obsérvese que

$$\frac{\partial f(x, \theta_1, \theta_2)}{\partial \theta_1} = \frac{x}{\theta_2 + x}$$

y que

$$\frac{\partial f(x, \theta_1, \theta_2)}{\partial \theta_2} = \frac{-\theta_1 x}{(\theta_2 + x)^2}$$

y como la primera observación de x es $x_1 = 0.02$, entonces

$$Z_{11}^0 = \left.\frac{x_1}{\theta_2 + x}\right|_{\theta_2 = 0.08} = \frac{0.02}{0.08 + 0.02} = 0.2000$$

$$Z_{12}^0 = \left.\frac{-\theta_1 x_1}{(\theta_2 + x_1)^2}\right|_{\theta_1 = 205,\, \theta_2 = 0.08} = \frac{(-205)(0.02)}{(0.08 + 0.02)^2} = -410.00$$

http://carlos2524.jimdo.com/

TABLA 12.2 Datos, valores ajustados y derivadas para el experimento de puromicina en $\hat{\theta}_0' = [205, 0.08]'$

i	x_i	y_i	f_i^0	$y_i - f_i^0$	Z_{i1}^0	Z_{i2}^0
1	0.02	76	41.00	35.00	0.2000	−410.00
2	0.02	47	41.00	6.00	0.2000	−410.00
3	0.06	97	87.86	9.14	0.4286	−627.55
4	0.06	107	87.86	19.14	0.4286	−627.55
5	0.11	123	118.68	4.32	0.5789	−624.65
6	0.11	139	118.68	20.32	0.5789	−624.65
7	0.22	159	150.33	8.67	0.7333	−501.11
8	0.22	152	150.33	1.67	0.7333	−501.11
9	0.56	191	179.38	11.62	0.8750	−280.27
10	0.56	201	179.38	21.62	0.8750	−280.27
11	1.10	207	191.10	15.90	0.9322	−161.95
12	1.10	200	191.10	8.90	0.9322	−161.95

Ahora se agrupan las derivadas Z_{ij}^0 en la matriz $\mathbf{Z}_0$ y el vector de incrementos se calcula con la ecuación (12.21), resultando

$$\hat{\boldsymbol{\beta}}_0 = \begin{bmatrix} 8.03 \\ -0.017 \end{bmatrix}$$

El estimado corregido $\hat{\theta}_1$ de la ecuación (12.22) es

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1 &= \hat{\boldsymbol{\beta}}_0 + \boldsymbol{\theta}_0 \\ &= \begin{bmatrix} 8.03 \\ -0.017 \end{bmatrix} + \begin{bmatrix} 205.00 \\ 0.08 \end{bmatrix} \\ &= \begin{bmatrix} 213.03 \\ 0.063 \end{bmatrix} \end{aligned}$$

La suma de cuadrados de residuales, en este punto, es $S(\hat{\theta}_1) = 1\,206$, bastante menor que $S(\theta_0)$. Por consiguiente, se adopta $\hat{\theta}_1$ como el estimado corregido de θ y se corre otra iteración.

El algoritmo de Gauss-Newton convergió hacia $\hat{\theta}' = [212.7, 0.0641]'$, con $S(\hat{\theta}) = 1\,195$. Por consiguiente, el modelo ajustado obtenido por linealización es

$$\hat{y} = \hat{\theta}_1 e^{\hat{\theta}_2 x} = 212.7 e^{0.0641 x}$$

La figura 12.4 muestra el modelo ajustado. Obsérvese que este modelo no lineal produce un ajuste mucho mejor a los datos que la transformación seguida por regresión lineal, en el ejemplo 12.2 (compárense las Figs. 12.4 y 12.3b).

Los residuales se pueden obtener a partir de un modelo no lineal de regresión en la forma acostumbrada, esto es

$$e_i = y_i - \hat{y}_i \qquad i = 1, 2, \ldots, n$$

http://carlos2524.jimdo.com/

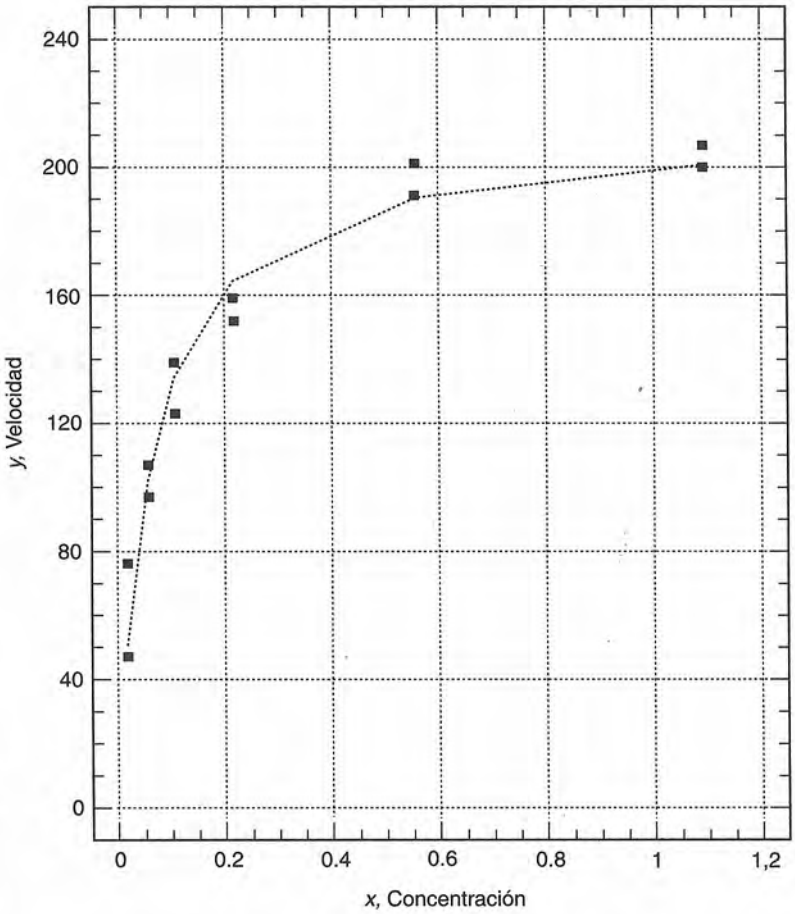


Figura 12.4 Gráfica del modelo ajustado de regresión no lineal, ejemplo 12.3.

En este ejemplo, los residuales se calculan a partir de

$$e_i = y_i - \hat{\theta}_1 e^{\hat{\theta}_2 x_i} = y_i - 212.7e^{0.0641 x_i}, \qquad i = 1, 2, \ldots, 10$$

Los residuales se grafican en función de los valores predichos o calculados, en la figura 12.5. En la figura 12.6 se ve una gráfica de probabilidad normal de los residuales. Hay un residual moderadamente grande; sin embargo, el ajuste general es satisfactorio y el modelo parece ser una mejora sustancial respecto al obtenido por el método de la transformación, en el ejemplo 12.2.

Estimación de σ^2

Cuando el procedimiento de estimación converge hacia un vector final de estimados de parámetro $\hat{\theta}$, se puede obtener un estimado de la varianza σ^2 del error a partir del cuadrado medio residual

$$\hat{\sigma}^2 = MS_{\text{Res}} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - p} = \frac{\sum_{i=1}^{n} \left[y_i - f(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\right]^2}{n - p} = \frac{S(\hat{\boldsymbol{\theta}})}{n - p} \qquad (12.24)$$

http://carlos2524.jimdo.com/

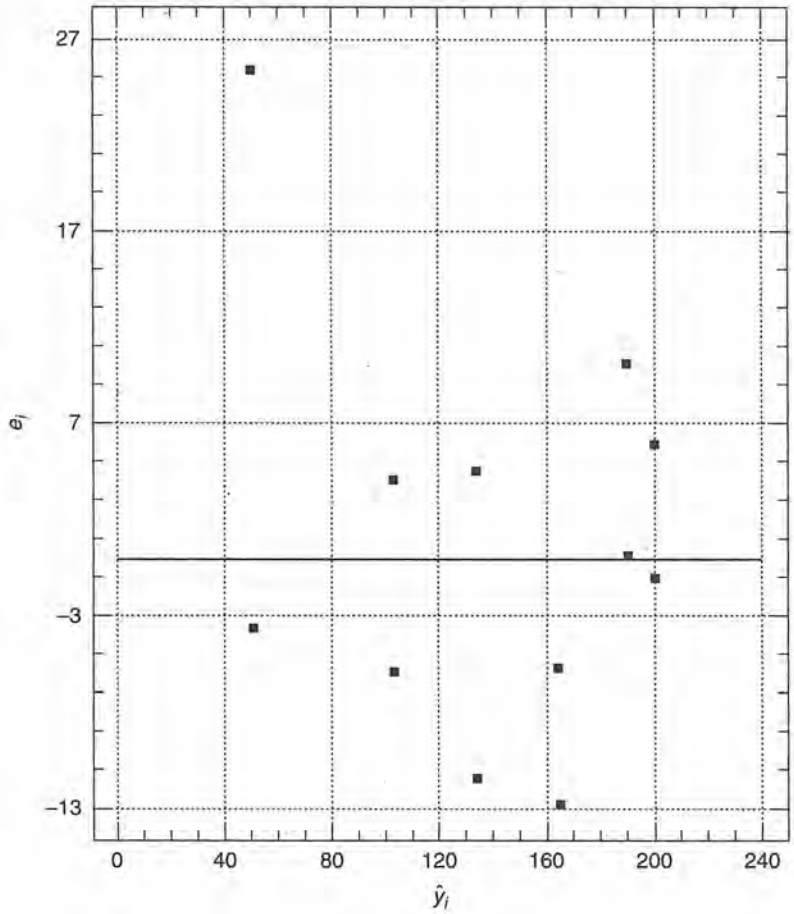


Figura 12.5 Gráfica de los residuales en función de los valores predichos, ejemplo 12.3.

en donde p es la cantidad de parámetros en el modelo de regresión no lineal. Para los datos de puromicina del ejemplo 12.3 se vio que la suma de cuadrados de residuales en la iteración final fue $S(\hat{\theta}) = 1\ 195$, así que el estimado de σ^2 es

$$\hat{\sigma}^2 = \frac{S(\hat{\theta})}{n - p} = \frac{1195}{12 - 2} = 1\ 19.5$$

También se puede estimar la **matriz de covarianza asintótica (de muestra grande)** del vector de parámetros $\hat{\theta}$ como sigue:

$$\mathrm{Var}(\hat{\theta}) = \hat{\sigma}^2(\mathbf{Z}'\mathbf{Z})^{-1} \tag{12.25}$$

siendo $\mathbf{Z}$ la matriz de derivadas parciales definida antes, evaluada en el estimado $\hat{\theta}$ de la iteración final por mínimos cuadrados.

La matriz de covarianza del vector $\hat{\theta}$ para el modelo de Michaelis-Menten, en el ejemplo 12.3, es

$$\begin{aligned}\mathrm{Var}(\hat{\theta}) &= \hat{\sigma}^2(\mathbf{Z}'\mathbf{Z})^{-1} \\ &= 119.5\begin{bmatrix} 0.4037 & 36.82 \times 10^{-5} \\ 36.82 \times 10^{-5} & 57.36 \times 10^{-8} \end{bmatrix}\end{aligned}$$

http://carlos2524.jimdo.com/

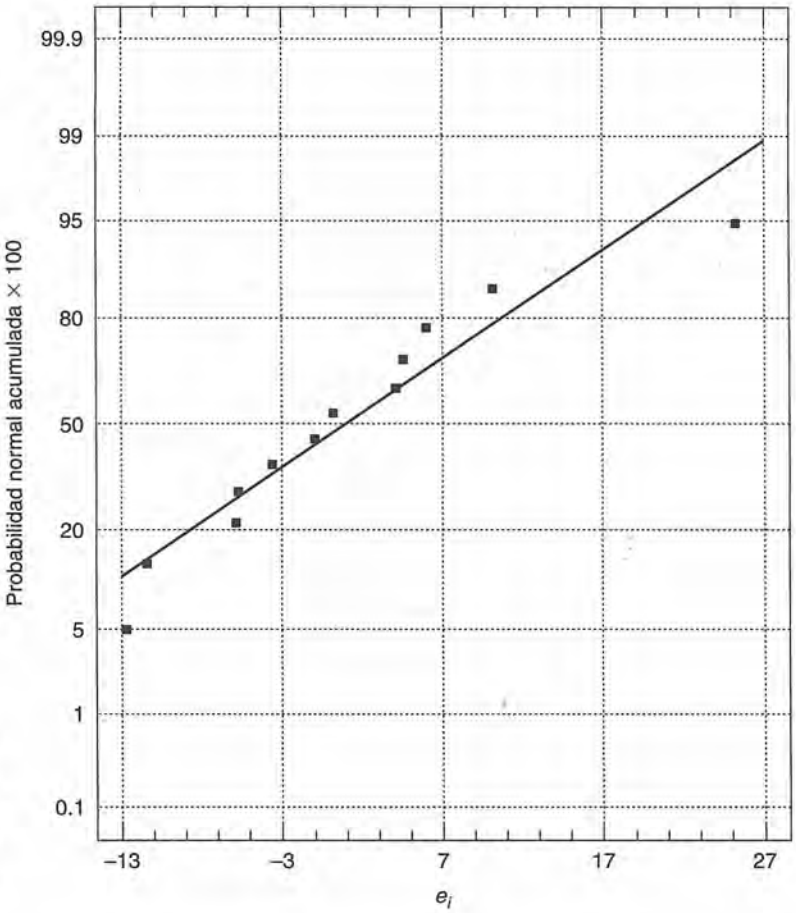


Figura 12.6 Gráfica de probabilidad normal de los residuales, ejemplo 12.3.

Los elementos de la diagonal principal de esta matriz son las varianzas aproximadas de los estimados de los coeficientes de regresión. En consecuencia, los **errores estándar** aproximados en los coeficientes son

$$\text{se}(\hat{\theta}_1) = \sqrt{\text{Var}(\hat{\theta}_1)} = \sqrt{119.5(0.4037)} = 6.95$$

y

$$\text{se}(\hat{\theta}_2) = \sqrt{\text{Var}(\hat{\theta}_2)} = \sqrt{119.5(57.36 \times 10^{-8})} = 8.28 \times 10^{-3}$$

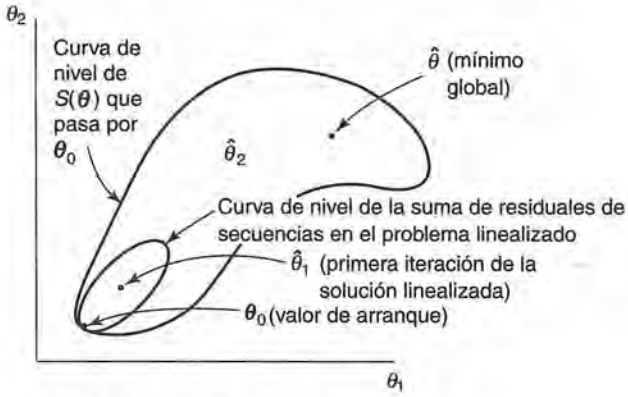
La correlación entre $\hat{\theta}_1$ y $\hat{\theta}_2$ es, más o menos,

$$\frac{36.82 \times 10^{-5}}{\sqrt{0.4037(57.36 \times 10^{-8})}} = 0.77$$

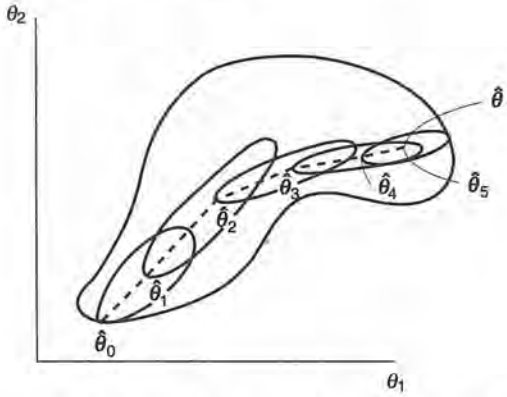
Perspectiva gráfica de la linealización

Se ha observado que la función suma de cuadrados de residuales, $S(\theta)$, para un modelo no lineal de regresión, suele ser "aplatanada" irregular, como se ve en las partes b y c de la

http://carlos2524.jimdo.com/



a) La primera iteración



b) Evolución de sucesivas iteraciones de linealización

Figura 12.7 Interpretación geométrica de la linealización: a) la primera iteración; b) evolución de iteraciones sucesivas de linealización.

figura 12-1. Por otra parte, la función suma de cuadrados de residuales para los cuadrados mínimos lineales se comporta muy bien; de hecho, es elíptica y tiene el mínimo global en el fondo de la "cuenca", véase la figura 12.1a. La técnica de linealización convierte el problema de regresión no lineal en una serie de problemas lineales que comienzan en el punto θ_0.

La primera iteración de la linealización cambia las curvas irregulares de nivel por un conjunto de curvas elípticas. Las curvas irregulares de $S(\theta)$ pasan exactamente a través del punto inicial θ_0, como se ve en la figura 12.7a. Cuando se resuelve el problema linealizado, uno se mueve hacia el mínimo global, sobre el conjunto de curvas de nivel elípticas, esto lo hacen los mínimos cuadrados ordinarios. A continuación, en la siguiente iteración sólo se repite el proceso, comenzando en la nueva solución $\hat{\theta}_1$. La evolución definitiva de la linealización es una secuencia de problemas lineales para los cuales las soluciones "se cierran" hacia el mínimo global de la función no lineal. Esto se ve en la figura 12.7b, siempre y cuando el problema no lineal sea demasiado deteriorado, sea debido a un modelo mal especificado o por datos inadecuados, el procedimiento de linealización debe converger hacia un buen estimado del mínimo global, en unas cuantas iteraciones.

La linealización se facilita con un buen valor inicial θ_0, esto es, uno que sea razonablemente cercano al mínimo global. Cuando θ_0 está cerca de $\hat{\theta}$, las curvas de nivel de la suma real de cuadrados de residuales para el problema no lineal, se aproxima bien, normalmente, mediante las curvas de nivel del problema linealizado. En la sección 12.4.3 se describirá la obtención de los valores iniciales.

http://carlos2524.jimdo.com/

12.4.2 Otros métodos de estimación de parámetros

El método básico de linealización que se describió en la sección 12.4.1 puede converger con mucha lentitud, en algunos problemas; en otros puede generar un movimiento en dirección incorrecta, cuando en realidad la función suma de cuadrados de residuales $S(\hat{\theta}_k)$ **aumente** en la k-ésima iteración. En los casos extremos no convergerá en absoluto, en consecuencia, se han desarrollado otras técnicas para resolver el problema de regresión no lineal; algunas de ellas son modificaciones y refinamientos del esquema de linealización. En esta sección se describirán en forma breve algunos de esos procedimientos.

El método del descenso más pronunciado

Este método trata de determinar el mínimo global de la función suma de cuadrados de residuales mediante minimización directa. El objetivo es pasar desde un punto de partida inicial θ_0, en una dirección del vector cuyos componentes se determinan con las derivadas de la función suma de cuadrados de residuales con respecto a los elementos de θ, por lo general, esas derivadas se estiman ajustando una aproximación de primer orden, o plana, en torno al punto θ_0. Los coeficientes de regresión del modelo de primer orden se toman como aproximaciones de las primeras derivadas.

El método de descenso más pronunciado se usa mucho en la metodología de superficies de respuesta, para pasar de un estimado inicial de las condiciones óptimas de un proceso, a una región que contenga al óptimo con más probabilidad. La desventaja principal de este método de solución del problema de regresión no lineal es que puede converger con mucha lentitud. El descenso más pronunciado suele funcionar mejor cuando el punto de partida está muy lejos del óptimo, sin embargo, a medida que la solución que se obtenga se acerque al óptimo, el procedimiento producirá movimientos cada vez más cortos, en un comportamiento de "zig-zag"; es el problema de convergencia que se mencionó arriba.

Incrementos fraccionados

Una modificación normal a la técnica de linealización es el uso de los **incrementos fraccionados**. Para describir este método sea $\hat{\beta}_k$ el vector normal de incremento de la ecuación (12.23). En la k-ésima iteración sólo se continúa a la siguiente si $S(\hat{\theta}_{k+1}) < S(\hat{\theta}_k)$. Si $S(\hat{\theta}_{k+1}) > S(\hat{\theta}_k)$, se usa $\hat{\beta}_k/2$ como vector de incrementos. Este corte a la mitad se podría usar varias veces durante una iteración, si es necesario. Si después de una cantidad especificada de intentos no se obtiene una reducción en $S(\hat{\theta}_{k+1})$, se termina el procedimiento. El concepto general de este método es evitar que el procedimiento de linealización haga un paso demasiado grande en cualquier iteración. La técnica de los incrementos fraccionados ayuda cuando se encuentran problemas de convergencia en el procedimiento básico de linealización.

Compromiso de Marquardt

Otra modificación conocida del algoritmo básico de linealización fue desarrollada por Marquardt [1963]. Propuso calcular el vector de incrementos en la k-ésima iteración con

$$(\mathbf{Z}_k'\mathbf{Z}_k + \lambda\mathbf{I}_p)\,\hat{\beta}_k = \mathbf{Z}_k'(\mathbf{y} - \mathbf{f}_k) \tag{12.26}$$

http://carlos2524.jimdo.com/

en donde $\lambda > 0$, nótese la semejanza con el estimador de regresión ridge en el capítulo 10. Como las variables regresoras son derivadas de la misma función, la función linealizada invita a la multicolinealidad; por eso es razonable, intuitivamente, el procedimiento parecido al ridge, de la ecuación (12.26). Marquardt [1963] usó un procedimiento de búsqueda para determinar un valor de λ que reduzca la suma de cuadrados de residuales en cada paso.

Los distintos programas de cómputo seleccionan a λ en distintas formas. Por ejemplo, en PROC NLIN del programa SAS se comienza con $\lambda = 10^{-8}$. Se hace una serie de cálculos de tanteo en cada iteración, en donde λ se multiplica repetidamente por 10 hasta que

$$S(\hat{\theta}_{k+1}) < S(\hat{\theta}_k) \tag{12.27}$$

El procedimiento también implica reducir λ por un factor de 10 en cada iteración, siempre que se satisfaga la ecuación (12.27). La estrategia es mantener a λ tan pequeña como sea posible y al mismo tiempo asegurar que la suma de cuadrados de residuales se reduzca en cada iteración. El procedimiento general se llama con frecuencia compromiso de Marquardt, porque el vector de incrementos resultante que se obtiene con este método suele estar entre el vector de Gauss-Newton, y el obtenido con la dirección de descenso más pronunciado.

12.4.3 Valores iniciales

Para ajustar un modelo no lineal de regresión se requieren valores de arranque θ_0 de los parámetros del modelo. Unos buenos valores iniciales, esto es, valores de θ_0 cercanos a los valores verdaderos de los parámetros, minimizarán las dificultades de convergencia. Se han hecho modificaciones al procedimiento de linealización, como el compromiso de Marquardt, con las que el procedimiento es menos sensible a la elección de los valores de arranque, pero siempre se aconseja seleccionar θ_0 con cuidado. Una mala elección podría causar la convergencia hacia un mínimo local de la función, y pasaría completamente desapercibido el que se haya obtenido una solución subóptima.

En los modelos no lineales de regresión, con frecuencia los parámetros tienen algún significado físico, lo cual mucho ayudaría para obtener los valores de arranque. También puede ayudar el graficar la función de valor esperado para varios valores de los parámetros, para familiarizarse con el comportamiento del modelo, y la forma en que los cambios en los parámetros afectan ese comportamiento.

Por ejemplo, en la función de Michaelis-Menten que se usó con los datos de puromicina, el parámetro θ_1 es la velocidad asintótica de la reacción, es decir, el valor máximo de f cuando $x \to \infty$. De igual modo, θ_2 representa la mitad de la concentración, que es el valor de x, tal que cuando la concentración llega a él, la velocidad es la mitad del valor máximo. Al examinar el diagrama de dispersión en la figura 12.2, parecería que $\theta_1 = 205$ y $\theta_2 = 0.08$; son valores iniciales razonables; son los valores que se usaron en el ejemplo 12.3.

En algunos casos se puede transformar la función de valor esperado para obtener los valores de arranque, por ejemplo, el modelo de Michaelis-Menten se puede "linealizar" sacando el recíproco de la función de valor esperado. Los mínimos cuadrados lineales se pueden usar con los datos recíprocos, como se hizo en el ejemplo 12.2, para producir estimados de los parámetros lineales. A continuación, esos estimados se pueden usar para obtener los valores iniciales necesarios de θ_0, también la transformación gráfica puede ser muy efectiva. Un buen ejemplo de lo anterior lo presentan Bates y Watts [1988, p. 47].

http://carlos2524.jimdo.com/

TABLA 12.3 Resultados de la regresión no lineal para los datos de puromicina, obtenidos con STATGRAPHICS

Resultados de ajuste del modelo para: REGRESS.Vel			
	estimado	**error estándar**	**relación**
Coeficiente 1	212.639337	6.95864670	30.5576
Coeficiente 2	0.064049	0.00833586	7.6835
Iteraciones totales = 3	Total de evaluaciones de función = 10		

Análisis de varianza para la regresión completa				
fuente	**suma de cuadrados**	**df**	**cuadrado medio**	**relación**
Modelo	270213.54	2	135106.77	1130.17
Error	1195.4572	10	119.5457	
Total	271409.00	12		
Total (Corr.)	30858.917	11		
R cuadrada = 0.961261				
R cuadrada (Aj. por gdl.) = 0.957387				

12.4.4 Programas de cómputo

Muchos paquetes de programas estadísticos tienen funciones de modelado con regresión no lineal, por ejemplo, el procedimiento PROC NLIN del programa SAS es un programa que se usa ampliamente; emplea el método del compromiso de Marquardt que se reseñó arriba, aunque también tiene otras opciones para estimar parámetros. El programa estima en forma numérica las derivadas necesarias, o permite que el usuario ingrese las derivadas en forma directa. Muchos paquetes estadísticos para PC también tienen posibilidades de ajustar con modelo no lineal de regresión. La tabla 12.3 muestra el resultado de la rutina no lineal de regresión de STATGRAPHICS para los datos de puromicina, en este programa también se usa el método de Marquardt.

Los resultados de computadora en la tabla 12.3, para los datos de puromicina, se parecen mucho a los de un programa estándar de cómputo para regresión lineal. La parte superior de la tabla presenta los estimados de los parámetros, donde los títulos "Coeficiente 1" y "Coeficiente 2" representan a θ_1 y θ_2, respectivamente. Se usaron en estos cálculos los mismos valores iniciales que en el ejemplo 12.3, $\theta_{10} = 205$ y $\theta_{20} = 0.08$. La solución que se determinó es casi idéntica a la obtenida con la iteración de Gauss-Newton en el ejemplo 12.3, también, los resultados muestran los errores estándar de los estimados de los parámetros; esos errores también son muy parecidos a los resultados presentados más arriba.

Nótese que los resultados presentan una "relación" para cada coeficiente, que se calcula como sigue:

$$\frac{\hat{\theta}}{\text{se}(\hat{\theta})} = \frac{\text{estimado del coeficiente}}{\text{error estándar del coeficiente}}$$

En la regresión lineal se reconoce que esto es un estadístico t para probar la hipótesis nula H_0: $\theta = 0$. La parte de abajo de la tabla 12.3 es una tabla de análisis de varianza, también contiene una relación del cuadrado medio del modelo entre el cuadrado medio del

http://carlos2524.jimdo.com/

error, o de los cuadrados de los residuales. En la regresión lineal eso sería el estadístico F para probar el significado de la regresión, es decir, probar H_0: $\theta_1 = \theta_2 = 0$. En la regresión lineal esas relaciones no son los estadísticos t o F exactos, pero se suelen interpretar como estadísticos t o F **aproximados**. La inferencia estadística en el modelo no lineal de regresión se basa en los resultados **asintóticos** o **de muestra grande**. En la siguiente sección se resumirán algunos aspectos útiles de la inferencia para modelos no lineales de regresión.

12.5 INFERENCIA ESTADÍSTICA EN REGRESIÓN NO LINEAL

En un **modelo lineal de regresión**, cuando los errores están normal e independientemente distribuidos, las pruebas estadísticas y los intervalos de coeficientes exactos se pueden conseguir, basados en las distribuciones t y F, y los estimados de parámetro tienen propiedades estadísticas útiles y atractivas. Sin embargo, éste no es el caso con la regresión no lineal, aun cuando los errores estén normal e independientemente distribuidos. Indica esto que, en la regresión no lineal, los estimados por mínimos cuadrados (o por la función de valor esperado), de los parámetros del modelo, no tienen ninguna de las atrayentes propiedades de sus contrapartes en la regresión lineal, como por ejemplo, el insesgo, la varianza mínima o las distribuciones normales de muestreo. La inferencia estadística en la regresión no lineal depende de los resultados con **muestra grande** o **asintóticos**. La teoría de muestras grandes se aplica en general a errores distribuidos normal y no normalmente.

Los resultados asintóticos clave se pueden resumir como sigue: en general, cuando el tamaño n de la muestra es grande, el valor esperado de $\hat{\theta}$ es aproximadamente igual a θ, el vector verdadero de estimados de los parámetros, y la matriz de covarianza de $\hat{\theta}$ es aproximadamente igual a $\sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}$, donde $\mathbf{Z}$ es la matriz de derivadas parciales evaluadas en el estimado $\hat{\theta}$ de la iteración final con mínimos cuadrados, además, la distribución de muestreo de $\hat{\theta}$ es aproximadamente normal, en consecuencia, la inferencia estadística para la regresión no lineal, cuando el tamaño de la muestra es grande, se hace exactamente como en la regresión lineal. Las pruebas estadísticas y los intervalos de confianza sólo son procedimientos aproximados.

Ejemplo 12.4 *Datos de puromicina*

Se regresará al modelo de Michaelis-Menten para los datos de puromicina, del ejemplo 12.3. El resultado de cómputo para el modelo se ve en la tabla 12.3. Para probar el significado de la regresión (esto es, H_0: $\theta_1 = \theta_2 = 0$) se podría usar la relación

$$F_0 = \frac{MS_{\text{Modelo}}}{MS_{\text{Error}}} = \frac{135\,106.77}{119.5457} = 1\,130.17$$

y calcular un valor P aproximado de acuerdo con la distribución $F_{2,\,10}$. Ese valor P es bastante menor que 0.0001 (es $1.66E{-}12$), por lo que es seguro rechazar la hipótesis nula y llegar a la conclusión de que al menos uno de los parámetros del modelo es distinto de cero. Para probar la hipótesis con los parámetros individuales del modelo, H_0: $\theta_1 = 0$ y H_0: $\theta_2 = 0$, se podrían calcular los estadísticos t aproximados como sigue:

$$t_0 = \frac{\hat{\theta}_1}{\text{se}(\hat{\theta}_1)} = \frac{212.639377}{6.95864670} = 30.5576$$

http://carlos2524.jimdo.com/

y

$$t_0 = \frac{\hat{\theta}_2}{\text{se}(\hat{\theta}_2)} = \frac{0.064049}{0.00833586} = 7.6835$$

Los valores P aproximados para esos dos estadísticos de prueba son $1.65E{-}11$ y $8.36E{-}6$, respectivamente, por consiguiente, se llegaría a la conclusión de que ambos parámetros son distintos de cero.

El 95% de los intervalos de confianza aproximados para θ_1 y θ_2 se calculan como sigue:

$$\begin{aligned} \hat{\theta}_1 - t_{0.025,\,10}\text{se}(\hat{\theta}_1) &\le \theta_1 \le \hat{\theta}_1 + t_{0.025,\,10}\text{se}(\hat{\theta}_1) \\ 212.639 - 2.228(6.959) &\le \theta_1 \le 212.639 + 2.228(6.959) \\ 197.110 &\le \theta_1 \le 228.158 \end{aligned}$$

y

$$\begin{aligned} \hat{\theta}_2 - t_{0.025,\,10}\text{se}(\hat{\theta}_2) &\le \theta_2 \le \hat{\theta}_2 + t_{0.025,\,10}\text{se}(\hat{\theta}_2) \\ 0.064 - 2.228(0.00833) &\le \theta_2 \le 0.064 + 2.228(0.00833) \\ 0.045 &\le \theta_2 \le 0.083 \end{aligned}$$

respectivamente. Al establecer esos intervalos se han usado los resultados de cómputo de la tabla 12.3. Otros intervalos aproximados de confianza y de predicción se podrían establecer insertando las cantidades adecuadas de regresión no lineal en las ecuaciones correspondientes de la regresión lineal.

Validez de la inferencia aproximada

Como las pruebas, procedimientos e intervalos de confianza en la regresión no lineal se basan en la teoría de muestra grande, y comúnmente el tamaño de muestra en una regresión lineal puede no ser tan grande, es lógico preguntar acerca de la validez de los procedimientos. Sería preferible tener una guía o "regla fácil" que dijera cuándo el tamaño de la muestra es suficientemente grande como para que sean válidos los resultados asintóticos, desafortunadamente no se dispone de esa guía general, sin embargo, hay algunos **indicadores** de que los resultados pueden ser válidos en determinada aplicación.

1. Si el algoritmo de estimación con regresión no lineal converge sólo después de pocas iteraciones, indica que la aproximación lineal que se usó para resolver el problema fue muy satisfactoria, y que es probable que se apliquen muy bien los resultados asintóticos. Cuando la convergencia requiere muchas iteraciones, es un síntoma de que pueden no aplicarse los resultados asintóticos, y se deberían considerar otras verificaciones de la adecuación.
2. Se han desarrollado varias medidas de curvatura y no linealidad del modelo, que describen Bates y Watts [1988]. Esas medidas explican, en forma cuantitativa, la adecuación de la aproximación lineal; de nuevo, una aproximación lineal inadecuada indicaría que son dudosos los resultados asintóticos de la inferencia.

http://carlos2524.jimdo.com/

3. En el capítulo 14 se ilustrará una técnica de remuestreo, llamada **bootstrap**, con la que se puede estudiar la distribución de muestreo de los estimadores, calcular errores estándar aproximados y determinar intervalos de confianza aproximados. Los estimados *bootstrap* de esas cantidades se calcularían y se compararían con los errores estándar e intervalos de confianza aproximados, obtenidos con los resultados asintóticos. Una buena concordancia con los estimados *bootstrap* es un indicio de que son válidos los resultados de la inferencia con muestra grande.

Cuando hay indicios de que no son válidos los resultados de inferencia asintótica, quien busca los modelos tiene algunas alternativas. Una posibilidad es considerar una forma alterna del modelo, si existe, o quizá un modelo distinto de regresión no lineal, a veces pueden ayudar en este respecto las gráficas de los datos y las de diversas funciones de valor esperado del modelo no lineal, también, se pueden usar los resultados de inferencia a partir del remuestreo o con *bootstrap*; sin embargo, si el modelo es incorrecto o está mal especificado, hay pocos motivos para creer que los resultados de remuestreo sean más válidos que los de la inferencia de muestra grande.

12.6 EJEMPLOS DE MODELOS DE REGRESIÓN NO LINEALES

En el caso ideal, se escoge un modelo de regresión no lineal con base en **argumentos teóricos** de la materia que se maneja, que nos dice que los conocimientos químicos, físicos o biológicos específicos conducen a un **modelo mecanístico** de la función de valor esperado, y no a un modelo empírico. Muchos modelos de regresión no lineales quedan dentro de categorías diseñadas para casos o ambientes específicos. En esta sección describiremos algunos de esos modelos.

Quizá la categoría más conocida de modelos no lineales es la de los **modelos de crecimiento**, que son los modelos que se usan para describir cómo crece algo, al cambiar la variable regresora, ya que ésta, con frecuencia, es el tiempo. Entre las aplicaciones características están las de biología, cuando las plantas y los organismos crecen al paso del tiempo, pero también hay muchas aplicaciones en economía e ingeniería, por ejemplo, el crecimiento de confiabilidad de un sistema complejo, al paso del tiempo, se puede describir frecuentemente con un modelo de regresión no lineal.

El modelo de crecimiento **logístico** es

$$y = \frac{\theta_1}{1 + \theta_2 \exp(-\theta_3 x)} + \varepsilon \tag{12.28}$$

Los parámetros en este modelo tienen una interpretación física sencilla. Para $x = 0$, $y = \theta_1/(1 + \theta_2)$ es el valor de y en el momento (o valor) cero. El parámetro θ_1 es el límite del crecimiento cuando $x \to \infty$. Los valores de θ_2 y θ_3 deben ser positivos, también, el término $-\theta_3 x$ en la exponencial del denominador de la ecuación (12.28) se puede reemplazar por una estructura más general, con varios regresores.

El modelo de **Gompertz** se expresa como sigue:

$$y = \theta_1 \exp(-\theta_2 e^{-\theta_3 x}) + \varepsilon \tag{12.29}$$

y es otro modelo de crecimiento que se usa mucho. Cuando $x = 0$, $y = \theta_1 e^{-\theta_3}$, y θ_1 es el límite del crecimiento cuando $x \to \infty$.

http://carlos2524.jimdo.com/

El modelo de crecimiento de **Weibull** es

$$y = \theta_1 - \theta_2 \exp(-\theta_3 x^{\theta_4}) + \varepsilon \tag{12.30}$$

Cuando $x = 0$, $y = \theta_1 - \theta_2$, mientras que el crecimiento límite es θ_1 cuando $x \to \infty$.

En algunas aplicaciones, la respuesta esperada se determina con la solución de un conjunto de ecuaciones diferenciales. Esos modelos se llaman comúnmente **modelos de compartimiento**, y como las reacciones químicas se describen con frecuencia mediante sistemas lineales de ecuaciones diferenciales de primer orden, se aplican a menudo en química, ingeniería química y farmacocinética. En otros casos se especifica la función del valor esperado como la solución de una ecuación diferencial no lineal, o de una ecuación integral que no tiene solución analítica. Hay técnicas especiales para modelar y resolver esos problemas; se remite al lector interesado a Bates y Watts [1988].

PROBLEMAS

12.1 Para el modelo de Michaelis-Menten presentado en la ecuación (12.17), graficar la función del valor esperado, para $\theta_1 = 200$ y $\theta_2 = 0.04$, 0.06, 0.08 y 0.10. Encimar esas curvas en el mismo conjunto de ejes x–y. ¿Qué efectos tiene el parámetro θ_2 sobre el comportamiento de la función de valor esperado?

12.2 Para el modelo de Michaelis-Menten presentado en la ecuación (12.17), graficar la función de valor esperado para $\theta_1 = 100$, 150, 200 y 250, con $\theta_2 = 0.06$. Encimar esas curvas en el mismo conjunto de ejes x–y. ¿Qué efecto tiene el parámetro θ_1 sobre el comportamiento de la función de valor esperado?

12.3 Graficar la función de valor esperado para el modelo de crecimiento logístico, ecuación (12.28), para $\theta_1 = 10$, $\theta_2 = 2$, con valores de $\theta_3 = 0.25$, 1, 2 y 3, respectivamente. Encimar esas gráficas en el mismo conjunto de ejes x–y. ¿Qué efecto tiene el parámetro θ_3 sobre la función valor esperado?

12.4 Trazar la función de valor esperado para el modelo de crecimiento logístico, ecuación (12.28), para $\theta_1 = 1$, $\theta_3 = 1$ y valores de $\theta_2 = 1$, 4 y 8, respectivamente. Encimar esas gráficas en los mismos ejes x–y. Describir el efecto de θ_2 sobre la forma de la función.

12.5 Examinar el modelo de Gompertz, ecuación (12.29). Graficar la función del valor esperado para $\theta_1 = 1$, $\theta_3 = 1$ y $\theta_2 = 1/8$, 1, 8 y 64, dentro del intervalo $0 \le x \le 10$.

a. Describir el comportamiento del modelo en función de θ_2.
b. Describir el comportamiento del modelo cuando $x \to \infty$.
c. ¿Cuál es $E(y)$ cuando $x = 0$?

12.6 Para cada uno de los modelos que se presentan abajo, determinar si es lineal, intrínsecamente lineal o no lineal. Si es intrínsecamente lineal, indicar cómo se puede linealizar mediante una transformación adecuada.

a. $y = \theta_1 e^{\theta_2 + \theta_3 x} + \varepsilon$
b. $y = \theta_1 + \theta_2 x_1 + \theta_2 x_2^{\theta_3} + \varepsilon$
c. $y = \theta_1 + \theta_2/\theta_1 x + \varepsilon$
d. $y = \theta_1 (x_1)^{\theta_2} (x_2)^{\theta_3} + \varepsilon$
e. $y = \theta_1 + \theta_2 e^{\theta_3 x} + \varepsilon$

http://carlos2524.jimdo.com/

12.7 Reconsiderar los modelos de regresión del problema 12.6 partes a-e. Suponer que en esos modelos, los términos de error son multiplicativos y no aditivos. Volver a hacer el problema con esta nueva hipótesis sobre la estructura de los errores.

12.8 Se tienen las siguientes observaciones:

x	y	
0.5	0.68	1.58
1	0.45	2.66
2	2.50	2.04
4	6.19	7.85
8	56.1	54.2
9	89.8	90.2
10	147.7	146.3

a. Ajustar el modelo no lineal de regresión

$$y = \theta_1 e^{\theta_2 x} + \varepsilon$$

para estos datos. Describir cómo se obtuvieron los valores de arranque.

b. Probar el significado de la regresión.

c. Estimar σ^2, la varianza del error.

d. Probar las hipótesis H_0: $\theta_1 = 0$ y H_0: $\theta_2 = 0$. ¿Son distintos de cero los dos parámetros del modelo? Si no lo son, ajustar un modelo adecuado.

e. Analizar los residuales de este modelo. Comentar la adecuación del modelo.

12.9 Con los mismos datos del problema anterior, la medición de la respuesta en las dos columnas se determinó en dos días distintos. Ajustar el nuevo modelo:

$$y = \theta_3 x_2 + \theta_1 e^{\theta_2 x_1} + \varepsilon$$

a esos datos, siendo x_1 el regresor original del problema 12.8, y x_2 una variable indicadora; $x_2 = 0$ si la observación se hizo en el día 1, y $x_2 = 1$, si se hizo en el día 2. ¿Hay algún indicio de una diferencia entre los dos días? Usar $\theta_{30} = 0$ como valor de arranque.

12.10 Se tiene el modelo

$$y = \theta_1 - \theta_2 e^{-\theta_3 x} + \varepsilon$$

Es la llamada ecuación de Mitcherlich, que se usa con frecuencia en ingeniería química. Por ejemplo, y puede ser el rendimiento y x puede ser el tiempo de reacción.

a. ¿Es éste un modelo no lineal de regresión?

b. Describa cómo se obtendrían valores de arranque razonables, de los parámetros θ_1, θ_2 y θ_3.

c. Graficar la función de valor esperado, para los valores de parámetro $\theta_1 = 0.5$, $\theta_2 = -0.10$ y $\theta_3 = 0.10$. Describir la forma de la función.

d. Graficar la función de valor esperado para los valores de parámetro $\theta_1 = 0.5$, $\theta_2 = 0.10$ y $\theta_3 = 0.10$. Comparar la forma con la obtenida en la parte c.

12.11 Los siguientes datos representan la fracción de cloro activo en un producto químico, en función del tiempo después de su fabricación.

http://carlos2524.jimdo.com/

Cloro disponible y_i	Tiempo x_i
0.49, 0.49	8
0.48, 0.47, 0.48, 0.77	10
0.46, 0.46, 0.45, 0.43	12
0.45, 0.43, 0.43	14
0.44, 0.43, 0.43	16
0.46, 0.45	18
0.42, 0.42, 0.43	20
0.41, 0.41, 0.40	22
0.42, 0.40, 0.40	24
0.41, 0.40, 0.41	26
0.41, 0.40	28
0.40, 0.40, 0.38	30
0.41, 0.40	32
0.40,	34
0.41, 0.38	36
0.40, 0.40	38
0.39,	40
0.39	42

a. Trazar un diagrama de dispersión de los datos.
b. Ajustar la ley de Mitcherlich (véase el Prob. 12.10) a estos datos. Describir cómo se obtuvieron los valores iniciales.
c. Probar el significado de la regresión.
d. Determinar intervalos aproximados de 95% de confianza para los parámetros θ_1, θ_2 y θ_3. ¿Hay indicios que respalden la afirmación que los tres parámetros son distintos de cero?
e. Analizar los residuales y comentar la adecuación del modelo.

12.12 Examinar los siguientes datos:

		x_2 = 50	75	100
x_1	2	4.70	5.52	3.98
		2.68	3.75	4.22
	4	6.35	5.88	6.28
		6.10	7.69	7.12
	6	7.85	9.00	11.43
		9.25	9.78	9.62

Estos datos se obtuvieron en un experimento en el que x_1 = tiempo de reacción, minutos, y x_2 = temperatura, grados Celsius. La variable de respuesta y es la concentración (gramos por litro). Se está considerando el modelo

$$y = \theta_1(x_1)^{\theta_2}(x_2)^{\theta_3} + \varepsilon$$

http://carlos2524.jimdo.com/

a. Nótese que se puede linealizar la función del valor esperado sacando logaritmos. Ajustar a los datos el modelo lineal de regresión que resulta.
b. Probar el significado de la regresión. ¿Parece que las dos variables, x_1 y x_2, tienen efectos importantes?
c. Analizar los residuales y comentar la adecuación del modelo.

12.13 Continuación del problema 12.12
a. Ajustar el modelo no lineal del problema 12.12, con la solución obtenida al linealizar la función del valor esperado como valores iniciales.
b. Probar el significado de la regresión. ¿Parece que los efectos de las dos variables, x_1 y x_2, son importantes?
c. Analizar los residuales y comentar la adecuación del modelo.
d. ¿Cuál modelo prefiere el lector, el no lineal o el lineal del problema 12.12?

12.14 Continuación del problema 12.12. Las dos observaciones de cada celda de la tabla de datos, en el problema 12.12, son dos réplicas del experimento. Cada réplica se corrió partiendo su lote correspondiente de una materia prima (lote 1 para réplicas 1 y lote 2 para réplicas 2). Ajustar el modelo

$$y = \theta_4 x_3 + \theta_1 (x_1)^{\theta_2} (x_2)^{\theta_3} + \varepsilon$$

donde $x_3 = 0$, si la observación proviene de la réplica 1, y $x_3 = 1$ si la observación pertenece a la réplica 2. ¿Hay indicios de una diferencia entre los dos lotes de materia prima?

http://carlos2524.jimdo.com/

CAPÍTULO 13

MODELOS LINEALES GENERALIZADOS

13.1 INTRODUCCIÓN

En el capítulo 5 se desarrolló e ilustró la transformación de datos, como método de ajustar modelos de regresión cuando no son adecuadas las hipótesis de una variable de respuesta normalmente distribuida, con varianza constante. Con frecuencia, la transformación de la variable de respuesta es una forma muy eficaz de manejar la no normalidad de la respuesta y también la desigualdad de la varianza. Los mínimos cuadrados ponderados también son una forma potencialmente útil para manejar el problema de varianza no constante. En este capítulo se presenta un método alternativo para transformar datos, cuando no se satisfacen las hipótesis "acostumbradas" de normalidad y de varianza constante. Este método se basa en el **modelo lineal generalizado** (GLM, de *generalized linear model*).

El GLM es una unificación de los modelos de regresión lineal y no lineal, que también permite incorporar distribuciones de respuesta no normales. En un modelo lineal generalizado la distribución de la variable de respuesta sólo necesita ser un miembro de la **familia exponencial**, que comprende las distribuciones normal, de Poisson, binomial, exponencial y gamma, entre sus miembros, además, el modelo lineal con error normal no es más que un caso especial del modelo lineal generalizado, por lo que en muchos aspectos se puede considerar que el lineal generalizado es un método unificador de muchos aspectos del modelado y análisis empíricos de datos.

Se iniciará la presentación de esos modelos examinando el caso de la **regresión logística**, que es una situación en la que la variable de respuesta sólo tiene dos resultados posibles, llamados, en forma genérica, "éxito" y "fracaso" y se representan por 0 y 1. Obsérvese que la respuesta es cualitativa, en esencia, porque los calificativos "éxito" y "fracaso" son totalmente arbitrarios. A continuación se examinará el caso en el que la variable de respuesta es un conteo, como por ejemplo la cantidad de defectos en una unidad de producto, o el número relativo de eventos raros (ejemplo de esto es la cantidad de huracanes en el Atlántico que llegan a Estados Unidos en un año). Por último, se describirá cómo se unifican todos estos casos mediante el modelo lineal generalizado.

13.2 MODELOS DE REGRESIÓN LOGÍSTICA

13.2.1 Modelos con una variable de respuesta binaria

Se considerará el caso en el que la variable de respuesta, en un problema de regresión, sólo asume dos valores posibles: 0 y 1; esos números podrían ser asignaciones arbitrarias a una respuesta cualitativa. Por ejemplo, la respuesta podría ser el resultado de una prueba de fun-

http://carlos2524.jimdo.com/

cionamiento eléctrico para un dispositivo semiconductor, que da como resultado un "éxito", que indica que el dispositivo trabaja bien, o un "fracaso" que podría deberse a un corto, un circuito abierto u otro problema de funcionamiento.

Supóngase que el modelo tiene la forma

$$y_i = \mathbf{x}_i'\beta + \varepsilon_i \tag{13.1}$$

en donde $\mathbf{x}'_i = [1, x_{i1}, x_{i2}, \ldots, x_{ik}]$, $\beta' = [\beta_0, \beta_1, \beta_2, \ldots, \beta_k]$, y la variable de respuesta y_i toma los valores 0 o 1. Se supondrá que la variable de respuesta y_i es una **variable aleatoria de Bernoulli**, cuya distribución de probabilidad es la siguiente:

y_i	Probabilidad
1	$P(y_i = 1) = \pi_i$
0	$P(y_i = 0) = 1 - \pi_i$

Ahora bien, como $E(\varepsilon_i) = 0$, el valor esperado de la variable de respuesta es

$$\begin{aligned} E(y_i) &= 1(\pi_i) + 0(1 - \pi_i) \\ &= \pi_i \end{aligned}$$

Esto implica que

$$E(y_i) = \mathbf{x}_i'\beta = \pi_i$$

Que quiere decir que la respuesta esperada, determinada con la función de respuesta $E(y_i) = \mathbf{x}_i'\beta$ no es más que la probabilidad de que la variable de respuesta tenga el valor de 1.

Hay algunos problemas sustantivos con el modelo de regresión en la ecuación (13.1). El primero es que se observa que si la respuesta es binaria, entonces los términos de error ε_i sólo pueden tener dos valores, que son

$$\begin{aligned} \varepsilon_i &= 1 - \mathbf{x}_i'\boldsymbol{\beta} \quad &\text{cuando } y_i = 1 \\ \varepsilon_i &= -\mathbf{x}_i'\boldsymbol{\beta} \quad &\text{cuando } y_i = 0 \end{aligned}$$

En consecuencia, no es posible que los errores en este modelo sean normales. En segundo lugar, la varianza del error no es constante, ya que

$$\begin{aligned} \sigma_{y_i}^2 &= E\{y_i - E(y_i)\}^2 \\ &= (1 - \pi_i)^2\pi_i + (0 - \pi_i)^2(1 - \pi_i) \\ &= \pi_i(1 - \pi_i) \end{aligned}$$

Obsérvese que esta última expresión equivale a

$$\sigma_{y_i}^2 = E(y_i)[1 - E(y_i)]$$

http://carlos2524.jimdo.com/

porque $E(y_i) = \mathbf{x}_i'\beta = \pi_i$, lo que indica que la varianza de las observaciones (que es igual a la varianza de los errores, porque $\varepsilon_i = y_i - \pi_i$, y π_i es constante) es una función de la media. Por último, hay una restricción para la función de respuesta, ya que

$$0 \leq E(y_i) = \pi_i \leq 1$$

Esta restricción puede causar graves problemas en la elección de una **función de respuesta lineal**, como se ha supuesto al principio, en la ecuación (13.1). Sería posible ajustar un modelo con los datos para los cuales los valores predichos de la respuesta salen del intervalo 0, 1.

En general, cuando la variable de respuesta es binaria, hay bastantes pruebas empíricas que indican que la forma de la función de respuesta debe ser no lineal. Una función monótonamente creciente (o decreciente), en forma de S (o de S invertida), como la de la figura 13.1, es la que se acostumbra emplear; esta función se llama **función de respuesta logística** y tiene la forma

$$E(y) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})} \tag{13.2}$$

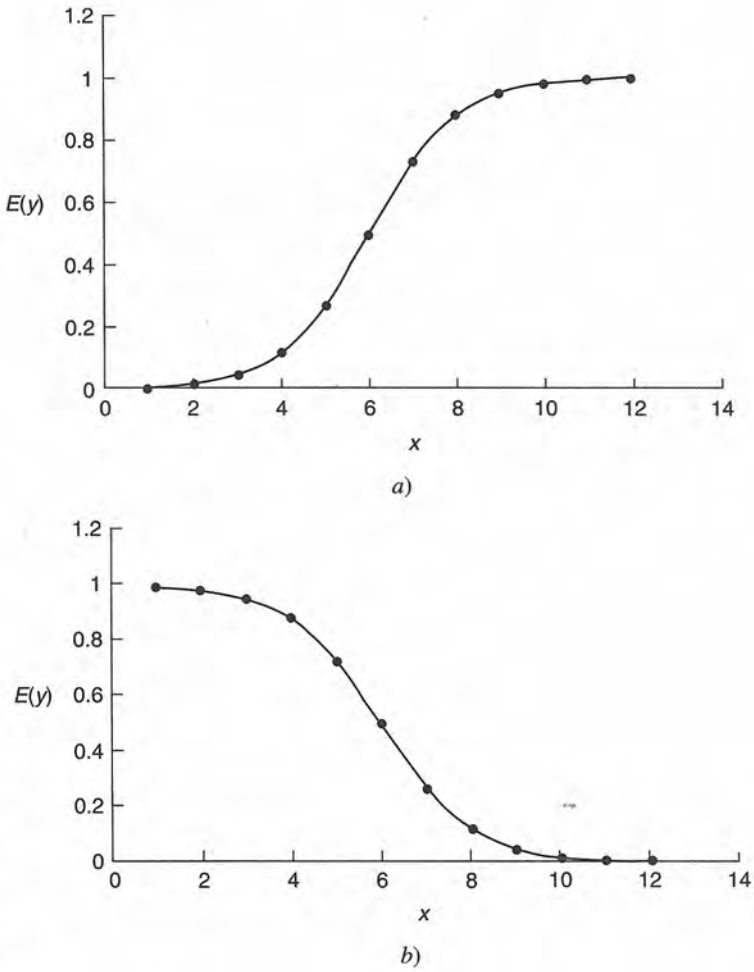
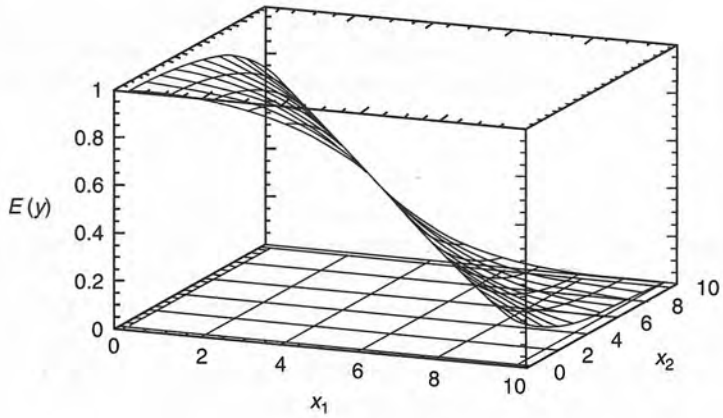
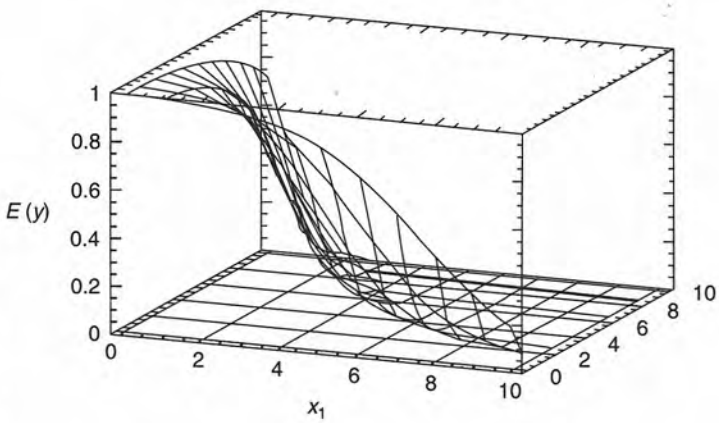


Figura 13.1 Ejemplos de la función de respuesta logística. a) $E(y) = 1/(1 + e^{-6.0-1.0x})$. b) $E(y) = 1/(1 + e^{-6.0+1.0x})$.

http://carlos2524.jimdo.com/



c)



d)

Figura 13.1 (*Continuación*) c) $E(y) = 1/(1 + e^{-5.0+0.65x_1+0.4x_2})$ y d) $E(y) = 1/(1 + e^{-5.0+0.65x_1+0.4x_2+0.15x_1x_2})$.

o bien, lo que es igual

$$E(y) = \frac{1}{1 + \exp(-\mathbf{x}'\boldsymbol{\beta})} \tag{13.3}$$

La función de respuesta logística se puede linealizar con facilidad. Un enfoque consiste en definir la porción estructural del modelo en términos de una función de la media de la función de respuesta. Sea

$$\eta = \mathbf{x}'\beta \tag{13.4}$$

el **predictor lineal**, estando definida η por la transformación

$$\eta = \ln \frac{\pi}{1 - \pi} \tag{13.5}$$

http://carlos2524.jimdo.com/

A esta transformación se le llama con frecuencia **transformación logit** de la probabilidad π, y la relación $\pi/(1-\pi)$ en la transformación se llama **ventaja**; a veces, a la transformación logit se le llama ventaja logarítmica.

Hay otras funciones que tienen la misma forma que la función logística, y también se pueden obtener transformando π, una de ellas es la transformación **probit**, obtenida transformando a π con la distribución normal acumulada. De esta manera se obtiene un **modelo de regresión probit**, este modelo es menos flexible que el de regresión logística, y es probable que no se use tanto, porque no puede incorporar con facilidad más de una variable predictora. Otra transformación posible es la **transformación log-log complementaria** de π, definida por $\ln[-\ln(1-\pi)]$, que produce una función de respuesta que no es simétrica respecto al valor $\pi = 0.5$.

13.2.2 Estimación de parámetros en un modelo de regresión logística

La forma general del modelo de regresión logística es

$$y_i = E(y_i) + \varepsilon_i \tag{13.6}$$

donde las observaciones y_i son variables aleatorias independientes de Bernoulli, cuyos valores esperados son

$$\begin{aligned} E(y_i) &= \pi_i \\ &= \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{x}_i'\boldsymbol{\beta})} \end{aligned} \tag{13.7}$$

Se usará el método de **máxima verosimilitud** para estimar los parámetros del predictor lineal $\mathbf{x}_i'\boldsymbol{\beta}$.

Cada observación de la muestra sigue la distribución de Bernoulli, por lo que la distribución de probabilidades de cada observación es

$$f_i(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}, \qquad i = 1, 2, \ldots, n$$

y naturalmente, cada observación y_i toma el valor 0 o 1. Como las observaciones son independientes, la función verosimilitud no es más que

$$\begin{aligned} L(y_1, y_2, \ldots, y_n, \boldsymbol{\beta}) &= \prod_{i=1}^{n} f_i(y_i) \\ &= \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \end{aligned}$$

Es más cómodo trabajar con el logaritmo de la verosimilitud:

$$\begin{aligned} \ln L(y_1, y_2, \ldots, y_n, \boldsymbol{\beta}) &= \ln \prod_{i=1}^{n} f_i(y_i) \\ &= \sum_{i=1}^{n}\left[y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right)\right] + \sum_{i=1}^{n} \ln(1-\pi_i) \end{aligned}$$

http://carlos2524.jimdo.com/

Ahora bien, como $1 - \pi_i = [1 + \exp(\mathbf{x}_i'\beta)]^{-1}$, y $\eta_i = \ln[\pi_i/(1 - \pi_i)] = \mathbf{x}_i'\beta$, el logaritmo de la verosimilitud se puede expresar como sigue:

$$\ln L(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \mathbf{x}_i' \boldsymbol{\beta} - \sum_{i=1}^{n} \ln \left[1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})\right] \tag{13.8}$$

Con frecuencia, en los modelos de regresión logística se tienen observaciones o intentos repetidos en cada nivel de las variables x, esto sucede mucho en los experimentos diseñados. Sea y_i la cantidad de 1 observado en i, y n_i la cantidad de intentos en cada observación, entonces, el logaritmo de la verosimilitud se transforma en

$$\ln L(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \pi_i + \sum_{i=1}^{n} n_i \ln(1 - \pi_i) - \sum_{i=1}^{n} y_i \ln(1 - \pi_i) \tag{13.9}$$

Se podrían usar métodos numéricos de búsqueda, para calcular los estimados $\hat{\boldsymbol{\beta}}$ por máxima verosimilitud (MLE, por *maximum likelihood estimates*); sin embargo, sucede que se pueden usar los mínimos cuadrados iterativamente reponderados (IRLS) para determinar los MLE. Para conocer los detalles de este procedimiento, véase el apéndice C.13. Hay varios excelentes programas de cómputo para implementar los IRLS para el modelo de regresión logística, como PROC GENMOD de SAS y *S*-PLUS.

Sea $\hat{\boldsymbol{\beta}}$ el estimado final de los parámetros del modelo que se obtiene con el algoritmo anterior. Si son correctas las hipótesis del modelo, se puede demostrar que, en forma asintótica,

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \quad \text{y} \quad \operatorname{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \tag{13.10}$$

El valor estimado del predictor lineal es $\hat{\eta}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}}$, y el valor esperado del modelo de regresión logística se escribe con frecuencia como sigue:

$$\begin{aligned} \hat{y}_i = \hat{\pi}_i &= \frac{\exp(\hat{\eta}_i)}{1 + \exp(\hat{\eta}_i)} \\ &= \frac{\exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{x}_i'\boldsymbol{\beta})} \\ &= \frac{1}{1 + \exp(-\mathbf{x}_i'\hat{\boldsymbol{\beta}})} \end{aligned} \tag{13.11}$$

Ejemplo 13.1 *Datos sobre la neumoconiosis*

Un artículo de la revista *Biométrica* en 1959 presentó datos acerca de los mineros de carbón que presentan síntomas de neumoconiosis grave, y de la cantidad de años de exposición, esos datos aparecen en la tabla 13.1. La variable de respuesta de interés, y, es la proporción de mineros que tienen síntomas graves. En la figura 13.2 se ve una gráfica de la variable de respuesta en función de la cantidad de años de exposición. Un modelo razonable de probabilidad de la cantidad de casos graves es el binomial, por lo que se ajustará un modelo de regresión logística a esos datos.

http://carlos2524.jimdo.com/

TABLA 13.1 Datos de neumoconiosis

Años de exposición	Cantidad de casos graves	Cantidad total de mineros	Proporción de casos graves, y
5.8	0	98	0
15.0	1	54	0.0185
21.5	3	43	0.0698
27.5	8	48	0.1667
33.5	9	51	0.1765
39.5	8	38	0.2105
46.0	10	28	0.3571
51.5	5	11	0.4545

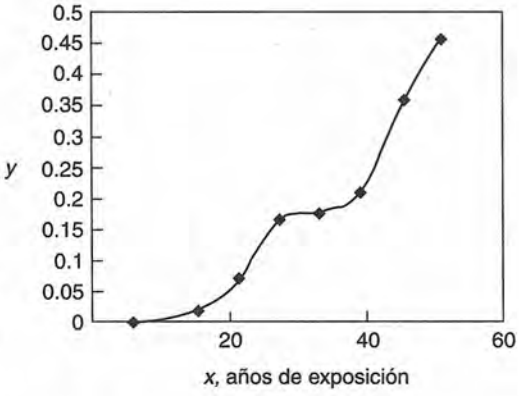


Figura 13.2 Gráfica de los datos de neumoconiosis de la tabla 13.1.

La tabla 13.2 contiene una porción seleccionada de los resultados de SAS PROC GENMOD, paquete de programas muy usado, con excelente capacidad para ajustar y analizar modelos de regresión logística. En las secciones que siguen se describirá con más detalle la información que contienen estos resultados. La parte de ellos titulada "Análisis de estimados de parámetro" presenta los estimados de los coeficientes de regresión en el predictor lineal.

El modelo ajustado de regresión logística es

$$\hat{y} = \frac{1}{1 + e^{-4.7965 + 0.0935x}}$$

donde x es la cantidad de años de exposición. La figura 13.3 presenta una gráfica de los valores ajustados de este modelo, sobrepuesta al diagrama de dispersión de los datos muestrales. Parece que el modelo de regresión logística proporciona un ajuste razonable a esos datos.

13.2.3 Interpretación de los parámetros en un modelo de regresión logística

Es relativamente fácil interpretar los parámetros de un modelo de regresión logística. Primero se examinará el caso en el que el predictor lineal sólo tiene un regresor, por lo que el valor ajustado del modelo en determinado valor de x, por ejemplo x_i, es

http://carlos2524.jimdo.com/

TABLA 13.2 Resultados de SAS PROC GENMOD del modelo de regresión logística en el ejemplo 13.1

The GENMOD Procedure

Model Information

Description	Value
Data set	WORK.LUNG
Distribution	BINOMIAL
Link Function	LOGIT
Dependent Variable	R
Dependent Variable	N
Observation Used	8
Number of Events	44
Number of Trails	371

Criteria for Assessing Goodness of Fit

Criterion	DF	Value	Value DF
Deviance	6	6.0508	1.0085
Scaled Deviance	6	6.0508	1.0085
Pearson Chi-Square	6	5.0285	0.8381
Scaled Pearson X2	6	5.0285	0.8381
Log Likelihood		– 109.6637	

Analysis of Parameter Estimates

Parameter	DF	Estimate	Std. Error	Chi Square	Pr > Chi
INTERCEPT	1	– 4.7965	0.5686	71.1606	0.0001
X	1	0.0935	0.0154	36.7084	0.0001
SCALE	0	1.0000	0.0000		

Note: The scale parameter was held fixed.

Last Evaluation of the Negative of the Hessian

Parameter	PRM1	PRM2	Scale
PRM1	32.87874	1153.472	0
PRM2	1153.472	44669.1	0
Scale	0 0	0	

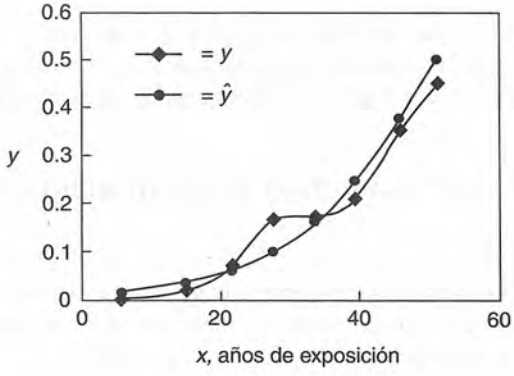


Figura 13.3 Modelo ajustado de regresión logística para los datos de neumoconiosis de la tabla 13.1.

http://carlos2524.jimdo.com/

$$\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

El valor ajustado en $x_i + 1$ es

$$\hat{\eta}(x_i + 1) = \hat{\beta}_0 + \hat{\beta}_1(x_i + 1)$$

y la diferencia entre los dos valores predichos es

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_1$$

Ahora, $\hat{\eta}(x_i)$ sólo es el logaritmo de la ventaja cuando la variable regresora es igual a x_i, y $\hat{\eta}(x_i + 1)$ es el logaritmo de la ventaja cuando el regresor es igual a $x_i + 1$. Por consiguiente, la diferencia entre los dos valores ajustados es

$$\begin{aligned}\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) &= \ln\left(\text{ventaja}_{x_i+1}\right) - \ln\left(\text{ventaja}_{x_i}\right)\\ &= \ln\left(\frac{\text{ventaja}_{x_i+1}}{\text{ventaja}_{x_i}}\right)\\ &= \hat{\beta}_1\end{aligned}$$

Si se sacan antilogaritmos se obtiene el **cociente de ventaja**

$$\hat{O}_R = \frac{\text{ventaja}_{x_i+1}}{\text{ventaja}_{x_i}} = e^{\hat{\beta}_1} \tag{13.12}$$

Se puede interpretar el cociente de ventaja como el aumento estimado en la probabilidad de éxito asociado con un cambio unitario en el valor de la variable predictora, en general, el aumento estimado del cociente de ventaja, asociado con un cambio de d unidades en la variable predictora, es $\exp(d\,\hat{\beta}_1)$.

La interpretación de los coeficientes de regresión en el modelo logístico múltiple se parece al caso en el que el predictor lineal sólo contiene un regresor, que nos indica que la cantidad $\exp(\hat{\beta}_j)$ es el cociente de ventaja para el regresor x_j, suponiendo que las demás variables predictoras son constantes.

Ejemplo 13.2 *Datos de neumoconiosis*

En el ejemplo 13.1 se ajustó el modelo de regresión logística

$$\hat{y} = \frac{1}{1 + e^{-4.7965 + 0.0935x}}$$

a los datos de neumoconiosis de la tabla 13.1. Como el predictor lineal sólo contiene una variable regresora, y $\hat{\beta}_1 = 0.0935$, se puede calcular el cociente de ventaja con la ecuación (13.12):

$$\hat{O}_R = e^{\hat{\beta}_1} = e^{0.0935} = 1.10$$

http://carlos2524.jimdo.com/

Esto implica que cada año adicional de exposición aumenta 10% la probabilidad de contraer una neumoconiosis grave; si el tiempo de exposición aumenta 10 años, la relación de ventaja se transforma en $\exp(d\,\hat{\beta}_1) = \exp[10(0.0935)] = 2.55$. Eso indica que las probabilidades más se duplican con una exposición de 10 años.

13.2.4 Pruebas de hipótesis para los parámetros del modelo

La prueba de hipótesis en la regresión logística (y en general, para el modelo lineal general) se basa en **pruebas de cociente de máxima verosimilitud**, que es un procedimiento para muestras grandes, por lo que los procedimientos de prueba se basan en la teoría asintótica. El método de la razón de verosimilitud conduce a un estadístico llamado **desviación**.

Desviación del modelo

La desviación del modelo compara el logaritmo de la verosimilitud del modelo ajustado con el logaritmo de la verosimilitud de un modelo saturado, que es un modelo que tiene exactamente n parámetros y se ajusta perfectamente a los datos de la muestra. Para el modelo de regresión logística eso significa que las probabilidades π_i son totalmente irrestrictas, por lo que al igualar $\pi_i = y_i$ (recuérdese que $y_i = 0$ o 1) se maximizaría la verosimilitud. Se puede demostrar que esto da como resultado el valor máximo de la función verosimilitud para el modelo saturado de unidad, por lo que el valor máximo de la función logaritmo de verosimilitud es cero.

Ahora se examinará la función logaritmo de verosimilitud para el modelo logístico ajustado. Cuando los estimados $\hat{\boldsymbol{\beta}}$ de máxima verosimilitud se usan en la función logaritmo de verosimilitud, ésta alcanza su valor máximo, el cual es

$$\ln L(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} y_i \mathbf{x}_i' \hat{\boldsymbol{\beta}}_i - \sum_{i=1}^{n} \ln\left[1 + \exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}})\right] \tag{13.13}$$

El valor de la función logaritmo de verosimilitud, para el modelo ajustado, nunca podrá ser mayor que el de esa función para el modelo saturado, porque el modelo ajustado contiene menos parámetros. La desviación compara al logaritmo de verosimilitud del modelo saturado con el logaritmo de verosimilitud del modelo ajustado. En forma específica, la **desviación del modelo** se define como sigue:

$$\begin{aligned} \lambda(\boldsymbol{\beta}) &= 2\ln L(\text{modelo saturado}) - 2\ln L(\hat{\boldsymbol{\beta}}) \\ &= 2\left[\ell(\text{modelo saturado}) - \ell(\hat{\boldsymbol{\beta}})\right] \end{aligned} \tag{13.14}$$

donde ℓ representa el logaritmo de la función verosimilitud. Ahora bien, si el modelo de regresión logística es la función correcta de regresión, y el tamaño n de la muestra es grande, la desviación del modelo tiene aproximadamente una distribución ji cuadrada, con $n - p$ grados de libertad. Valores grandes de la desviación del modelo indican que el modelo no es correcto, mientras que un valor pequeño implica que el modelo ajustado (que tiene menos parámetros que el modelo saturado) se ajusta a los datos casi tan bien como el modelo saturado. Los criterios formales de prueba son los siguientes:

http://carlos2524.jimdo.com/

$$\begin{aligned} &\text{si } \lambda(\beta) \leq \chi^2_{\alpha,n-p} \quad \text{se concluye que el modelo ajustado es adecuado} \\ &\text{si } \lambda(\beta) > \chi^2_{\alpha,n-p} \quad \text{se concluye que el modelo ajustado no es adecuado} \end{aligned} \tag{13.15}$$

La desviación está relacionada con una cantidad muy conocida. Si se considera el error normal estándar del modelo de regresión lineal, sucede que la desviación es el error de la suma de cuadrados de residuales dividido entre la varianza del error σ^2.

Ejemplo 13.3 *Datos de la neumoconiosis*

La tabla 13.2 contiene el resultado obtenido con SAS PROC GENMOD para el modelo de regresión logística que se ajustó a los datos de neumoconiosis. En la sección de resultados titulada "Criterios para evaluar bondad de ajuste" la desviación del modelo aparece como $\lambda(\beta) = 6.0508$. Esto se podría comparar con un punto porcentual ji cuadrada con $n - p = 8 - 2 = 6$ grados de libertad, o bien se podría calcular el valor P. Si se usa el método del valor P se ve que $P = 0.4175$, por lo que según los criterios de la ecuación (13.15), se llega a la conclusión de que el modelo ajustado es adecuado.

Una buena regla fácil y general es dividir la desviación del modelo, $\lambda(\beta)$, entre sus grados de libertad, $n - p$. Si el cociente $\lambda(\beta)/(n - p)$ es cercano a la unidad, se considera que el modelo es adecuado. Un valor grande de esta relación puede indicar que el modelo se especificó en forma incorrecta.

Prueba de hipótesis sobre subconjuntos de parámetros usando la desviación

También se puede usar la desviación para probar hipótesis sobre subconjuntos de los parámetros del modelo, tal como se usaron diferencia de (o error de) las sumas de cuadrados, para probar hipótesis en el caso del modelo de regresión lineal con errores normales. Recuérdese que el modelo se puede escribir en la forma

$$\begin{aligned} \eta &= \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 \end{aligned} \tag{13.16}$$

donde el **modelo completo** tiene p parámetros, β_1 contiene a $p - r$ de esos parámetros, β_2 contiene a r de esos parámetros, y las columnas de las matrices $\mathbf{X}_1$ y $\mathbf{X}_2$ contienen las variables asociadas con esos parámetros. Supóngase que se desea probar las hipótesis

$$\begin{aligned} H_0&: \boldsymbol{\beta}_2 = \mathbf{0} \\ H_1&: \boldsymbol{\beta}_2 \neq \mathbf{0} \end{aligned} \tag{13.17}$$

Por consiguiente, el **modelo reducido** es

$$\eta = \mathbf{X}_1\beta_1 \tag{13.18}$$

Ahora se ajusta el modelo reducido y se define a $\lambda(\beta_1)$ como la desviación para el modelo reducido, esta desviación siempre será mayor que la del modelo completo, porque el mode-

http://carlos2524.jimdo.com/

lo reducido contiene menos parámetros; sin embargo, si la desviación del modelo reducido no es mucho más grande que la del modelo completo, quiere decir que el modelo reducido tiene un ajuste más o menos tan bueno como el modelo completo, por lo que es probable que los parámetros en β_2 sean igual a cero. Esto es, no se puede rechazar la hipótesis nula anterior, sin embargo, si la diferencia en la desviación es grande, al menos uno de los parámetros en β_2 probablemente no sea cero, y se debe rechazar la hipótesis nula. Formalmente, la diferencia en la desviación es

$$\lambda(\beta_2|\beta_1) = \lambda(\beta_1) - \lambda(\beta) \tag{13.19}$$

y esta cantidad tiene $n - (p - r) - (n - p) = r$ grados de libertad. Si es cierta la hipótesis nula y si n es grande, la diferencia de la desviación en la ecuación 13.19 tiene una distribución ji cuadrada con r grados de libertad. Por consiguiente, el estadístico de prueba y los criterios de decisión son

$$\begin{aligned} &\text{si } \lambda(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) \geq \chi^2_{\alpha,r} && \text{rechazar la hipótesis nula} \\ &\text{si } \lambda(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) < \chi^2_{\alpha,r} && \text{no rechazar la hipótesis nula} \end{aligned} \tag{13.20}$$

A veces, la diferencia de desviación $\lambda(\beta_2|\beta_1)$ se llama **desviación parcial**, que es una prueba de cociente de verosimilitud. Para visualizarlo, sea $L(\hat{\beta})$ el valor máximo de la función verosimilitud para el modelo completo, y sea $L(\hat{\beta}_1)$ el valor máximo de la función de verosimilitud para el modelo reducido. El **cociente de verosimilitud** es

$$\frac{L(\hat{\boldsymbol{\beta}}_1)}{L(\hat{\boldsymbol{\beta}})} \tag{13.21}$$

El estadístico para la prueba de cociente de verosimilitud es igual a –2 multiplicado por el logaritmo del cociente de verosimilitud, es decir

$$\chi^2 = -2 \ln \frac{L(\hat{\boldsymbol{\beta}}_1)}{L(\hat{\boldsymbol{\beta}})} \tag{13.22}$$

Sin embargo, es exactamente igual que la diferencia de desviación. Para visualizarlo, se sustituye la definición de la desviación, ecuación (13.14), en la ecuación (13.22), observando que se anulan los logaritmos de verosimilitud para el modelo saturado.

Ejemplo 13.4 *Datos de neumoconiosis*

De nuevo se considerarán los datos de neumoconiosis de la tabla 13.1. El modelo que se ajustó inicialmente a los datos es

$$\hat{y} = \frac{1}{1 + e^{-4.7965 + 0.0935x}}$$

Supóngase que se quiere determinar si al agregar un término cuadrático en el predictor lineal mejora el modelo. En consecuencia, se definirá al modelo completo como

$$y = \frac{1}{1 + e^{\beta_0 + \beta_1 x + \beta_{11} x^2}} + \varepsilon$$

http://carlos2524.jimdo.com/

TABLA 13.3 Resultados con SAS PROC GENMOD para el ejemplo 13.4

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.LUNG
Distribution	BINOMIAL
Link Funtion	LOGIT
Dependent Variable	R
Dependent Variable	N
Observation Used	8
Number of Events	44
Number of Trails	371

Criteria for Assessing Goodness of Fit

Criterion	DF	Value	Value/DF
Deviance	5	3.2816	0.6563
Scaled Deviance	5	3.2816	0.6563
Pearson Chi-Square	5	2.9448	0.5890
Scaled Pearson X2	5	2.9448	0.5890
Lug Likelihood		– 108.2791	

Analysis of Parameter Estimates

Parameter	DF	Estimate	Std. Error	Chi Square	Pr > Chi
INTERCEPT	1	– 6.7108	1.5352	19.1075	0.0001
X	1	0.2276	0.0928	6.0213	0.0141
X*X	1	– 0.0021	0.0014	2.3323	0.1267
SCALE	0	1.0000	0.0000		

NOTE: The scale parameter was held fixed.

LR Statistics for Type 1 Analysis

Source	Deviance	DF	Chi Square	Pr > Chi
INTERCEPT	56.9028	0		
X	6.0508	1	50.8520	0.0001
X*X	3.2816	1	2.7691	0.0961

LR Statistics for Type 3 Analysis

Souce	DF	Chi Square	Pr > Chi
X	1	8.9918	0.0027
X*X	1	2.7691	0.0961

Last Evaluation of the Negative of the Hessian

Parameter	PRM1	PRM2	PRM3	Scale
PRM1	33.43534	1199.595	46195.0	0
PRM2	1199.595	46195.0	1868432	0
PRM3	46195.0	1868432	7847716	0
Scale	0	0	0	0

http://carlos2524.jimdo.com/

La tabla 13.3 contiene los resultados abreviados de SAS PROC GENMOD para este modelo. Ahora bien, el predictor lineal para el modelo completo se puede escribir como sigue:

$$\begin{aligned}\eta &= \mathbf{X}\boldsymbol{\beta}\\ &= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2\\ &= \beta_0 + \beta_1 x + \beta_{11}x^2\end{aligned}$$

En esta tabla se ve que la desviación del modelo completo es

$$\lambda(\beta) = 3.2816$$

con $n - p = 8 - 3 = 5$ grados de libertad. Ahora, el modelo reducido tiene $\mathbf{X}_1\beta_1 = \beta_0 + \beta_1 x$, por lo que $\mathbf{X}_2\beta_2 = \beta_{11}x^2$, con $r = 1$ grado de libertad. El modelo reducido se ajustó originalmente en el ejemplo 13.1, y la tabla 13.2 muestra que la desviación del modelo reducido es

$$\lambda(\beta_1) = 6.0508$$

con $p - r = 3 - 1 = 2$ grados de libertad. Por lo anterior, la diferencia en desviaciones del modelo completo y el reducido se calcula con la ecuación (13.19), y resulta

$$\begin{aligned}\lambda(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1) &= \lambda(\boldsymbol{\beta}_1) - \lambda(\boldsymbol{\beta})\\ &= 6.0508 - 3.2816\\ &= 2.7692\end{aligned}$$

que se compararía con una distribución ji cuadrada con $r = 1$ grado de libertad. Como el valor P asociado con la diferencia de desviaciones es 0.0961, se podría concluir que hay cierto valor marginal al incluir el término cuadrático en la variable regresora x = años de exposición, en el predictor lineal, para el modelo de regresión logística.

El programa SAS PROC GENMOD calcula en forma automática la desviación parcial, con la ecuación (13.19), para cada variable regresora individual, en el predictor lineal, como si se agregara por último al modelo. A esto se le llama análisis tipo 3 en SAS, y aparece en los resultados de la tabla 13.3, bajo el encabezado "Estadísticos LR para análisis tipo 3". Los resultados también contienen estadísticos de la desviación parcial para el análisis tipo 1, que en la terminología de SAS significa que las variables regresoras se agregan en orden al predictor lineal.

Pruebas de los coeficientes individuales del modelo

Se pueden hacer pruebas de los coeficientes individuales del modelo, como

$$\begin{aligned}H_0&: \beta_j = 0\\ H_1&: \beta_j \neq 0\end{aligned} \tag{13.23}$$

aplicando el método de la diferencia de la desviación, como se ilustró en el ejemplo 13.4. Hay otro método que también se basa en la teoría de los estimadores de máxima verosimilitud. Para muestras grandes, la distribución de un estimador de máxima verosimilitud es

http://carlos2524.jimdo.com/

aproximadamente normal, con poco o ningún sesgo, además, las varianzas y covarianzas de un conjunto de estimadores de máxima verosimilitud se pueden determinar a partir de las segundas derivadas parciales de la función logaritmo de verosimilitud, con respecto a los parámetros del modelo, evaluadas en los estimados de máxima verosimilitud, entonces se puede hacer un estadístico t para probar las hipótesis de arriba. A esto a veces se le llama **inferencia de Wald**.

Sea **G** la matriz de $p \times p$ de las segundas derivadas parciales de la función logaritmo de verosimilitud, esto es,

$$G_{ij} = \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j}, \qquad i, j = 0, 1, \ldots, k$$

G se llama **matriz hessiana** o **de Hess**. Si los elementos de la hessiana se evalúan en los estimadores de máxima verosimilitud $\beta = \hat{\beta}$, la matriz de covarianza para muestra grande, de los coeficientes de regresión, es

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \hat{\Sigma} = -\mathbf{G}(\hat{\boldsymbol{\beta}})^{-1} \tag{13.24}$$

Las raíces cuadradas de los elementos diagonales de esta matriz son los errores estándar de muestras grandes de los coeficientes de regresión, por lo que el estadístico de prueba para la hipótesis nula en

$$H_0: \beta_j = 0$$
$$H_1: \beta_j \neq 0$$

es

$$Z_0 = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} \tag{13.25}$$

La distribución de referencia para este estadístico es la distribución normal estándar. Algunos programas de cómputo elevan al cuadrado el estadístico Z_0 y lo comparan con una distribución ji cuadrada con un grado de libertad.

Ejemplo 13.5 *Datos de neumoconiosis*

La tabla 13.3 contiene los resultados obtenidos con SAS PROC GENMOD y los datos de neumoconiosis, que se dieron originalmente en la tabla 13.1. El modelo ajustado es

$$\hat{y} = \frac{1}{1 + e^{-6.7108 + 0.2276x - 0.0021x^2}}$$

Los resultados de SAS muestran la matriz hessiana, junto con los errores estándar de cada coeficiente del modelo, y un estadístico de prueba equivalente al Z_0 de la ecuación (13.25). SAS presenta el cuadrado de esta ecuación, y calcula el valor P de una distribución ji cuadrada. Obsérvese que el valor P para β_1 es $P = 0.0141$, que implica que los años de exposición son un regresor importante, sin embargo, obsérvese que el valor P para $\beta_{11} = 0.1267$, lo cual sugiere que el término de años de exposición al cuadrado no contribuye en forma significativa al ajuste.

http://carlos2524.jimdo.com/

Recuérdese que en el ejemplo anterior, al probar el significado de β_{11} con el método de la desviación parcial, se obtuvo un valor P diferente. Ahora bien, en la regresión lineal, la prueba t con un solo regresor equivale a la prueba F parcial con una sola variable (recuérdese que el cuadrado del estadístico t es igual al estadístico parcial F). Sin embargo, esta equivalencia sólo es cierta en los **modelos lineales**, y el modelo lineal generalizado es un **modelo no lineal**.

También es directo el uso de la inferencia de Wald para establecer límites de confianza de los coeficientes individuales de regresión. Un intervalo aproximado de confianza de 100(1 – a) por ciento para el j-ésimo coeficiente del modelo es

$$\hat{\beta}_j - Z_{\alpha/2}\text{se}\left(\hat{\beta}_j\right) \leq \beta_j \leq \hat{\beta}_j + Z_{\alpha/2}\text{se}\left(\hat{\beta}_j\right) \tag{13.26}$$

Ejemplo 13.6 *Datos sobre neumoconiosis*

Al usar el resultado de SAS, en la tabla 13.3, se puede determinar un intervalo de confianza aproximado de 95% para β_{11}, de acuerdo con la ecuación (13.26), como sigue:

$$\begin{aligned} \hat{\beta}_{11} - Z_{0.025}\text{se}\left(\hat{\beta}_{11}\right) \leq \beta_{11} &\leq \hat{\beta}_{11} + Z_{0.025}\text{se}\left(\hat{\beta}_{11}\right) \\ -0.0021 - 1.96(0.0014) \leq \beta_{11} &\leq -0.0021 + 1.96(0.0014) \\ -0.0048 \leq \beta_{11} &\leq 0.0006 \end{aligned}$$

Obsérvese que el intervalo de confianza incluye al cero, por lo que al nivel de 5% de significancia no se rechazaría la hipótesis de que este coeficiente del modelo es cero.

13.3 REGRESIÓN DE POISSON

Ahora se describirá otro escenario de modelado de regresión, en el que la variable de respuesta de interés no tiene distribución normal. En este caso, la variable de respuesta representa un conteo de algún elemento relativamente raro, como por ejemplo, defectos en una unidad de producto manufacturado, errores en un programa de cómputo, o conteo de partículas de materia, u otros contaminantes, en el ambiente. El analista desea modelar la relación entre los conteos observados y variables regresoras o predictoras potencialmente útiles. Por ejemplo, a un ingeniero le podría interesar modelar la relación entre la cantidad observada de defectos en una unidad de producto, y las condiciones de producción con las que se produjo realmente la unidad.

Se supondrá que y_i, la variable de respuesta, es un conteo, como por ejemplo la observación $y_i = 0, 1, \ldots$. Un modelo probabilístico razonable para los datos de conteo es, con frecuencia, la distribución de Poisson:

$$f(y) = \frac{e^{-\mu}\mu^y}{y!}, \qquad y = 0, 1, \ldots \tag{13.27}$$

donde el parámetro $\mu > 0$. La distribución de Poisson es otro ejemplo de distribuciones de probabilidades en donde se relacionan el promedio y la varianza. De hecho, para la distribución de Poisson se demuestra en forma directa que

http://carlos2524.jimdo.com/

$$E(y) = \mu \qquad \text{y que} \qquad \text{Var}(y) = \mu$$

Esto es, tanto la media **como** la varianza de la distribución de Poisson son iguales al parámetro μ.

El modelo de regresión de Poisson se puede escribir en la forma

$$y_i = E(y_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{13.28}$$

Se supondrá que el valor esperado de la respuesta observada se puede escribir como sigue:

$$E(y_i) = \mu_i$$

y que hay una función g que relaciona la media de la respuesta con un predictor lineal, digamos

$$\begin{aligned} g(\mu_i) &= \eta_i \\ &= \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \\ &= \mathbf{x}_i' \boldsymbol{\beta} \end{aligned} \tag{13.29}$$

A la función g se le suele llamar **función cadena**. La relación entre el promedio y el predictor lineal es

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i'\beta) \tag{13.30}$$

Hay varias funciones cadena que se usan con frecuencia con la distribución de Poisson. Una de ellas es la **cadena identidad**

$$g(\mu_i) = \mu_i = \mathbf{x}_i'\beta \tag{13.31}$$

Cuando se usa esta cadena, $E(y_i) = \mu_i = \mathbf{x}_i'\beta$, porque $\mu_i = g^{-1}(\mathbf{x}_i'\beta) = \mathbf{x}_i'\beta$. Otra función cadena muy usada para la distribución de Poisson es la **cadena logarítmica**

$$g(\mu_i) = \ln(\mu_i) = \mathbf{x}_i'\beta \tag{13.32}$$

Para la cadena logarítmica en esta ecuación, la relación entre la media de la variable de respuesta y el predictor lineal es

$$\begin{aligned} \mu_i &= g^{-1}(\mathbf{x}_i' \boldsymbol{\beta}) \\ &= e^{\mathbf{x}_i' \boldsymbol{\beta}} \end{aligned} \tag{13.33}$$

La cadena logarítmica tiene especial atractivo para la distribución de Poisson porque asegura que todos los valores predichos de la variable de respuesta serán no negativos.

El método de máxima verosimilitud se usa para estimar los parámetros de la regresión de Poisson. El desarrollo se parece mucho al método que se usó en la regresión logística. Si se tiene una muestra aleatoria de n observaciones de la respuesta y y los predictores x, la función de verosimilitud es

http://carlos2524.jimdo.com/

$$
\begin{aligned}
L(\mathbf{y}, \beta) &= \prod_{i=1}^{n} f_i(y_i) \\
&= \prod_{i=1}^{n} \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \qquad (13.34) \\
&= \frac{\prod_{i=1}^{n} \mu_i^{y_i} \exp\left(-\sum_{i=1}^{n} \mu_i\right)}{\prod_{i=1}^{n} y_i!}
\end{aligned}
$$

en donde $\mu_i = g^{-1}(\mathbf{x}_i'\beta)$. Una vez especificada la función cadena, se maximiza el logaritmo de la verosimilitud

$$
\ln L(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \ln(\mu_i) - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} \ln(y_i!) \qquad (13.35)
$$

Para determinar los estimados de máxima verosimilitud para los parámetros, en la regresión de Poisson, se pueden usar los mínimos cuadrados iterativamente reponderados, de acuerdo con un método parecido al que se usó para la regresión logística; una vez obtenidos los estimados de parámetro $\hat{\beta}$, el modelo ajustado de regresión de Poisson es

$$
\hat{y}_i = g^{-1}\left(\mathbf{x}_i' \hat{\boldsymbol{\beta}}\right) \qquad (13.36)
$$

Por ejemplo, si se usa la cadena identidad, la ecuación de predicción se transforma en

$$
\begin{aligned}
\hat{y}_i &= g^{-1}\left(\mathbf{x}_i' \hat{\boldsymbol{\beta}}\right) \\
&= \mathbf{x}_i' \hat{\boldsymbol{\beta}}
\end{aligned}
$$

y si se usa la cadena logarítmica, entonces

$$
\begin{aligned}
\hat{y}_i &= g^{-1}\left(\mathbf{x}_i' \hat{\boldsymbol{\beta}}\right) \\
&= \exp\left(\mathbf{x}_i' \hat{\boldsymbol{\beta}}\right)
\end{aligned}
$$

La inferencia acerca del modelo y sus parámetros sigue exactamente el mismo método que se usó para la regresión logística, esto es, la desviación del modelo es una medida general de bondad de ajuste, y se pueden hacer pruebas sobre subconjuntos de los parámetros del modelo usando la diferencia de la desviación entre los modelos completo y reducido. Éstas son pruebas de cociente de verosimilitud. Para probar hipótesis y establecer intervalos de confianza para los parámetros individuales del modelo se puede usar la inferencia de Wald, basada en las propiedades de estimadores de máxima verosimilitud con muestra grande.

http://carlos2524.jimdo.com/

Ejemplo 13.7 *Datos de daños a aviones*

Durante la guerra de Vietnam, la Marina de Estados Unidos usó varios tipos de aviones de ataque (**bombarderos** en la jerga de la Marina), con frecuencia en misiones de golpes a baja altura contra puentes, carreteras y otras instalaciones de transporte; dos de los aviones eran el Skyhawk A-4 de McDonnell Douglas y el Intruder A-6 de Grumman. El A-4 es un avión monomotor, monoplaza y ligero, usado principalmente de día, también lo volaron los Ángeles Azules, equipo de demostración de vuelos de la Marina, durante muchos años. El A-6 es un bimotor, biplaza, para todo tiempo y de tamaño mediano, con excelentes posibilidades en día y noche. Sin embargo, el Intruder no podía maniobrarse desde los portaaviones de clase Essex, más pequeños, muchos de los cuales todavía estaban en servicio durante el conflicto.

Se adjudicaron considerables recursos contra el A-4 y el A-6, incluyendo armas pequeñas, artillería AAA (antiaérea) y cohetes de superficie a aire. La tabla 13.4 contiene datos tomados en 30 misiones de ataque donde intervinieron estos dos tipos de avión. El regresor x_1 es una variable indicadora (A-4 = 0 y A-6 = 1), y los demás regresores x_2 y x_3 son la carga de bombas (en toneladas) y el total de meses de experiencia de la tripulación. La variable de respuesta fue la cantidad de lugares donde se infligieron daños al avión.

Se modelará la respuesta daños en función de los tres regresores; como la respuesta es un conteo, se usará un modelo de regresión de Poisson, con cadena logarítmica. La tabla 13.5 presenta parte del reporte entregado por SAS PROC GENMOD. El modelo en la primera página de la tabla usa los tres regresores. La prueba de adecuación del modelo, basada en la desviación, es satisfactoria, pero se observa que x_3 = experiencia de la tripulación, no es significativa, de acuerdo con la prueba de Wald y con la desviación parcial tipo 3. Es una indicación razonable de que se puede sacar a x_3 del modelo, sin embargo, cuando se quita x_3, sucede que ahora ya no es significativa x_1 = tipo de avión (se puede comprobar con facilidad que la desviación parcial tipo 3 para x_1 en este modelo tiene un valor $P = 0.1582$). Una reflexión sobre los datos de la tabla 13.4 revelará que hay mucha multicolinealidad en ellos. En esencia, el A-6 es un avión más grande, por lo que transporta mayor carga de bombas, y como tiene una tripulación de dos hombres, tenderá a tener más meses totales de experiencia de tripulación. Por consiguiente, al aumentar x_1 hay una tendencia de que los otros dos regresores también aumenten.

Para investigar la utilidad potencial de diversos modelos de subconjunto se ajustarán los tres modelos de dos variables, y los tres modelos de una variable, a los datos de la tabla 13.4. Un resumen de los resultados obtenidos así es el siguiente:

Modelo	Desviación	Diferencia de la desviación en comparación con el modelo completo	Valor P
$x_1x_2x_3$	28.4906		
x_1x_2	31.0223	2.5316	0.1116
x_1x_3	32.8817	4.3911	0.0361
x_2x_3	31.6062	3.1155	0.0775
x_1	38.3497	9.8591	0.0072
x_2	38.0137	4.5251	0.1041
x_3	54.9653	26.4747	< 0.0001

http://carlos2524.jimdo.com/

TABLA 13.4 Datos de daños a aviones

Observación	y	x_1	x_2	x_3
1	0	0	4	91.5
2	1	0	4	84.0
3	0	0	4	76.5
4	0	0	5	69.0
5	0	0	5	61.5
6	0	0	5	80.0
7	1	0	6	72.5
8	0	0	6	65.0
9	0	0	6	57.5
10	2	0	7	50.0
11	1	0	7	103.0
12	1	0	7	95.5
13	1	0	8	88.0
14	1	0	8	80.5
15	2	0	8	73.0
16	3	1	7	116.1
17	1	1	7	100.6
18	1	1	7	85.0
19	1	1	10	69.4
20	2	1	10	53.9
21	0	1	10	112.3
22	1	1	12	96.7
23	1	1	12	81.1
24	2	1	12	65.6
25	5	1	8	50.0
26	1	1	8	120.0
27	1	1	8	104.4
28	5	1	14	88.9
29	5	1	14	73.7
30	7	1	14	57.8

Al examinar la diferencia de la desviación entre cada uno de los modelos de subconjunto y el modelo completo, se observa que al sacar x_1 o x_2 se produce un modelo con dos variables que es bastante peor que el modelo completo. Al eliminar x_3 se obtiene un modelo que no tiene diferencia significativa con el modelo completo, pero como ya se observó, x_1 no es significativa en este modelo. Esto conduce a considerar los modelos de una variable. Sólo uno de ellos, el que contiene a x_2, no tiene diferencia significativa con el modelo completo. Los resultados de SAS PROC GENMOD para este modelo se ven en la segunda parte de la tabla 13.5.

El modelo de regresión de Poisson para predecir los daños es

$$\hat{y} = e^{-1.6491 + 0.2282x_2}$$

La desviación de este modelo es $\lambda(\beta) = 33.0137$, con 28 grados de libertad, y el valor P es 0.2352, por lo que se llega a la conclusión que el modelo es un ajuste adecuado para los datos.

http://carlos2524.jimdo.com/

TABLA 13.5 Resultados de SAS PROC GENMOD para los datos de daños a aviones en el ejemplo 13.5.

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.PLANE
Distribution	POISSON
Link Function	LOG
Dependent Value	Y
Observations Used	30

Criteria for Assessing Goodness of Fit

Criterion	DF	Value	Value/DF
Deviance	26	28.4906	1.0958
Scaled Deviance	26	28.4906	1.0958
Pearson Chi-Square	26	25.4279	0.9780
Scaled Pearson X2	26	25.4279	0.9780
Log Likelihood		– 11.3455	

Analysis of Parameter Estimates

Parameter	DF	Estimate	Std. Error	Chi Square	Pr > Chi
INTERCEPT	1	– 0.3824	0.8630	0.1964	0.6577
X1	1	0.8805	0.5010	3.0892	0.0788
X2	1	0.1352	0.0653	4.2842	0.0385
X3	1	– 0.0127	0.0080	2.5283	0.1118
SCALE	0	1.0000	0.0000		

NOTE: The scale parameters was held fixed.

LR Statistics for Type 1 Analysis

Source	Deviance	DF	Chi Square	Pr > Chi
INTERCEPT	57.5983	0		
X1	38.3497	1	19.2486	0.0001
X2	31.0223	1	7.3274	0.0068
X3	28.4906	1	2.5316	0.1116

LR Statistics for Type 3 Analysis

Source	DF	Chi Square	Pr > Chi
X1	1	3.1155	0.0775
X2	1	4.3911	0.0361
X3	1	2.5316	0.1116

http://carlos2524.jimdo.com/

TABLA 13.5 (Continuación)

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.PLANE
Distribution	POISSON
Link Function	LOG
Dependent Variable	Y
Observations Used	30

Criteria for Assessing Goodness of Fit

Criterion	DF	Value	Value/DF
Deviance	28	33.0137	1.1791
Scaled Deviance	28	33.0137	1.1791
Pearson Chi-Square	28	33.4108	1.1932
Scaled Pearson X2	28	33.4108	1.1932
Log Likelihood		– 13.6071	

Analysis of Parameter Estimates

Parameter	DF	Estimate	Std. Err	Chi Square	Pr > Chi
INTERCEPT	1	– 1.6491	0.4996	10.8980	0.0010
X2	1	0.2282	0.0462	24.3904	0.0001
SCALE	0	1.0000	0.0000		

NOTE: The scale parameter was held fixed.

LR Statistics for Type 1 Analysis

Source	Deviance	DF	Chi cuadrada	Pr > Chi
INTERCEPT	57.5983	0		
X2	33.0137	1	24.5846	0.0001

LR Statistics for Type 3 Analysis

Source	DF	Chi Square	Pr > Chi
X2	1	24.5846	0.0001

13.4 EL MODELO LINEAL GENERALIZADO

Todos los modelos que se han examinado en las dos secciones anteriores de este capítulo pertenecen a una familia de modelos de regresión, llamada **modelo lineal generalizado** (GLM, de *generalized linear model*). Este modelo es en realidad una técnica unificadora de modelos de regresión y de diseño de experimentos, que une los modelos acostumbrados de regresión con la teoría normal y los modelos no lineales, como el logístico y el de Poisson.

Una hipótesis clave en el GLM es que la distribución de la variable de respuesta es un miembro de la **familia exponencial** de distribuciones, que incluye, entre otras, la normal, binomial, de Poisson, normal inversa, exponencial y gamma. Las distribuciones que son miembros de la familia exponencial tienen la forma general

http://carlos2524.jimdo.com/

$$f(y_i, \theta_i, \phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + h(y_i, \phi)\} \tag{13.37}$$

donde ϕ es un parámetro de escala, y θ_i se llama parámetro natural de localización. Para los miembros de la familia exponencial,

$$\begin{aligned} \mu = E(y) &= \frac{db(\theta_i)}{d\theta_i} \\ \text{Var}(y) &= \frac{d^2b(\theta_i)}{d\theta_i^2}a(\phi) \\ &= \frac{d\mu}{d\theta_i}a(\phi) \end{aligned} \tag{13.38}$$

Sea

$$\text{Var}(\mu) = \frac{\text{Var}(y)}{a(\phi)} = \frac{d\mu}{d\theta_i}, \tag{13.39}$$

donde Var(μ) representa la dependencia de la varianza de la respuesta a su media. Es una característica de todas las distribuciones de la familia exponencial, a excepción de la distribución normal. Como resultado de la ecuación (13.39), se tiene

$$\frac{d\theta_i}{d\mu} = \frac{1}{\text{Var}(\mu)} \tag{13.40}$$

En el apéndice C.13 se demostrará que las distribuciones normal, binomial y de Poisson son miembros de la familia exponencial.

13.4.1 Funciones cadena y predictores lineales

El concepto básico de un GLM es desarrollar un modelo lineal para una función adecuada del valor esperado de la variable de respuesta. Sea η_i el **predictor lineal**, definido por

$$\eta_i = g[E(y_i)] = g(\mu_i) = \mathbf{x}_i'\beta \tag{13.41}$$

Nótese que la respuesta esperada no es más que

$$E(y_i) = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i'\beta) \tag{13.42}$$

La función g se llama **función cadena**. Recuérdese que se presentó el concepto de una función cadena al describir la regresión de Poisson en la sección 13. Hay muchas alternativas posibles para la función cadena, pero si se escoge

$$\eta_i = \theta_i \tag{13.43}$$

se dice que η_i es una **cadena canónica**. La tabla 13.6 muestra las cadenas canónicas para las alternativas más comunes de distribuciones que se emplean con el modelo lineal general (GLM).

http://carlos2524.jimdo.com/

TABLA 13.6 Cadenas canónicas para el modelo lineal generalizado

Distribución	Cadena canónica
Normal	$\eta_i = \mu_i$ (cadena identidad)
Binomial	$\eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$ (cadena logística)
Poisson	$\eta_i = \ln(\lambda)$ (cadena logarítmica)
Exponencial	$\eta_i = \frac{1}{\lambda_i}$ (cadena recíproca)
Gamma	$\eta_i = \frac{1}{\lambda_i}$ (cadena recíproca

Hay otras funciones cadena que se pueden usar con un GLM, incluyendo:

1. La cadena probit,

$$\eta_i = \Phi^{-1}[E(y_i)]$$

en la que Φ representa la función distribución normal estándar acumulada.

2. La cadena log-log complementaria,

$$\eta_i = \ln\{\ln[1 - E(y_i)]\}$$

3. La cadena de la familia de potencias,

$$\eta_i = \begin{cases} E(y_i)^{\lambda}, & \lambda \neq 0 \\ \ln\left[E(y_i)\right], & \lambda = 0 \end{cases}$$

Una idea muy básica es que hay dos componentes en un sistema lineal general: la distribución de la respuesta y la función cadena. Se puede considerar la selección de la función cadena en una forma muy parecida a la de una transformación en la respuesta, sin embargo, a diferencia de una transformación, la función cadena aprovecha la distribución **natural** de la respuesta. Así como el no usar una transformación adecuada puede acarrear problemas con un modelo lineal ajustado, la elección inadecuada de la función cadena también puede ocasionar grandes problemas con un modelo lineal general.

13.4.2 Estimación de parámetros e inferencia en el modelo lineal general

El método de máxima verosimilitud es la base teórica de la estimación de parámetros en el modelo lineal general, sin embargo, la implementación real de la máxima verosimilitud da como resultado un algoritmo basado en los mínimos cuadrados iterativamente reponderados (IRLS). Es exactamente lo que se vio antes para los casos especiales de la regresión logística y la de Poisson. Los detalles del procedimiento se presentan en el apéndice C.13.

http://carlos2524.jimdo.com/

En este capítulo nos basaremos en SAS PROC GENMOD para ajustar modelos y para hacer inferencia.

Si $\hat{\beta}$ es el valor final de los coeficientes de regresión que produce el algoritmo IRLS, y si son correctas las hipótesis del modelo, incluyendo la elección de la función cadena, se puede demostrar que, asintóticamente

$$E(\hat{\beta}) = \beta \qquad \text{y que} \qquad \mathrm{Var}(\hat{\beta}) = a(\phi)(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \tag{13.44}$$

siendo la matriz $\mathbf{V}$ una matriz diagonal formada con las varianzas de los parámetros estimados en el predictor lineal, aparte de $a(\phi)$.

Observaciones importantes acerca del modelo lineal general:

1. En el caso normal, cuando los experimentadores y los analistas de datos usan una transformación, usan mínimos cuadrados ordinarios para ajustar realmente el modelo en la escala transformada.
2. En un modelo lineal general se reconoce que la varianza de la respuesta no es constante y se usan mínimos cuadrados ponderados como base de la estimación de parámetros.
3. Lo anterior sugiere que un modelo lineal general debe ser mejor que el análisis estándar, usando transformaciones cuando un problema queda con varianza constante después de hacer la transformación.
4. Toda la inferencia que se describió antes acerca de la regresión logística se aplica en forma directa al modelo lineal general. Esto es, se puede usar la desviación del modelo para probar el ajuste general del modelo, y la diferencia de las desviaciones entre un modelo completo y uno reducido se puede usar para probar hipótesis acerca de subconjuntos de parámetros en el modelo. Se puede aplicar la inferencia de Wald para probar hipótesis y establecer intervalos de confianza para parámetros individuales del modelo.

Ejemplo 13.8 *El experimento de fibras de estambre*

La tabla 13.7 contiene datos de un experimento para investigar tres factores, x_1 = longitud, x_2 = amplitud y x_3 = carga, sobre los ciclos a la falla y de las fibras de estambre. Las variables regresoras se codificaron, y los lectores familiarizados con el diseño de experimentos reconocerán que se usó un diseño factorial 3^3. Estos datos también aparecen en Box y Draper (1987) y en Myers y Montgomery (1995). Estos autores usan los datos para ilustrar la utilidad de las transformaciones estabilizadoras de varianza. Los cuatro autores demuestran que la transformación logarítmica es muy eficiente para estabilizar la varianza de la respuesta ciclos a la falla. El modelo de mínimos cuadrados es

$$\hat{y} = \exp(6.33 + 0.83x_1 - 0.63x_2 - 0.39x_3)$$

En este experimento, la variable de respuesta es un ejemplo de respuesta no negativa que cabría esperar tenga una distribución asimétrica con una cola derecha larga. Con frecuencia, los datos de falla se modelan con distribuciones exponenciales, de Weibull, lognormal o gamma, porque poseen la forma anticipada y también porque a veces hay justificación teórica o empírica para una determinada distribución.

http://carlos2524.jimdo.com/

TABLA 13.7 Datos del experimento de lana peinada

x_1	x_2	x_3	y
−1	−1	−1	674
0	−1	−1	1414
1	−1	−1	3636
−1	0	−1	338
0	0	−1	1022
1	0	−1	1568
−1	1	−1	170
0	1	−1	442
1	1	−1	1140
−1	−1	0	370
0	−1	0	1198
1	−1	0	3184
−1	0	0	266
0	0	0	620
1	0	0	1070
−1	1	0	118
0	1	0	332
1	1	0	884
−1	−1	1	292
0	−1	1	634
1	−1	1	2000
−1	0	1	210
0	0	1	438
1	0	1	566
−1	1	1	90
0	1	1	220
1	1	1	360

Aquí se modelarán los datos de ciclos a la falla con un modelo lineal general que usa la distribución gamma y la cadena logarítmica. De acuerdo con la tabla 13.6, se observa que en este caso la cadena canónica es la inversa; sin embargo, la cadena logarítmica suele ser una elección muy efectiva con la distribución gamma.

La tabla 13.8 presenta algo de la información resumida de resultados de SAS PROC GENMOD para los datos de la fibra de estambre. Nótese que el modelo ajustado es

$$\hat{y} = \exp(6.35 + 0.84x_1 - 0.63x_2 - 0.39x_3)$$

que es virtualmente idéntico al que se obtuvo por transformación de datos. En realidad, como la transformación logarítmica funciona muy bien en este caso, no debe sorprender que el modelo lineal general produzca un modelo casi idéntico. Recuérdese que se observó que es muy probable que el modelo lineal general sea una alternativa eficaz de una transformación de datos cuando la transformación no puede producir las propiedades deseadas de varianza constante, y normalidad aproximada en la variable de respuesta.

Para el caso de la respuesta gamma es adecuado usar la **desviación escalada** de los resultados de SAS como medida de ajuste general del modelo. Esta cantidad se compararía con la distribución ji cuadrada con $n - p$ grados de libertad, como se acostumbra. En la tabla 13.8 se ve que la desviación escalada es 27.1276, y comparándola con una distribución ji cuadrada con 23 grados de libertad se obtiene un valor P aproximado de 0.25, por lo que no hay

http://carlos2524.jimdo.com/

TABLA 13.8 Resultados SAS PROC GENMOD para el experimento de lana peinada

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.WOOL
Distribution	GAMMA
Link Function	LOG
Dependent Variable	CYCLES
Observations Used	27

Criteria for Assessing Goodness of Fit

Criteria	DF	Value	Value/DF
Deviance	23	0.7694	0.0335
Scaled Deviance	23	27.1276	1.1795
Pearson Chi[Square	23	0.7274	0.0316
Scaled Pearson X2	23	25.6456	1.1150
Log Likelihood		– 161.3784	

Analysis of Parameter Estimates

Parameter	DF	Estimate	Std Err	Chi Square	Pr > Chi
INTERCEPT	1	6.3489	0.0324	38373.0419	0.0001
A	1	0.8425	0.0402	438.3606	0.0001
B	1	– 0.6313	0.0396	253.7576	0.0001
C	1	– 0.3851	0.0402	91.8566	0.0001
SCALE	1	35.2585	9.5511		

NOTE: The scale parameter was estimated by maximum likelihood.

LR Statistics for Type 1 Analysis

Source	Deviance	DF	Chi cuadrada	Pr > Chi
INTERCEPT	22.8861	0		
A	10.2104	1	23.6755	0.0001
B	3.3459	1	31.2171	0.0001
C	0.7694	1	40.1106	0.0001

LR Statistics for Type 3 Analysis

Source	DF	Chi Square	Pr > Chi
A	1	77.2935	0.0001
B	1	63.4324	0.0001
C	1	40.1106	0.0001

indicación de que el modelo sea inadecuado, según el criterio de la desviación, nótese que la desviación escalada, dividida entre sus grados de libertad igualmente se acerca a la unidad. También se muestran, en la tabla 13.8, las pruebas de Wald y los estadísticos de desviación parcial (los análisis tipo 1, o "efectos agregados en orden" y tipo 3, o "efectos agregados al último") para cada regresor en el modelo. Los estadísticos de prueba indican que todos los regresores son predictores importantes, y se deben incluir en el modelo.

http://carlos2524.jimdo.com/

13.4.3 Predicción y estimación con el modelo lineal general

Para cualquier modelo lineal generalizado, el estimado de la respuesta media en un punto de interés, por ejemplo $\mathbf{x}_0$, es

$$\hat{y}_0 = \hat{\mu}_0 = g^{-1}\left(\mathbf{x}_0' \hat{\boldsymbol{\beta}}\right) \tag{13.45}$$

en donde g es la función cadena y se sobreentiende que $\mathbf{x}_0$ se puede desarrollar según la forma del modelo, si es necesario, para acomodar términos como interacciones que se hayan incluido en el predictor lineal. Un intervalo aproximado de confianza para la respuesta media en ese punto se determina como sigue: Sea Σ la matriz varianza-covarianza asintótica para $\hat{\boldsymbol{\beta}}$, por lo que

$$\Sigma = a(\phi)(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

La varianza asintótica del predictor lineal estimado en $\mathbf{x}_0$ es

$$\mathrm{Var}(\hat{\eta}_0) = \mathrm{Var}(\mathbf{x}_0' \hat{\boldsymbol{\beta}}) = \mathbf{x}_0' \Sigma \mathbf{x}_0$$

Así, un estimado de esta varianza es $\mathbf{x}_0' \hat{\Sigma}\, \mathbf{x}_0$, donde $\hat{\Sigma}$ es la matriz estimada de varianza-covarianza de $\hat{\boldsymbol{\beta}}$. El intervalo de confianza de $100(1-\alpha)$ por ciento de la respuesta media real en el punto $\mathbf{x}_0$ es

$$L \leq \mu(\mathbf{x}_0) \leq U \tag{13.46}$$

en donde

$$L = g^{-1}(\mathbf{x}_0' \hat{\boldsymbol{\beta}} - Z_{\alpha/2}\mathbf{x}_0' \hat{\Sigma}\, \mathbf{x}_0) \quad \text{y} \quad U = g^{-1}(\mathbf{x}_0' \hat{\boldsymbol{\beta}} + Z_{\alpha/2}\mathbf{x}_0' \hat{\Sigma}\, \mathbf{x}_0) \tag{13.47}$$

Este método lo usa SAS PROC GENMOD para calcular los intervalos de confianza de la respuesta media. Este método para determinar los intervalos de confianza suele funcionar bien en la práctica, porque $\hat{\boldsymbol{\beta}}$ es un estimado de verosimilitud máxima, y en consecuencia cualquier función de $\hat{\boldsymbol{\beta}}$ también es un estimado de máxima verosimilitud. El procedimiento anterior tan solo establece un intervalo de confianza en el espacio definido por el predictor lineal, para después transformar de regreso ese intervalo a la métrica original.

También es posible usar la inferencia de Wald para deducir otras ecuaciones de intervalos aproximados de confianza para la respuesta media. Véanse los detalles en Myers y Montgomery (1997).

Ejemplo 13.9 *El experimento de la lana peinada*

La tabla 13.9 presenta tres conjuntos de intervalos de confianza para la respuesta media del experimento de fibras de lana descrito originalmente en el ejemplo 13.8. En esta tabla se han mostrado los intervalos de 95% de confianza para la respuesta promedio en todos los 27 puntos de los datos experimentales originales para tres modelos: mínimos cuadrados en escala logarítmica, respuesta no transformada de este modelo de mínimos cuadrados y el modelo lineal generalizado (distribución gamma de la respuesta y cadena logarítmica). Los intervalos de confianza para el modelo lineal generalizado se calcularon con la ecuación

http://carlos2524.jimdo.com/

TABLA 13.9 Comparación de los intervalos de confianza de 95% de la respuesta media, para los datos de la lana peinada

	Por métodos de mínimos cuadrados con transformación logarítmica de datos				Con modelo lineal generalizado		Longitud del intervalo de confianza de 95%	
	Transformados		No transformados					
Observación	Valor predicho	Intervalo de confianza de 95%	Valor predicho	Intervalo de confianza de 95%	Valor predicho	Intervalo de confianza de 95%	Mínimos cuadrados	GLM
1	2.83	(2.76, 2.91)	682.50	(573.85, 811.52)	680.52	(583.83, 793.22)	237.67	209.39
2	2.66	(2.60, 2.73)	460.26	(397.01, 533.46)	463.00	(407.05, 526.64)	136.45	119.50
3	2.49	(2.42, 2.57)	310.38	(260.98, 369.06)	315.01	(271.49, 365.49)	108.09	94.00
4	2.56	(2.50, 2.62)	363.25	(313.33, 421.11)	361.96	(317.75, 412.33)	107.79	94.58
5	2.39	(2.34, 2.44)	244.96	(217.92, 275.30)	246.26	(222.55, 272.51)	57.37	49.96
6	2.22	(2.15, 2.28)	165.20	(142.50, 191.47)	167.55	(147.67, 190.10)	48.97	42.42
7	2.29	(2.21, 2.36)	193.33	(162.55, 229.93)	192.52	(165.69, 223.70)	67.38	58.01
8	2.12	(2.05, 2.18)	130.38	(112.46, 151.15)	130.98	(115.43, 148.64)	38.69	33.22
9	1.94	(1.87, 2.02)	87.92	(73.93, 104.54)	89.12	(76.87, 103.32)	30.62	26.45
10	3.20	(3.13, 3.26)	1569.28	(1353.94, 1819.28)	1580.00	(1390.00, 1797.00)	465.34	407.00
11	3.02	(2.97, 3.08)	1058.28	(941.67, 1189.60)	1075.00	(972.52, 1189.00)	247.92	216.48
12	2.85	(2.79, 2.92)	713.67	(615.60, 827.37)	731.50	(644.35, 830.44)	211.77	186.09
13	2.92	(2.87, 2.97)	835.41	(743.19, 938.86)	840.54	(759.65, 930.04)	195.67	170.39
14	2.75	(2.72, 2.78)	563.25	(523.24, 606.46)	571.87	(536.67, 609.38)	83.22	72.70
15	2.58	(2.63, 2.63)	379.84	(337.99, 426.97)	389.08	(351.64, 430.51)	88.99	78.87
16	2.65	(2.58, 2.71)	444.63	(383.53, 515.35)	447.07	(393.81, 507.54)	131.82	113.74
17	2.48	(2.43, 2.53)	299.85	(266.75, 336.98)	304.17	(275.13, 336.28)	70.23	61.15
18	2.31	(2.24, 2.37)	202.16	(174.42, 234.37)	206.95	(182.03, 235.27)	59.95	53.23
19	3.56	(3.48, 3.63)	3609.11	(3034.59, 4292.40)	3670.00	(3165.00, 4254.00)	1257.81	1089.00
20	3.39	(3.32, 3.45)	2433.88	(2099.42, 2821.63)	2497.00	(2200.00, 2833.00)	722.21	633.00
21	3.22	(3.14, 3.29)	1641.35	(1380.07, 1951.64)	1699.00	(1462.00, 1974.00)	571.57	512.00
22	3.28	(3.22, 3.35)	1920.88	(1656.91, 2226.90)	1952.00	(1720.00, 2215.00)	569.98	495.00
23	3.11	(3.06, 3.16)	1295.39	(1152.66, 1455.79)	1328.00	(1200.00, 1470.00)	303.14	270.00
24	2.94	(2.88, 3.01)	873.57	(753.53, 1012.74)	903.51	(793.15, 1029.00)	259.22	235.85
25	3.01	(2.93, 3.08)	1022.35	(859.81, 1215.91)	1038.00	(894.79, 1205.00)	356.10	310.21
26	2.84	(2.77, 2.90)	689.45	(594.70, 799.28)	706.34	(620.99, 803.43)	204.58	182.44
27	2.67	(2.59, 2.74)	464.94	(390.93, 552.97)	480.57	(412.29, 560.15)	162.04	147.86

http://carlos2524.jimdo.com/

(13.47). Las dos últimas columnas de la tabla 13.9 comparan las longitudes de los intervalos de confianza por mínimos cuadrados con teoría normal, de la respuesta no transformada, con los del modelo lineal generalizado; nótese que las longitudes de los intervalos del modelo lineal generalizado siempre son más cortos, que los del análisis de mínimos cuadrados basado en transformaciones. Así, aun cuando las ecuaciones de predicción obtenidas con estas dos técnicas son muy parecidas (como se hizo notar en el Ej. 13.8), hay cierta evidencia que indica que las predicciones obtenidas con el modelo lineal generalizado son más precisas, porque los intervalos de confianza son más cortos.

13.4.4 Análisis de residuales en el modelo lineal generalizado

Como en cualquier procedimiento de ajuste de modelo, es importante el análisis de residuales para ajustar el modelo lineal generalizado. Los residuales pueden guiar sobre la adecuación general del modelo, para verificar las hipótesis y para dar una indicación sobre la adecuación de la función de cadena seleccionada.

Los **residuales originales** u ordinarios del modelo lineal generalizado son las diferencias entre las observaciones y los valores ajustados,

$$\begin{aligned} e_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\mu}_i \end{aligned} \tag{13.48}$$

Se recomienda en general hacer el análisis de residuales en el modelo lineal generalizado mediante **residuales de desviación**. El i-ésimo residual de desviación se define como la raíz cuadrada de la contribución de la i-ésima observación a la desviación, multiplicada por el signo del residual original, es decir:

$$r_{Di} = \sqrt{d_i}\,\text{signo}(y_i - \hat{y}_i) \tag{13.49}$$

donde d_i es la contribución de la i-ésima observación a la desviación. Para el caso de la regresión logística (un modelo lineal generalizado con errores binomiales y cadena logit), se puede demostrar que

$$d_i = y_i \ln\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i)\left[\frac{1 - (y_i/n_i)}{1 - \hat{\pi}_i}\right], \quad i = 1, 2, \ldots, n$$

donde

$$\hat{\pi}_i = \frac{1}{1 + e^{-\mathbf{x}'_i\hat{\boldsymbol{\beta}}}}$$

Nótese que a medida que mejora el ajuste del modelo con los datos, se ve que $\hat{\pi}_i \equiv y_i/n_i$, y que los residuales de desviación se hacen más pequeños, cercanos a cero. Para la regresión de Poisson con cadena logarítmica,

$$d_i = y_i \ln\left(\frac{y_i}{e^{\mathbf{x}'_i\hat{\boldsymbol{\beta}}}}\right) - \left(y_i - e^{\mathbf{x}'_i\hat{\boldsymbol{\beta}}}\right), \quad i = 1, 2, \ldots, n$$

http://carlos2524.jimdo.com/

De nuevo, nótese que a medida que el valor observado de la respuesta y_i y el valor predicho $\hat{y}_i = e^{\mathbf{x}_i'\hat{\beta}}$ se acercan entre sí, los residuales de la desviación tienden a cero.

En general, los residuales de desviación se comportan en forma muy parecida a los residuales ordinarios en un modelo de regresión lineal con la teoría normal estándar. Así, graficar los residuales de desviación, contra valores ajustados en una escala de probabilidad normal es un diagnóstico lógico. Cuando se grafican los residuales de desviación en función de valores ajustados se acostumbra transformar los primeros a una escala constante de información. Así,

1. Para respuestas normales, usar $\hat{y}_i$.
2. Para respuestas binomiales, usar $2\,\text{sen}^{-1}\sqrt{\hat{\pi}_i}$.
3. Para respuestas de Poisson, usar $2\sqrt{\hat{y}_i}$.
4. Para respuestas gamma, usar $2\ln(\hat{y}_i)$.

Ejemplo 13.10

La tabla 13.10 representa las observaciones reales del experimento de lana peinada en el ejemplo 13.8, junto con los valores predichos con el modelo lineal generalizado (respuesta gamma con cadena logarítmica) ajustado a los datos, los residuales originales y los residuales de desviación. Estas cantidades se calcularon con SAS PROC GENMOD. La figura 13.4a es una gráfica de probabilidad normal de los residuales de desviación, y la figura 13.4b es una gráfica de los residuales de desviación en función de los valores ajustados de "información constante", $2\ln(\hat{y}_i)$. La gráfica de probabilidad normal de los residuales de desviación es en general satisfactoria, mientras que la de los residuales de desviación en función de los valores ajustados indica que una de las observaciones podría ser un valor atípico no muy externo; sin embargo, ninguna de las gráficas proporciona indicación significativa alguna sobre la inadecuación del modelo, por lo que se llega a la conclusión que el modelo lineal generalizado con distribución gamma de la variable de respuesta y una cadena logarítmica es un modelo muy satisfactorio de la respuesta ciclos a la falla.

13.4.5 Sobredispersión

La sobredispersión es un fenómeno que sucede a veces cuando se modelan datos de respuesta con una distribución binomial o de Poisson. Básicamente quiere decir que la varianza de la respuesta es mayor de la que se podría esperar para esa elección de distribución estadística de la respuesta. Con frecuencia la presencia de sobredispersión se determina evaluando el valor de la desviación del modelo dividido entre los grados de libertad. Si el cociente es mucho mayor que la unidad, es posible que haya sobredispersión.

La forma más directa de modelar este caso es permitir que la función de varianza de las distribuciones binomial o de Poisson tenga un factor multiplicativo de dispersión ϕ, de modo que

$$\text{Var}(y) = \phi\mu(1-\mu) \qquad \text{distribución binomial}$$

$$\text{Var}(y) = \phi\mu \qquad \text{distribución de Poisson}$$

http://carlos2524.jimdo.com/

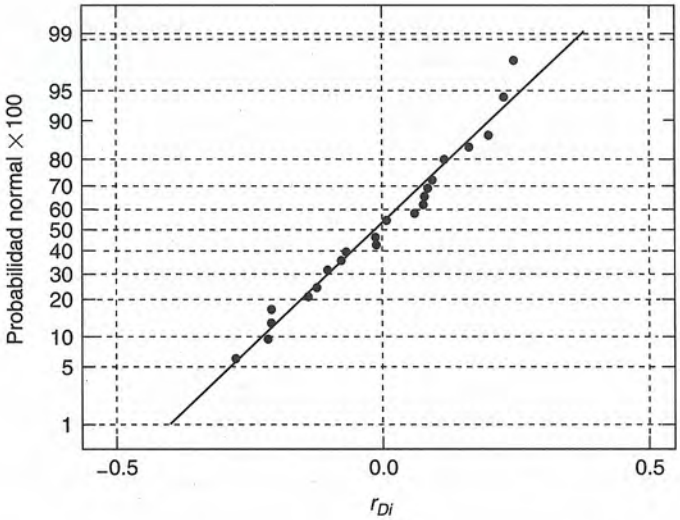
TABLA 13.10 Valores predichos y residuales del experimento de lana peinada.

y_i	$\hat{y}$	Predicción lineal $x'\beta$	e_i	r_{Di}
674	680.5198	6.5229	– 6.5198	– 0.009611
370	462.9981	6.1377	– 92.9981	– 0.2161
292	315.0052	5.7526	– 23.0052	– 0.0749
338	361.9609	5.8915	– 23.9609	– 0.0677
266	246.2636	5.5064	19.7364	0.0781
210	167.5478	5.1213	42.4522	0.2347
170	192.5230	5.2602	– 22.5230	– 0.1219
118	130.9849	4.8751	– 12.9849	– 0.1026
90	89.1168	4.4899	0.8832	0.009878
1414	1580.2950	7.3654	– 166.2950	– 0.1092
1198	1075.1687	6.9802	122.8313	0.1102
634	731.5013	6.5951	– 97.5013	– 0.1397
1022	840.5414	6.7340	181.4586	0.2021
620	571.8704	6.3489	48.1296	0.0819
438	389.0774	5.9638	48.9226	0.1208
442	447.0747	6.1027	– 5.0747	– 0.0114
332	304.1715	5.7176	27.8285	0.0888
220	206.9460	5.3325	13.0540	0.0618
3636	3669.7424	8.2079	– 33.7424	– 0.009223
3184	2496.7442	7.8227	687.2558	0.2534
2000	1698.6836	7.4376	301.3164	0.1679
1568	1951.8954	7.5766	– 383.8954	– 0.2113
1070	1327.9906	7.1914	– 257.9906	– 0.2085
566	903.5111	6.8063	– 337.5111	– 0.4339
1140	1038.1916	6.9452	101.8084	0.0950
884	706.3435	6.5601	177.6565	0.2331
360	480.5675	6.1750	– 120.5675	– 0.2756

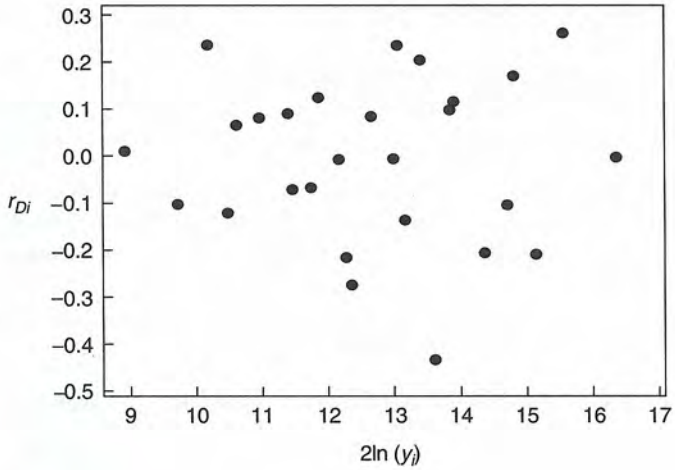
Los modelos se ajustan en la forma acostumbrada y los valores de los parámetros del modelo no se afectan por el valor de ϕ. El parámetro ϕ se puede especificar en forma directa si se conoce su valor, o quizá se pueda estimar replicando algunos de los puntos de datos, también, se puede estimar en forma directa. Un estimado lógico de ϕ es la desviación dividida entre sus grados de libertad. La matriz de covarianza de los coeficientes del modelo se multiplica por ϕ y la desviación escalada y los logaritmos de verosimilitud que se usaron en pruebas de hipótesis se dividen entre ϕ.

La función que se obtiene al dividir un logaritmo de verosimilitud entre ϕ para el caso de distribución de los errores binomial o de Poisson ya no es una función adecuada de logaritmo de verosimilitud, y es un ejemplo de una **función de cuasi verosimilitud**. Por fortuna, la mayor parte de la teoría asintótica de los logaritmos de verosimilitud se aplica a las cuasi verosimilitudes, por lo que se puede justificar el cálculo aproximado de errores estándar y de estadísticos de desviación exactamente como se hizo antes.

http://carlos2524.jimdo.com/



a)



b)

Figura 13.4 Gráficas de los residuales de desviación del modelo lineal general para los datos de la fibra de lana. a) Gráfica de probabilidad normal con los residuales de desviación. b) Gráfica de los residuales de desviación en función de 2 ln($\hat{y}$).

PROBLEMAS

13.1 La tabla siguiente presenta los resultados de disparo de 25 misiles antiaéreos, de superficie a aire, con blancos de diversas velocidades. El resultado de cada prueba es un éxito ($y = 1$) o un fracaso ($y = 0$).

http://carlos2524.jimdo.com/

Prueba	Velocidad del blanco (x), nudos	y	Prueba	Velocidad del blanco (x), nudos	y
1	400	0	14	330	1
2	220	1	15	280	1
3	490	0	16	210	1
4	210	1	17	300	1
5	500	0	18	470	1
6	270	0	19	230	0
7	200	1	20	430	0
8	470	0	21	460	0
9	480	0	22	220	1
10	310	1	23	250	1
11	240	1	24	200	1
12	490	0	25	390	0
13	420	0			

a. Ajustar un modelo de regresión logístico a la variable de respuesta y. Usar un modelo de regresión lineal simple como estructura del predictor lineal.

b. La desviación del modelo ¿indica que es adecuado el modelo logístico de regresión de la parte a?

c. Dé una interpretación del parámetro β_1 en este modelo.

d. Desarrollar el predictor lineal para incluir un término cuadrático en la velocidad del blanco. ¿Hay algún indicio de que ese término cuadrático se requiere en el modelo?

13.2 Se hizo un estudio tratando de relacionar la propiedad de casas con el ingreso familiar. Se seleccionaron 20 familias, se estimó su ingreso y otra información acerca de la propiedad de la vivienda ($y = 1$ indica sí, y $y = 0$ indica no). Los datos se ven a continuación.

Familia	Ingreso	Estado de propiedad	Familia	Ingreso	Estado de propiedad
1	38 000	0	11	38 700	1
2	51 200	1	12	40 100	0
3	39 600	0	13	49 500	1
4	43 400	1	14	38 000	0
5	47 700	0	15	42 000	1
6	53 000	0	16	54 000	1
7	41 500	1	17	51 700	1
8	40 800	0	18	39 400	0
9	45 400	1	19	40 900	0
10	52 400	1	20	52 800	1

a. Ajustar un modelo de regresión logístico a la variable de respuesta y. Usar un modelo de regresión lineal simple como estructura para el predictor lineal.

b. ¿Indica la desviación del modelo que es adecuado el modelo logístico de regresión de la parte a?

c. Dé una interpretación del parámetro β_1 en este modelo.

d. Desarrollar el predictor lineal para incluir un término cuadrático en el ingreso. ¿Hay algún indicio de que se requiere este término cuadrático en el modelo?

http://carlos2524.jimdo.com/

13.3 Se estudió la resistencia a la compresión de un tornillo de aleación, usado en construcción de aviones. Se seleccionaron diez cargas en el intervalo de 2 500 a 4 300 psi, y se probaron con ellas varios tornillos. Se anotó la cantidad de tornillos que fallaron en cada carga. Los datos completos de la prueba se ven a continuación.

Carga (psi), x	Tamaño de muestra, n	Cantidad de fallas, r
2 500	50	10
2 700	70	17
2 900	100	30
3 100	60	21
3 300	40	18
3 500	85	43
3 700	90	54
3 900	50	33
4 100	80	60
4 300	65	51

a. Ajustar un modelo de regresión logístico a los datos. Usar un modelo de regresión lineal simple como estructura para el predictor lineal.

b. ¿Indica la desviación del modelo que es adecuado el modelo logístico de regresión de la parte a?

c. Desarrollar el predictor lineal para incluir un término cuadrático. ¿Hay indicios de que este término cuadrático se requiera en el modelo?

d. Para el modelo cuadrático de la parte c, calcular el estadístico de Wald para cada parámetro individual del modelo.

e. Determinar intervalos aproximados de 95% de confianza para los parámetros del modelo cuadrático de la parte c.

13.4 El departamento de investigación de mercado de un fabricante de bebidas gaseosas investiga la eficacia de un cupón de descuento en la compra de un envase con dos litros de la bebida. A una muestra de 5 500 compradores se les dieron cupones por descuentos variables, entre 5 y 25 centavos. La variable de respuesta fue la cantidad de cupones de cada categoría de descuento presentados después de un mes. Los datos obtenidos se muestran a continuación.

Descuento, x	Tamaño de muestra, n	Cantidad redimida, r
5	500	100
7	500	122
9	500	147
11	500	176
13	500	211
15	500	244
17	500	277
19	500	310
21	500	343
23	500	372
25	500	391

http://carlos2524.jimdo.com/

a. Ajustar a los datos un modelo de regresión logístico. Usar un modelo de regresión lineal simple como estructura del predictor lineal.
b. ¿Indica la desviación del modelo que es adecuado el modelo logístico de regresión de la parte a?
c. Trazar una gráfica de los datos y del modelo ajustado de regresión logística.
d. Desarrollar el predictor lineal para incluir un término cuadrático. ¿Hay algún indicio de que se requiere este término cuadrático en el modelo?
e. Trazar una gráfica de este nuevo modelo, sobre la que se preparó en la parte c. ¿Proporciona visualmente el modelo ampliado mejor ajuste a los datos que el modelo original de la parte a?
f. Para el modelo cuadrático de la parte d, calcular el estadístico de Wald para cada parámetro individual del modelo.
g. Calcular los intervalos aproximados de 95% de confianza para los parámetros del modelo cuadrático de regresión logística de la parte d.

13.5 Se hizo un estudio para investigar compras de automóviles nuevos. Se seleccionó una muestra de 20 familias. Cada familia se encuestó para determinar la edad de su coche más viejo y su ingreso familiar total. Se hizo una encuesta de seguimiento 6 meses después, para determinar si en realidad habían comprado un nuevo vehículo en ese lapso ($y = 1$ indica sí, y $y = 0$ indica no). Los datos del estudio aparecen en la siguiente tabla.

Ingreso, x_1	Edad, x_2	y	Ingreso, x_1	Edad, x_2	y
45 000	2	0	37 000	5	1
40 000	4	0	31 000	7	1
60 000	3	1	40 000	4	1
50 000	2	1	75 000	2	0
55 000	2	0	43 000	9	1
50 000	5	1	49 000	2	0
35 000	7	1	37 500	4	1
65 000	2	1	71 000	1	0
53 000	2	0	34 000	5	0
48 000	1	0	27 000	6	0

a. Ajustar un modelo de regresión logístico a los datos.
b. ¿Indica la desviación del modelo que es adecuado el modelo de regresión logístico de la parte a?
c. Interpretar los coeficientes β_1 y β_2 del modelo.
d. ¿Cuál es la probabilidad estimada de que una familia con ingreso de 45 000 pesos y un coche de 5 años de antigüedad compre un nuevo vehículo durante los siguientes 6 meses?
e. Desarrollar el predictor lineal para incluir un término de interacción. ¿Hay indicios de que se requiera este término en el modelo?
f. Para el modelo de la parte a, calcular los estadísticos para cada parámetro del modelo.
g. Calcular intervalos aproximados de 95% de confianza de los parámetros del modelo, para el modelo logístico de regresión de la parte a.

13.6 Un fabricante de productos químicos ha mantenido registros de la cantidad de fallas de determinada clase de válvulas usada en sus unidades de procesamiento, y el tiempo (meses) transcurrido desde que se instaló la válvula. Los datos se ven a continuación.

http://carlos2524.jimdo.com/

Válvula	Cantidad de fallas	Meses	Válvula	Cantidad de fallas	Meses
1	5	18	9	0	7
2	3	15	10	0	12
3	0	11	11	0	3
4	1	14	12	1	7
5	4	23	13	0	2
6	0	10	14	7	30
7	0	5	15	0	9
8	1	8			

a. Ajustar un modelo de regresión de Poisson a los datos.

b. ¿Indica la desviación del modelo que el modelo de regresión de Poisson de la parte a es adecuado?

c. Trazar una gráfica de los valores ajustados del modelo en función de los meses. También, graficar la cantidad observada de fallas en esa gráfica.

d. Desarrollar el predictor lineal para incluir un término cuadrático. ¿Hay algún indicio de que este término se requiere en el modelo?

e. Para el modelo de la parte a, calcular el estadístico de Wald para cada parámetro del modelo.

f. Calcular intervalos aproximados de 95% de confianza de los parámetros del modelo de regresión de Poisson de la parte a.

13.7 Myers [1990] presenta datos sobre la cantidad de fracturas (y) que se presentan en los mantos superiores de minas de carbón en la región de los Apalaches, del oeste de Virginia. Se determinaron cuatro regresores: x_1 = espesor (pies) de la carga interior, que es la distancia más corta entre el piso del estrato y el manto inferior; x_2 = porcentaje de extracción del manto previamente explotado; x_3 = altura del manto inferior (pies) y x_4 = tiempo (años) de operación de la mina. Los datos obtenidos se muestran a continuación.

Observación	y (cantidad de fracturas por subregión)	x_1	x_2	x_3	x_4
1	2	50	70	52	1.0
2	1	230	65	42	6.9
3	0	125	70	45	1.0
4	4	75	65	68	0.5
5	1	70	65	53	0.5
6	2	65	70	46	3.0
7	0	65	60	62	1.0
8	0	350	60	54	0.5
9	4	350	90	54	0.5
10	4	160	80	38	0.0
11	1	145	65	38	10.0
12	4	145	85	38	0.0
13	1	180	70	42	2.0
14	5	43	80	40	0.0
15	2	42	85	51	12.0
16	5	42	85	51	0.0

http://carlos2524.jimdo.com/

Observación	y (cantidad de fracturas por subregión)	x_1	x_2	x_3	x_4
17	5	45	85	42	0.0
18	5	83	85	48	10.0
19	0	300	65	68	10.0
20	5	190	90	84	6.0
21	1	145	90	54	12.0
22	1	510	80	57	10.0
23	3	65	75	68	5.0
24	3	470	90	90	9.0
25	2	300	80	165	9.0
26	2	275	90	40	4.0
27	0	420	50	44	17.0
28	1	65	80	48	15.0
29	5	40	75	51	15.0
30	2	900	90	48	35.0
31	3	95	88	36	20.0
32	3	40	85	57	10.0
33	3	140	90	38	7.0
34	0	150	50	44	5.0
35	0	80	60	96	5.0
36	2	80	85	96	5.0
37	0	145	65	72	9.0
38	0	100	65	72	9.0
39	3	150	80	48	3.0
40	2	150	80	48	0.0
41	3	210	75	42	2.0
42	5	11	75	42	10.0
43	0	100	65	60	25.0
44	3	50	88	60	20.0

a. Ajustar un modelo de regresión de Poisson a estos datos, con cadena logarítmica.
b. ¿Indica la desviación del modelo que es satisfactorio el modelo de la parte a?
c. Hacer un análisis parcial de desviación tipo 3 de los parámetros del modelo. ¿Indica el análisis que se pueden eliminar algunos regresores del modelo?
d. Calcular estadísticos de Wald para probar la contribución de cada regresor al modelo. Interpretar los resultados de esos estadísticos de prueba.
e. Calcular intervalos aproximados de 95% de confianza de Wald, para los parámetros del modelo.

13.8 Regresar a los datos de fractura de mina del problema 13.7. Eliminar del modelo original todos los regresores que se crea no tienen importancia y volver a hacer las partes b a e de ese problema. Comente sus resultados.

13.9 Regresar a los datos de fractura de mina de los problemas 13.7 y 13.8. Trazar gráficas de residuales de desviaciones para el mejor modelo que se haya encontrado, y comentar las gráficas. ¿Parece ser satisfactorio el modelo, desde el punto de vista del análisis de residuales?

13.10 Regresar al modelo de los datos de compra de automóviles, en el problema 13.5, parte a. Trazar gráficas de residuales de desviación para el modelo, y comentar esas gráficas. ¿Parece satisfactorio el modelo, desde el punto de vista de análisis de residuales?

http://carlos2524.jimdo.com/

13.11 Volver al modelo para los datos de los cupones de bebidas gaseosas en el problema 13.4, parte a. Trazar gráficas de residuales de desviación para el modelo, y comentar esas gráficas. ¿Parece satisfactorio el modelo, desde el punto de vista de análisis de residuales?

13.12 Volver al modelo de los datos de tornillos para aviones, del problema 13.3, parte a. Trazar gráficas de los residuales de desviación para el modelo, y comentar esas gráficas. ¿Parece adecuado el modelo, desde el punto de vista de análisis de residuales?

13.13 La función de densidad gamma es

$$f(y, r, \lambda) = \frac{\lambda^r}{\Gamma(r)} e^{-\lambda y} y^{r-1}, \quad \text{para } y, \lambda \geq 0$$

Demostrar que la distribución gamma es un miembro de la familia exponencial.

13.14 La función de densidad exponencial es

$$f(y, \lambda) = \lambda e^{-\lambda y}, \quad \text{para } y, \lambda \geq 0$$

Demostrar que la distribución exponencial es un miembro de la familia exponencial.

13.15 La función de masa binomial negativa es

$$f(y, \pi, \alpha) = \binom{y + \alpha - 1}{\alpha - 1} \pi^{\alpha} (1 - \pi)^{y},$$

$$\text{para } y = 0, 1, 2, \ldots, \alpha > 0 \quad \text{y} \quad 0 \leq \pi \leq 1$$

Demostrar que la binomial negativa es un miembro de la familia exponencial.

13.16 Los datos de la tabla de abajo son de un experimento para estudiar la velocidad de avance y de una barrena. Los cuatro factores de diseño son x_1 = carga, x_2 = flujo, x_3 = rapidez de taladrado y x_4 = tipo de lodo de perforación (el experimento original se describe en el libro de Cuthbert Daniel, de 1976, sobre experimentación industrial).

Observación	x_1	x_2	x_3	x_4	Velocidad de avance, y
1	−	−	−	−	1.68
2	+	−	−	−	1.98
3	−	+	−	−	3.28
4	+	+	−	−	3.44
5	−	−	+	−	4.98
6	+	−	+	−	5.70
7	−	+	+	−	9.97
8	+	+	+	−	9.07
9	−	−	−	+	2.07
10	+	−	−	+	2.44
11	−	+	−	+	4.09
12	+	+	−	+	4.53
13	−	−	+	+	7.77
14	+	−	+	+	9.43
15	−	+	+	+	11.75
16	+	+	+	+	16.30

http://carlos2524.jimdo.com/

a. Ajustar un modelo lineal generalizado de la respuesta velocidad de avance. Usar una distribución gamma de respuestas, y una cadena logarítmica, e incluir los cuatro regresores en el predictor lineal.

b. Calcular la desviación para el modelo lineal generalizado de la parte a. ¿Indica que el modelo es satisfactorio?

c. Hacer un análisis parcial de desviación tipo 3 de los parámetros del modelo. ¿Indica el análisis que se podrían sacar del modelo algunos regresores?

d. Calcular el estadístico de Wald para probar la contribución de cada regresor al modelo. Interpretar los resultados de esos estadísticos de prueba.

e. Calcular intervalos aproximados de 95% de confianza de Wald para los parámetros del modelo.

13.17 Regresar a los datos de perforación del problema 13.16. Eliminar del modelo original los regresores que se consideren sin importancia, y volver a hacer las partes b a e de ese problema. Comentar los resultados.

13.18 Regresar a los datos de perforación en el problema 13.16. Ajustar un modelo lineal general con la cadena logarítmica y la distribución gamma, pero desarrollar el predictor lineal para incluir las seis interacciones de dos factores con los cuatro regresores originales. Comparar la desviación de este modelo con la del "modelo sólo con los efectos principales" del problema 13.16. ¿Parece útil agregar los términos de interacción?

13.19 Regresar a los datos de perforación en el problema 13.16. Trazar gráficas de los residuales de desviación para el modelo, y comentarlas. ¿Parece satisfactorio el modelo, desde el punto de vista del análisis de residuales?

13.20 La tabla siguiente muestra los valores calculados y los residuales de desviación para el modelo de regresión de Poisson, usando x_2 = carga de la bomba como regresor de ajuste para los datos de daños por bombardeo en el ejemplo 13.7. Graficar los residuales y comentar la adecuación del modelo.

y	$\hat{y}$	$x_i'\hat{\beta}$	e_i	r_{Pi}
0	0.4789	−0.7364	−0.4789	−0.9786
1	0.4789	−0.7364	0.5211	0.6561
0	0.4789	−0.7364	−0.4789	−0.9786
0	0.6016	−0.5083	−0.6016	−1.0969
0	0.6016	−0.5082	−0.6016	−1.0969
0	0.6016	−0.5082	−0.6016	−1.0969
1	0.7558	−0.2800	0.2442	0.2675
0	0.7558	−0.2800	−0.7558	−1.2295
0	0.7558	−0.2800	−0.7558	−1.2295
2	0.9495	−0.0518	1.0505	0.9374
0	0.9495	−0.0518	−0.9495	−1.3781
1	0.9495	−0.0518	0.0505	0.0513
1	1.1929	0.1764	−0.1929	−0.1818
1	1.1929	0.1764	−0.1929	−0.1818
2	1.1929	0.1764	0.8071	0.6729
3	0.9495	−0.0518	2.0505	1.6737
1	0.9495	−0.0518	0.0505	0.0513
1	0.9495	−0.0518	0.0505	0.0513
1	1.8829	0.6328	−0.8829	−0.7072
2	1.8829	0.6328	0.1171	0.0845
0	1.8829	0.6328	−1.8829	−1.9406

http://carlos2524.jimdo.com/

y	$\hat{y}$	$x_i'\hat{\beta}$	e_i	r_{Pi}
1	2.9719	1.0892	– 1.9719	– 1.3287
1	2.9719	1.0892	– 1.9719	– 1.3287
2	2.9719	1.0892	– 0.9719	– 0.5996
5	1.1929	0.1764	3.8071	2.5915
1	1.1929	0.1764	– 0.1929	– 0.1818
3	1.1929	0.1764	1.8071	1.3853
5	4.6907	1.5456	0.3093	0.1413
5	4.6907	1.5456	0.3093	0.1413
7	4.6907	1.5456	2.3093	0.9930

13.21 Un modelo logístico de regresión con un predictor lineal que incluye un término de interacción es $\mathbf{x}'\beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$. Deducir una ecuación del cociente de ventaja para el regresor x_1. ¿Tiene este cociente la misma interpretación que en el caso en el que el predictor lineal **no** tiene término de interacción?

13.22 La teoría de la máxima verosimilitud establece que la covarianza estimada con muestras grandes, para los estimados de máxima verosimilitud, es la inversa de la **matriz de información**, en la que sus elementos son los negativos de los valores esperados de las segundas derivadas parciales de la función de logaritmo de verosimilitud, evaluada en los estimados de máxima verosimilitud. Considérese el modelo de regresión lineal con errores normales. Determinar la matriz de información y la matriz de covarianza de los estimados de máxima verosimilitud.

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

CAPÍTULO 14

OTROS TEMAS ACERCA DEL USO DEL ANÁLISIS DE REGRESIÓN

En este capítulo se presentan diversos temas que surgen al usar el análisis de regresión. En algunos casos sólo se reseña un vistazo sobre el tema, con referencias de presentaciones más completas.

14.1 MODELOS DE REGRESIÓN CON ERRORES AUTOCORRELACIONADOS

14.1.1 Causas y efectos de la autocorrelación

Las hipótesis fundamentales de la regresión lineal son que los términos de error ε_i tienen media cero, varianza constante, y no están correlacionados: [$E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2$ y $E(\varepsilon_i\varepsilon_j) = 0$]. Para fines de probar hipótesis y establecer intervalos de confianza, con frecuencia se agrega la suposición de normalidad, por lo que las ε_i son NID(0, σ^2). Algunas aplicaciones de la regresión implican variables regresoras y de respuesta que tienen un orden natural en el tiempo; esos datos se llaman **datos de serie de tiempo**, o **de serie temporal**. Los modelos de regresión que usan datos de series de tiempo se presentan con relativa frecuencia en economía, administración y algunos campos de la ingeniería. La suposición de errores independientes o no correlacionados muchas veces no es adecuada para los datos de series de tiempo. Por lo general, los errores en los datos de series de tiempo tienen **correlación seriada**, esto es, $E(\varepsilon_i\varepsilon_{i+j}) \neq 0$. Se dice que esos términos de error están **autocorrelacionados**.

Existen varias causas de autocorrelación, quizá la principal, en los problemas de regresión donde intervienen datos de series de tiempo es el no incluir uno o más regresores importantes en el modelo. Por ejemplo, supóngase que se trata de hacer la regresión de las ventas anuales de un concentrado de gaseosa en función de los gastos anuales de promoción de ese producto. Ahora bien, el crecimiento de la población durante el tiempo usado en el estudio, también influirá sobre las ventas del producto. Si el modelo no incluye el tamaño de la población, puede hacer que sus errores (del modelo) tengan autocorrelación positiva, porque el tamaño de la población tiene correlación positiva con las ventas del producto.

La presencia de autocorrelación en los errores tiene diversos efectos sobre el procedimiento ordinario de regresión por mínimos cuadrados. Estos efectos se resumen a continuación:

1. Los coeficientes de regresión ordinaria por mínimos cuadrados siguen siendo insesgados, pero ya no son estimados con varianza mínima. Se dice entonces que esos estimados son ineficientes.

http://carlos2524.jimdo.com/

2. Cuando los errores están autocorrelacionados positivamente, el cuadrado medio de los residuales, MS_{Res}, puede subestimar mucho a σ^2, en consecuencia, los errores estándar de los coeficientes de regresión pueden ser muy pequeños. Así, los intervalos de confianza son más cortos de lo que deberían ser, y las pruebas de hipótesis acerca de los coeficientes individuales de regresión pueden indicar que uno o más de los regresores contribuyen en forma significativa al modelo, cuando en realidad no contribuyen. En general, al subestimar σ^2, el analista se queda con una falsa impresión de exactitud.
3. Los intervalos de confianza y las pruebas de hipótesis basados en las distribuciones t y F ya no son adecuados, en el sentido estricto.

Hay tres métodos para manejar el problema de la autocorrelación. Si hay autocorrelación porque se haya omitido un regresor, y se puede identificar e incluir ese regresor en el modelo, la autocorrelación aparente debe desaparecer. Si no se puede resolver el problema de autocorrelación incluyendo factores omitidos antes, el analista debe buscar un modelo que incorpore en forma específica la estructura de autocorrelación. Esos modelos suelen requerir técnicas especiales de estimación de parámetros. Por último, se podrían usar métodos de mínimos cuadrados ponderados o generalizados si se conociera suficientemente la estructura de autocorrelación.

Como en administración y economía se presentan con frecuencia series de tiempo, gran parte de la metodología básica aparece en las publicaciones sobre economía. Unas buenas referencias sobre **econometría** (métodos matemáticos y estadísticos en economía) son las de Johnston [1972], Wonnacott y Wonnacott [1970] y Pindyck y Rubinfeld [1991]. Se pueden ver otros métodos para modelar y analizar series de tiempo en Box, Jenkins y Reinsel [1994], Fuller [1976] y Montgomery, Johnson y Gardiner [1990].

14.1.2 Detección de la presencia de autocorrelación

Para detectar la autocorrelación suelen ser útiles las **gráficas de residuales**. La presentación más adecuada es la de los residuales en función del tiempo. Si hay autocorrelación positiva, los residuales de igual signo se presentarán en grupos, que indica que no hay los suficientes cambios de signo en la secuencia de los residuales. Por otra parte, si hay autocorrelación negativa, los residuales cambiarán de signo con demasiada rapidez.

Para detectar la presencia de la autocorrelación se pueden aplicar diversas **pruebas estadísticas**. La que desarrollaron Durbin y Watson [1950, 1951, 1971] se usa ampliamente, ya que se basa en la hipótesis de que los errores del modelo de regresión se generan en un **proceso autorregresivo de primer orden**, que se observa a intervalos de tiempo igualmente espaciados, esto es,

$$\varepsilon_t = \rho\varepsilon_{t-1} + a_t \tag{14.1}$$

donde ε_t es el término de error en el modelo, en el periodo t; a_t es una variable aleatoria NID$(0, \sigma^2{}_a)$ y ρ $(|\rho| < 1)$ es el **parámetro de autocorrelación**. Así, un modelo de regresión lineal simple con **errores autorregresivos de primer orden** sería

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \varepsilon_t \\ \varepsilon_t &= \rho\varepsilon_{t-1} + a_t \end{aligned} \tag{14.2}$$

http://carlos2524.jimdo.com/

siendo y_t y x_t las observaciones de las variables de respuesta y regresión en el periodo t.

Cuando los errores del modelo de regresión se generan por el proceso autorregresivo de primer orden, como en la ecuación (14.1), varias de sus propiedades interesantes se pueden desarrollar. Al sustituir en forma sucesiva $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots,$ en el lado derecho de la ecuación (14.1), se obtiene

$$\varepsilon_t = \sum_{u=0}^{\infty} \rho^u a_{t-u}$$

Así, el término de error para el periodo t sólo es una combinación lineal de la realización actual y todas las anteriores, de las variables aleatorias a_t, NID(0, σ_a^2), además, se puede demostrar que

$$E(\varepsilon_t) = 0 \tag{14.3a}$$

$$\text{Var}(\varepsilon_t) = \sigma_a^2 \left(\frac{1}{1-\rho^2} \right) \tag{14.3b}$$

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+u}) = \rho^{|u|}\sigma^2 \left(\frac{1}{1-\rho^2} \right) \tag{14.3c}$$

Esto es, los errores tienen media cero y varianza constante, pero están autocorrelacionados, a menos que $\rho = 0$.

Como la mayor parte de los problemas de regresión donde intervienen las series de tiempo tienen autocorrelación positiva, las hipótesis que se suelen considerar en la prueba de Durbin-Watson son:

$$H_0: \rho = 0$$
$$H_1: \rho > 0 \tag{14.4}$$

El estadístico de prueba es

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{14.5}$$

en donde las e_t, $t = 1, 2, \ldots, n$ son los residuales de un análisis de mínimos cuadrados ordinarios aplicado a los datos (y_t, x_t). Desafortunadamente, la distribución de d depende de la matriz **X**, sin embargo, Durbin y Watson [1951] demostraron que d está entre dos cotas, digamos que d_L y d_U, tales que si d sale de esos límites, se puede llegar a una conclusión acerca de las hipótesis de las ecuaciones (14.4). El procedimiento de decisión es el siguiente:

Si $d < d_L$	rechazar $H_0: \rho = 0$
Si $d > d_U$	no rechazar $H_0: \rho = 0$
Si $d_L \leq d \leq d_U$	la prueba no es concluyente

http://carlos2524.jimdo.com/

Es claro que los valores pequeños de d implican que H_0: $\rho = 0$ se deben rechazar, porque la autocorrelación positiva indica que los términos consecutivos de error son de magnitud parecida, y las diferencias en los residuales, $e_t - e_{t-1}$ serán pequeñas. Durbin y Watson sugieren varios procedimientos para decidir cuándo los resultados son no concluyentes. En esos casos, un método razonable es analizar los datos con los métodos de la sección 14.1.3 para ver si se presentan cambios importantes en los resultados. La tabla A.6 del apéndice muestra los límites d_L y d_U para varios tamaños de muestra, diversas cantidades de regresores y tres tasas de error tipo I ($\alpha = 0.05$, $\alpha = 0.025$ y $\alpha = 0.01$).

No se encuentran con frecuencia los casos con autocorrelación negativa, sin embargo, si se desea una prueba para la autocorrelación negativa, se puede usar el estadístico $4 - d$, definiendo a d con la ecuación (14.5). Entonces, las reglas de decisión para H_0: $\rho = 0$ contra H_1: $\rho < 0$ son las mismas que las que se usan para probar una autocorrelación positiva. También es posible hacer una prueba bilateral (H_0: $\rho = 0$ contra H_1: $\rho \uparrow 0$) usando ambas pruebas unilaterales al mismo tiempo. Al hacerlo, el procedimiento bilateral tiene error de tipo 1 igual a 2α, siendo α el error de tipo I que se usó para cada prueba unilateral.

Ejemplo 14.1 *Ventas de concentrado para bebidas gaseosas*

Una empresa fabricante de bebidas gaseosas desea pronosticar las ventas anuales regionales del concentrado de uno de sus productos, en función de los gastos de promoción regional de ese producto. En las columnas 1 y 2 de la tabla 14.1 se ven los datos de 20 años. Suponiendo que sea adecuada una relación lineal, se ajustó un modelo lineal de regresión con los mínimos cuadrados ordinarios. En la columna 3 de la tabla 14.1 se ven los residuales de este modelo rectilíneo, y en la tabla 14.2 se presentan otros estadísticos de resumen para el modelo. Como las variables regresoras y de respuestas son de serie temporal, se cree que pueda haber autocorrelación. En la figura 14.1 se ve la gráfica de los residuales en función del tiempo. Esa gráfica es perturbadora, porque hay un desplazamiento definido, primero hacia arriba y después hacia abajo, en los residuales. La autocorrelación podría ser la responsable de ese comportamiento.

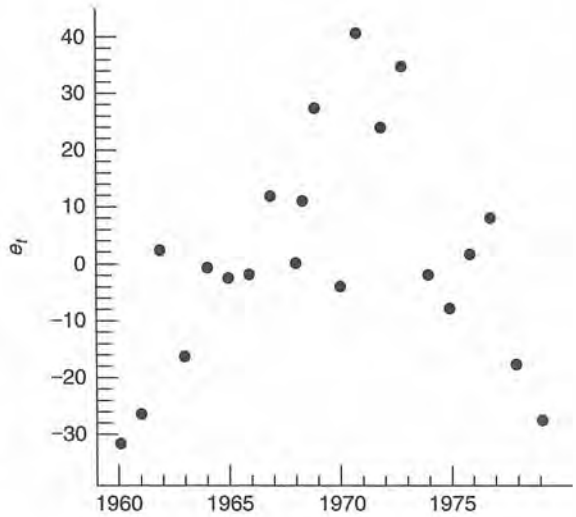


Figura 14.1 Residuales, e_t, en función del tiempo, ejemplo 14.1.

http://carlos2524.jimdo.com/

TABLA 14.1 Datos del ejemplo de ventas de concentrado de bebida gaseosa

		(1) Ventas anuales regionales de concentrado (unidades)	(2) Gastos anuales de publicidad ($ × 1 000)	(3) Residuales de mínimos cuadrados	(4)	(5)	(6) Población regional anual
	t	y_t	x_t	e_t	e_t^2	$(e_t - e_{t-1})^2$	z_t
1960	1	3083	75	– 32.330	1045.2289		825000
1961	2	3149	78	– 26.603	707.7196	32.7985	830445
1962	3	3218	80	2.215	4.9062	830.4771	838750
1963	4	3239	82	– 16.967	287.8791	367.9491	842940
1964	5	3295	84	– 1.148	1.3179	250.2408	846315
1965	6	3374	88	– 2.512	6.3101	1.8605	852240
1966	7	3475	93	– 1.967	3.8691	0.2970	860760
1967	8	3569	97	11.669	136.1656	185.9405	865925
1968	9	3597	99	– 0.513	0.2632	148.4011	871640
1969	10	3725	104	27.032	730.7290	758.7270	877745
1970	11	3794	109	– 4.422	19.5541	989.3541	886520
1971	12	3959	115	40.032	1602.5610	1976.1581	894500
1972	13	4043	120	23.577	555.8749	270.7670	900400
1973	14	4194	127	33.940	1151.9236	107.3918	904005
1974	15	4318	135	– 2.787	7.7674	1348.8725	908525
1975	16	4493	144	– 8.606	74.0632	33.8608	912160
1976	17	4683	153	0.575	0.3306	84.2908	917630
1977	18	4850	161	6.848	46.8951	39.3505	922220
1978	19	5005	170	– 18.971	359.8988	666.6208	925910
1979	20	5236	182	– 29.063	844.6580	101.8485	929610
		$\sum_{t=1}^{20} e_t^2 = 7\,587.9154$		$\sum_{t=2}^{20} (e_t - e_{t-1})^2 = 8\,195.2065$			

TABLA 14.2 Estadísticos resumen para el modelo de mínimos cuadrados del ejemplo 14.1

Parámetro	Estimado	Error estándar	Estadístico t
β_0	1608.508	17.0223	94.49
β_1	20.091	0.1428	140.71
$n = 20$	$R^2 = 0.9991$	$MS_{\text{Res}} = 421.5485$	

También se usará la prueba de Durbin-Watson para

$$H_0\colon \rho = 0$$

$$H_1\colon \rho > 0$$

Las columnas 4 y 5 de la tabla 1 contienen los cálculos necesarios para obtener el estadístico de prueba

http://carlos2524.jimdo.com/

$$d = \frac{\sum_{t=2}^{20} (e_t - e_{t-1})^2}{\sum_{t=1}^{20} e_t^2} = \frac{8\,195.2065}{7\,587.9154} = 1.08$$

Si se opta por $\alpha = 0.05$, la tabla del apéndice A.6 muestra los valores críticos correspondientes para $n = 20$ y un regresor, que son $d_L = 1.20$ y $d_U = 1.41$. Como el valor observado de $d = 1.08$ es menor que $d_L = 1.20$, se rechaza H_0 y se llega a la conclusión que los errores tienen autocorrelación positiva.

Aunque la prueba de Durbin-Watson tiene una gran utilidad, también tiene sus limitaciones. En el procedimiento se supone que los errores se generan mediante un proceso **autorregresivo de primer orden**. La prueba no necesariamente detecta la autocorrelación cuando la estructura en los errores sea distinta de la autorregresiva de primer orden.

Cuando se manejan datos de serie de tiempo, a veces se introducen como regresores **valores retrasados** de la variable de respuesta. Por ejemplo, se podría usar

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + \varepsilon_t$$

como modelo. Si se aplican los mínimos cuadrados ordinarios en modelos de este tipo, el estimador $\hat{\beta}$ que resulta es sesgado, sin embargo, si los errores no están correlacionados, $\hat{\beta}$ es un estimador consistente de β. Si los términos de error están correlacionados, lo anterior no necesariamente es cierto. Para conocer más acerca de los modelos con variables retrasadas, véase Goldberger [1964]. Además, si el modelo contiene variables retrasadas, la prueba de Durbin-Watson ya no es adecuada. Durbin [1970] presenta una prueba de autocorrelación con muestra grande para modelos con variables retrasadas.

14.1.3 Métodos de estimación de parámetros

Un valor significativo en el estadístico de Durbin-Watson, o una gráfica dudosa de residuales, indica que hay un error de especificación del modelo. Esta mala especificación podría ser una dependencia real de los errores respecto al tiempo, o una dependencia "artificial", causada por la omisión de un regresor importante. Si la autocorrelación aparente se debe a regresores faltantes, y si se pueden identificar e incorporar al modelo esos regresores faltantes, se podrá eliminar la autocorrelación aparente. Esto se ilustra en el siguiente ejemplo.

Ejemplo 14.2 *Ventas de concentrado para bebidas gaseosas*

Se tienen los datos de las ventas de concentrado para bebidas gaseosas que se presentaron en el ejemplo 14.1. La prueba de Durbin-Watson ha indicado que los errores en el modelo de regresión lineal, que relaciona las ventas de concentrado con los gastos de promoción, tienen autocorrelación positiva. En este problema es relativamente fácil imaginar otros regresores probables que pueden estar positivamente correlacionados con las ventas. Por ejemplo, es muy probable que la población de la región afecte las ventas de concentrado. En la columna 6 de la tabla 14.1 se muestran datos sobre la población de la región durante los años 1960 a 1979. Si se agrega este regresor al modelo, la ecuación tentativa será

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \varepsilon_t$$

http://carlos2524.jimdo.com/

La tabla 14.3 contiene los estadísticos de resumen para el análisis de esos datos por mínimos cuadrados. Se ve en esa tabla que el estadístico de Durbin-Watson es $d = 3.06$, porque el 5% de los valores críticos, que son $d_L = 1.10$ y $d_U = 1.54$, harían llegar a la conclusión de que no hay autocorrelación positiva en los errores. La gráfica de residuales en función del tiempo se ve en la figura 14.2, y mejoró mucho, en comparación con la figura 14.1; por consiguiente, al agregar el tamaño de la población al modelo se ha eliminado el problema aparente de autocorrelación.

TABLA 14.3 Estadísticos de resumen para el modelo de mínimos cuadrados del ejemplo 14.2

Parámetro	Estimado	Error estándar	Estadístico t
β_0	320.340	217.3278	1.47
β_1	18.434	0.2915	63.23
β_2	0.002	0.0003	5.93
$n = 20$	$R^2 = 0.9997$	$d = 3.06$	$MS_{\text{Res}} = 145.3408$

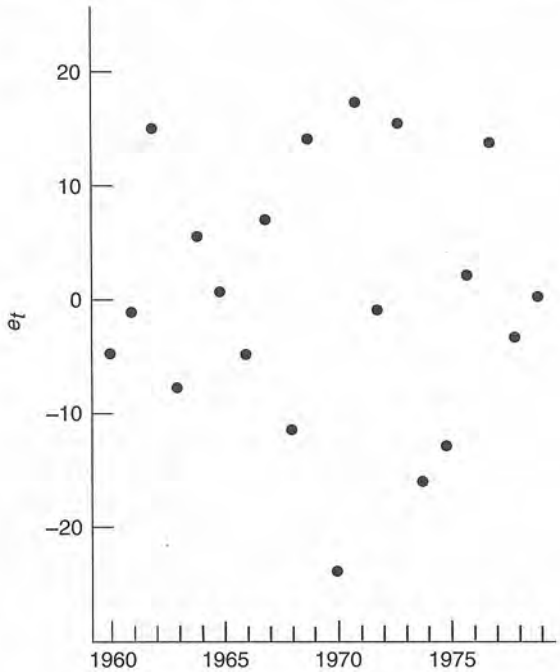


Figura 14.2 Residuales, e_t, en función del tiempo, ejemplo 14.2.

Cuando no se puede eliminar la autocorrelación aparente de los modelos agregando uno o más regresores nuevos al modelo, es necesario reconocer en forma explícita la estructura autocorrelativa del modelo, e inventar un método adecuado de estimación de parámetros. Hay varios procedimientos de estimación que se pueden usar (por ejemplo, véanse Johnson [1972] y Pesaran y Slater [1980]). Aquí se presentará el método descrito por Cochrane y Orcutt [1949].

Considérese el modelo de regresión lineal simple con errores autocorrelacionados de primer orden, que siguen la ecuación (14.2). Supóngase que se transforma la variable de respuesta, de tal manera que $y_t' = y_t - \rho\, y_{t-1}$. Al sustituir y_t y y_{t-1}, el modelo se transforma en

http://carlos2524.jimdo.com/

$$
\begin{aligned}
y_t' &= y_t - \rho y_{t-1} \\
&= \beta_0 + \beta_1 x_t + \varepsilon_t - \rho(\beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}) \\
&= \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + \varepsilon_t - \rho\varepsilon_{t-1} \\
&= \beta_0' + \beta_1' x_t' + a_t
\end{aligned} \tag{14.6}
$$

en donde $\beta_0' = \beta_0(1-\rho)$, $\beta_1' = \beta_1$, $x_t' = x_t - \rho x_{t-1}$ y $a_t = \varepsilon_t - \rho\varepsilon_{t-1}$. Nótese que los términos a_t de error en el modelo reparametrizado son variables aleatorias independientes [véase la Ec. (14.1)]. Por consiguiente, al transformar las variables regresoras y de respuesta se produce un modelo que satisface las suposiciones acostumbradas de la regresión y se pueden aplicar mínimos cuadrados ordinarios.

Desafortunadamente no se puede usar en forma directa el modelo reparametrizado, ecuación (14.6), porque x_t' y y_t' son funciones del parámetro desconocido ρ; sin embargo, se puede considerar que el proceso autorregresivo de primer orden $\varepsilon_t = \rho\varepsilon_{t-1} + a_t$ es una regresión que pasa por el origen. Así, se puede estimar ρ calculando los residuales e_t de una regresión por mínimos cuadrados ordinarios, de y_t respecto a x_t, para después hacer la regresión de e_t respecto a e_{t-1}. El estimado ρ que resulta con mínimos cuadrados es

$$
\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \tag{14.7}
$$

con este esquema de ρ se obtienen las variables regresoras y de respuesta, transformadas:

$$
\begin{aligned}
x_t' &= x_t - \hat{\rho} x_{t-1} \\
y_t' &= y_t - \hat{\rho} y_{t-1}
\end{aligned}
$$

y se aplican mínimos cuadrados ordinarios a los datos transformados. Se debe aplicar la prueba de Durbin-Watson a los residuales del modelo reparametrizado. Si de acuerdo con este procedimiento los residuales no están correlacionados, no se necesita más análisis; no obstante, si todavía hay autocorrelación positiva, será necesaria otra iteración. En la segunda iteración se estima ρ con nuevos residuales obtenidos con los coeficientes de regresión del modelo reparametrizado, con las variables originales regresoras y de respuesta. Se continúa este proceso iterativo, hasta que los términos de error en el modelo reparametrizado no estén correlacionados.

No siempre es bueno este procedimiento iterativo, una razón de que el procedimiento no elimine la autocorrelación en los errores, es que $\hat{\rho}$, el estimado de ρ, esté sesgado hacia abajo. Se sugiere que si una o dos iteraciones no producen errores no correlacionados, el analista debería considerar otras técnicas de estimación. Una posibilidad es estimar en forma simultánea a β_0, β_1 y ρ, minimizando

$$
S(\beta_0, \beta_1, \rho) = \sum_{t=2}^{n} [y_t - \rho y_{t-1} - \beta_0(1-\rho) - \beta_1(x_t - \rho x_{t-1})]^2 \tag{14.8}
$$

Éste es un problema no lineal de mínimos cuadrados. Nótese que se pueden aplicar procedimientos directos de búsqueda para minimizar $S(\beta_0, \beta_1, \rho)$. Por ejemplo, se podría aplicar

http://carlos2524.jimdo.com/

cualquier técnica de búsqueda unidimensional (véase Wilde y Beightler [1967, Cap. 6]) para seleccionar valores de ρ, para después aplicar mínimos cuadrados lineales para estimar β_0 y β_1. Este proceso se repite usando los valores de ρ especificados por la búsqueda unidimensional hasta llegar a un valor mínimo de $S(\beta_0, \beta_1, \rho)$. Otra posibilidad es suponer que $\rho = 1$, obtener las variables transformadas $y_t' = y_t - y_{t-1}$, y $x_t' = x_t - x_{t-1}$, para hacer la regresión de y_t' respecto a x_t', pasando por el origen. A este procedimiento se le llama **método de las primeras diferencias**. En Neter, Kutner, Nachtsheim y Wasserman [1966] se describe un ejemplo.

Ejemplo 14.3 *El segmento del mercado de una crema dental*

Los datos de las columnas 1 y 2 de la tabla 14.4 representan el segmento del mercado y_t, de cierta marca de crema dental, y el precio de venta, x_t, por libra, para 20 meses consecutivos. Se desea formar un modelo de regresión que relacione la parte del mercado en el periodo t con el precio de venta en ese periodo; en forma tentativa se supone un modelo rectilíneo. Al aplicar mínimos cuadrados, el modelo ajustado resulta ser

$$\hat{y}_t = 26.90989 - 24.28977x_t$$

En la tabla 14.5 se muestran los estadísticos de resumen para este modelo.

TABLA 14.4 Datos para el ejemplo 14.3

Periodo t	(1) Parte del mercado (%) y_t	(2) Precio por libra ($) x_t	(3) e_t	(4) $e_t e_{t-1}$	(5) e_t^2
1	3.63	0.97	0.2812		0.0791
2	4.20	0.95	0.3654	0.1028	0.1335
3	3.33	0.99	0.4670	0.1706	0.2181
4	4.54	0.91	−0.2662	−0.1243	0.0709
5	2.89	0.98	−0.2159	0.0575	0.0466
6	4.87	0.90	−0.1791	0.0387	0.0321
7	4.90	0.89	−0.3920	0.0702	0.1537
8	5.29	0.86	−0.7307	0.2864	0.5339
9	6.18	0.85	−0.0836	0.0611	0.0070
10	7.20	0.82	0.2077	−0.0174	0.0431
11	7.25	0.79	−0.4710	−0.0978	0.2218
12	6.09	0.83	−0.6594	0.3106	0.4348
13	6.80	0.81	−0.4352	0.2870	0.1894
14	8.65	0.77	0.4432	−0.1929	0.1964
15	8.43	0.76	−0.0197	−0.0087	0.0004
16	8.29	0.80	0.8119	−0.0160	0.6592
17	7.18	0.83	0.4306	0.3496	0.1854
18	7.90	0.79	0.1790	0.0771	0.0320
19	8.45	0.76	0.0003	0.0001	0.0000
20	8.23	0.78	0.2661	0.0001	0.0708
			$\sum_{t=2}^{20} e_t e_{t-1} = 1.3547$	$\sum_{t=1}^{20} e_t^2 = 3.3082$	

http://carlos2524.jimdo.com/

TABLA 14.5 Estadísticos de resumen para el modelo de mínimos cuadrados ordinarios del ejemplo 14.3

Coeficiente	Estimado	Error estándar	t_0
β_0	26.90989	1.1099	24.25
β_1	− 24.28977	1.2978	− 18.72
	$MS_{Res} = 0.1838$	$R^2 = 0.95$	$F_0 = 350.29$

La columna 3 de la tabla 14.4 contiene los residuales para este modelo, y una gráfica de los residuales en secuencia temporal se ve en la figura 14.3. La gráfica de residuales sugiere algo de autocorrelación positiva. Esto se confirma al calcular el estadístico Durbin-Watson, $d = 1.14$, que al compararse con los valores críticos para 5% y $n = 20$, $d_L = 1.20$ y $d_U = 1.41$, se ve que los residuales están positivamente autocorrelacionados.

Se usará el procedimiento Cochrane-Orcutt para estimar los parámetros del modelo. Las columnas 4 y 5 de la tabla 14.4 contienen los detalles necesarios para estimar el parámetro de autocorrelación ρ con la ecuación (14.7):

$$\hat{\rho} = \frac{\sum_{t=2}^{20} e_t e_{t-1}}{\sum_{t=1}^{20} e_t^2} = \frac{1.3547}{3.3082} = 0.409$$

A continuación, las variables transformadas se calculan con

$$x_t' = x_t - 0.409x_{t-1}, \qquad y_t' = y_t - 0.409y_{t-1}$$

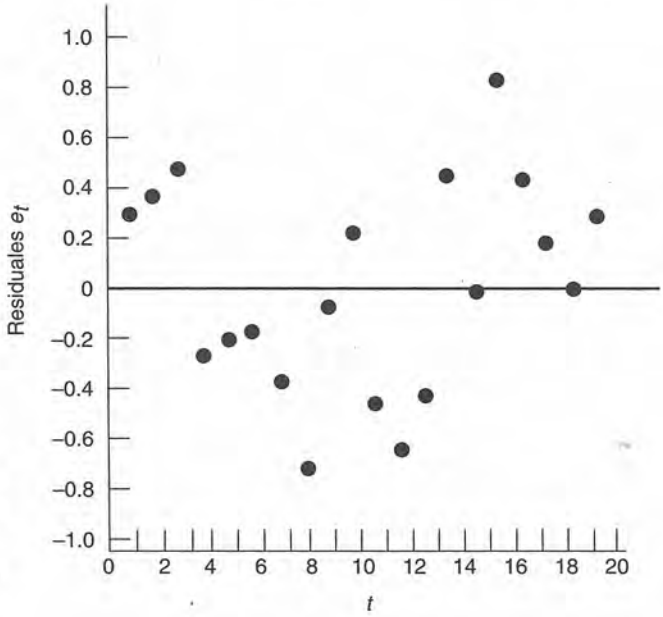


Figura 14.3 Gráfica de los residuales e_t en función del tiempo, ejemplo 14.3.

http://carlos2524.jimdo.com/

TABLA 14.6 Variables transformadas, regresora y de respuesta, y residuales

t	(1) x'_t	(2) y'_t	(3) e'_t
2	0.553	2.715	0.2504
3	0.601	1.612	0.3176
4	0.505	3.178	−0.4572
5	0.608	1.033	−0.1070
6	0.499	3.688	−0.0908
7	0.522	2.908	−0.3187
8	0.496	3.286	−0.5704
9	0.498	4.016	0.2153
10	0.472	4.672	0.2419
11	0.455	4.305	−0.5559
12	0.507	3.125	−0.4668
13	0.471	4.309	−0.1655
14	0.439	5.869	0.6212
15	0.445	4.892	−0.2010
16	0.489	4.842	0.8200
17	0.503	3.789	0.0985
18	0.451	4.963	0.0029
19	0.437	5.219	−0.0729
20	0.469	4.774	0.2660

para $t = 2, 3, \ldots, 20$. Estas variables transformadas se ven en las columnas 1 y 2 de la tabla 14.6. Un ajuste por mínimos cuadrados a las variables transformadas produce

$$y'_t = 15.85043 - 24.19991x'_t$$

Los residuales de este modelo se encuentran en la columna 3 de la tabla 14.6. Otros estadísticos de resumen para el modelo están en la tabla 14.7.

El estadístico Durbin-Watson para el modelo transformado es $d = 1.94$. Al compararlo con los valores críticos de 5% para $n = 19$, $d_L = 1.18$ y $d_U = 1.40$, se llega a la conclusión que los errores en el modelo transformado no están correlacionados; por lo anterior, el procedimiento Cochrane-Orcutt eliminó el problema original de autocorrelación.

Nótese que β'_1, en el modelo transformado, es igual a β_1 en términos de las variables originales. Por consiguiente, al comparar las tablas 14.5 y 14.7 se ve que con el procedimiento Cochrane-Orcutt se ha obtenido un estimado de la pendiente que sólo difiere un poco al determinado por mínimos cuadrados ordinarios, sin embargo; cuando se comparan

TABLA 14.7 Estadísticos de resumen para el ajuste de modelo a las variables transformadas en el ejemplo 14.3.

Coeficiente	Estimado	Error estándar	t_0
β'_0	15.85043	0.9471	16.74
β'_1	−24.19991	1.9015	−12.73
	$MS_{Res} = 0.1547$	$R^2 = 0.91$	$F_0 = 161.96$

http://carlos2524.jimdo.com/

los errores estándar, se ve que ese estimado de pendiente, con el método iterativo de Cochrane-Orcutt, tuvo un error estándar mayor que con el estimado por mínimos cuadrados ordinarios. Esto recalca nuestro comentario anterior, que si se usan mínimos cuadrados ordinarios cuando los errores están positivamente autocorrelacionados, es probable que los errores estándar de los coeficientes de regresión estén subestimados. En términos de las variables originales, la ordenada al origen y su error estándar son

$$\hat{\beta}_0 = \frac{\hat{\beta}_0}{1-\hat{\rho}} = \frac{15.85043}{1-0.409} = 26.81968$$

y

$$\text{se}\left(\hat{\beta}_0\right) = \frac{\text{se}\left(\hat{\beta}_0'\right)}{1-\hat{\rho}} = \frac{0.9471}{1-0.409} = 1.6025$$

Al comparar estos valores con los de la tabla 14.5 se ve que el estimado iterativo de la ordenada al origen no difiere mucho del obtenido por mínimos cuadrados, pero el primero tiene un error estándar mayor.

14.2 EFECTO DE LOS ERRORES DE MEDICIÓN SOBRE LOS REGRESORES

En casi todos los modelos de regresión se supone que la variable de respuesta y está sujeta al término de error ε, y que las variables regresoras $x_1, x_2, \ldots, x_k$ son **variables deterministas**, o **variables matemáticas**, no afectadas por errores. Hay dos variantes de este caso: la primera es el caso en que la respuesta y los regresores son **variables aleatorias con distribución conjunta**; esta hipótesis da lugar al **modelo de correlación** descrito en el capítulo 2 (véase la Sec. 2.10). El segundo caso es cuando hay **errores de medición** en la respuesta y en los regresores. Ahora bien, si los errores de medición sólo están en la variable de respuesta y, no se presentan nuevos problemas, siempre y cuando esos errores sean no correlacionados, y no tengan sesgo (es decir tengan valor esperado cero), sin embargo, el caso es distinto cuando hay errores de medición en las x. En esta sección se examinará este problema.

14.2.1 Regresión lineal simple

Supóngase que se quiere ajustar el modelo de regresión lineal simple, pero que se mide el regresor con errores, por lo que el regresor observado es

$$X_i = x_i + a_i, \qquad i = 1, 2, \ldots, n$$

donde x_i es el valor verdadero del regresor, X_i es su valor observado y a_i es el error de medición con $E(a_i) = 0$ y $\text{Var}(a_i) = \sigma_a^2$. La variable de respuesta y_i está sujeta al error acostumbrado, ε_i, $i = 1, 2, \ldots, n$, por lo que el modelo de regresión es

$$y_i = \beta_0 + \beta_i x_i + \varepsilon_i \tag{14.9}$$

http://carlos2524.jimdo.com/

Se supondrá que los errores ε_i y a_i no están correlacionados, ya que $E(\varepsilon_i a_i) = 0$. A esto se le llama a veces **errores en el modelo de ambas variables**. Como X_i es el valor observado del regresor, se puede escribir

$$\begin{aligned} y_i &= \beta_0 + \beta_1(X_i - a_i) + \varepsilon_i \\ &= \beta_0 + \beta_1 X_i + (\varepsilon_i - \beta_1 a_i) \end{aligned} \tag{14.10}$$

Al principio, la ecuación (14.10) podrá parecer como un modelo de regresión lineal ordinaria en la que el término del error es $\gamma_i = \varepsilon_i - \beta_1 a_i$; sin embargo, la variable regresora X_i es aleatoria, y se correlaciona con el término de error $\gamma_i = \varepsilon_i - \beta_1 a_i$. Se ve con facilidad la correlación entre X_i y γ_i, porque

$$\begin{aligned} \text{Cov}(X_i, \gamma_i) &= E\{[X_i - E(X_i)][\gamma_i - E(\gamma_i)]\} \\ &= E[(X_i - x_i)\gamma_i] \\ &= E[(X_i - x_i)(\varepsilon_i - \beta_1 a_i)] \\ &= E(a_i\varepsilon_i - \beta_1 a_i^2) \\ &= -\beta_1\sigma_a^2 \end{aligned}$$

Así, si $\beta_1 \neq 0$, el regresor X_i observado y el término de error γ_i están correlacionados.

La suposición acostumbrada cuando el regresor es una variable aleatoria, es que ese regresor y el componente del error son independientes. La violación de esta suposición introduce varias complicaciones en el problema, por ejemplo, si se aplican los métodos usuales de mínimos cuadrados a los datos (es decir, ignorar el error de medición), los estimadores de los parámetros del modelo ya no son insesgados. De hecho, se puede demostrar que si $\text{Cov}(x_i, \alpha_i) = 0$, entonces

$$E(\hat{\beta}_1) = \frac{\beta_1}{1 + \theta}$$

en donde

$$\theta = \frac{\sigma_a^2}{\sigma_x^2}$$

y

$$\sigma_x^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$$

Esto es, $\hat{\beta}_1$ siempre es un estimador sesgado de β_1, a menos que $\sigma_a^2 = 0$, lo cual sólo sucede cuando no hay errores de medición en las x_i.

Ya que el error de medición está presente hasta cierto grado en casi todos los casos prácticos de regresión, sería bueno contar con algún modo de manejar este problema. Nótese que si σ_a^2 es pequeño en relación con σ_x^2, el sesgo de $\hat{\beta}_1$ será pequeño. Eso implica que si la variabilidad de los errores de medición es pequeña en relación con la variabilidad de las x, se pueden ignorar los errores de medición, y aplicar mínimos cuadrados estándar.

http://carlos2524.jimdo.com/

Se han propuesto varios **métodos alternativos de estimación** para manejar el problema de los errores de medición de las variables. A veces, esas técnicas se describen bajo los temas de **relaciones estructurales** o **funcionales** en la regresión. Los economistas han usado una técnica llamada **mínimos cuadrados en dos etapas**, para estos casos. Con frecuencia esos métodos requieren suposiciones adicionales o información más detallada acerca de los parámetros de la distribución de los errores de medición. Las descripciones de esos métodos se encuentran en Graybill [1961], Johnston [1972], Sprent [1969] y Wonnacott y Wonnacott [1970]. Entre otras referencias útiles están las de Davies y Hutton [1975], Dolby [1976], Halperin [1961], Hodges y Moore [1972], Lindley [1974], Mandansky [1959] y Sprent y Dolby [1980], también se encuentran excelentes discusiones del tema en Draper y Smith [1998] y en Seber [1977].

14.2.2 El modelo de Berkson

Berkson [1950] investigó un caso que implicaba errores de medición en x_i, en el que se puede aplicar en forma directa el método de mínimos cuadrados. Su técnica consiste en ajustar el valor del regresor X_i a un **valor objetivo**. Esto obliga a que X_i se considere como fijo, mientras que el valor verdadero del regresor $x_i = X_i - a_i$ se vuelve una variable aleatoria. Como ejemplo de una situación donde se puede usar este método, supóngase que la corriente que pasa por un circuito eléctrico es una variable regresora. El flujo de corriente se mide con un amperímetro, que no es totalmente preciso, por lo que se presentan errores de medición; sin embargo, si se ajusta el flujo observado de corriente a valores objetivo de 100, 125, 150 y 175 A (por ejemplo), el **flujo observado de corriente** se puede considerar como **fijo**, y el **flujo real de corriente** se vuelve una **variable aleatoria**. Esta clase de problemas se presenta con frecuencia en ingeniería y en las ciencias físicas. El regresor es una variable como temperatura, presión o flujo, y hay error en el instrumento de medición que se usa para observar la variable. A este método también a veces se le llama **modelo de la variable independiente controlada**.

Si se considera que X_i es fija en un valor preasignado y deseado, entonces todavía es adecuada la ecuación (14.10), deducida con la relación $X_i = x_i + a_i$. Sin embargo, el término de error en este modelo, $\gamma_i = \varepsilon_i - \beta_1 a_i$ es ahora independiente de X_i, porque se considera que X_i es una variable fija, no estocástica. De este modo, los errores no están correlacionados con el regresor, y se satisfacen las hipótesis normales de los mínimos cuadrados. En consecuencia, en este caso es adecuado un análisis normal por mínimos cuadrados.

14.3 ESTIMACIÓN INVERSA - EL PROBLEMA DE LA CALIBRACIÓN

La mayor parte de los problemas que implican predicción o estimación requieren determinar el valor de y correspondiente a una x dada, por ejemplo, x_0. En esta sección se examinará el **problema inverso**; esto es, dado que se ha observado un valor de y, como y_0, determinar el valor de x que le corresponde. Por ejemplo, cuando se trata de calibrar un termopar y se sabe que la indicación que produce es una función lineal de la temperatura real, digamos

$$\text{temperatura observada} = \beta_0 + \beta_1(\text{temperatura real}) + \varepsilon$$

es decir

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{14.11}$$

http://carlos2524.jimdo.com/

Ahora supóngase que se mide una temperatura desconocida con el termopar, y que la indicación es y_0. Se desea estimar la temperatura real que indica que la temperatura x_0 corresponde a la lectura observada de temperatura y_0. Este caso se presenta con frecuencia en las ingenierías y las ciencias físicas, y a veces se le llama **problema de calibración**, también se presenta en las pruebas biológicas, donde se construye una curva estándar contra la cual se evalúan todas las pruebas o **discriminaciones** futuras.

Supóngase que el termopar se ha sometido a un conjunto de temperaturas controladas y conocidas, $x_1, x_2, \ldots, x_n$, y que se obtuvo un conjunto de temperaturas correspondientes $y_1, y_2, \ldots, y_n$. Un método para estimar x dada y sería ajustar el modelo (14.11), dando como resultado

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{14.12}$$

Ahora sea y_0 el valor observado de y. Un estimado natural de punto del valor correspondiente de x es

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} \tag{14.13}$$

suponiendo que $\hat{\beta}_1 \neq 0$. A este método se le llama con frecuencia **estimador clásico**. Se puede demostrar (véase Graybill [1976], Seber[1977]) que una región de confianza de $100(1-\alpha)$ por ciento para x_0 es el conjunto de los valores de x que satisfacen la desigualdad

$$\frac{\left(y_0 - \hat{\beta}_0 - \hat{\beta}_1 x\right)^2}{\hat{\sigma}^2 A^2} \leq t^2_{\alpha/2, n-2} \tag{14.14}$$

en la que

$$A^2 = 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Desafortunadamente, el conjunto de x que satisface la ecuación (14.14) puede producir un intervalo finito, dos intervalos semiinfinitos, o toda la recta real. Sólo en el primer caso se obtiene un intervalo de confianza útil, y únicamente se presentará cuando el estadístico F (o en forma equivalente, el estadístico t) para la hipótesis H_0: $\beta_1 = 0$ implique rechazo. Esto es, cuando $\hat{\beta}_1$ esté muy cerca de cero (cuando la línea de regresión es muy horizontal), el intervalo de confianza que resulta no tiene sentido. Por lo anterior, se recomienda al analista que encara una estimación inversa o un problema de calibración, que ajuste el modelo (14.11) y pruebe H_0: $\beta_1 = 0$. Si no se rechaza esta hipótesis, quiere decir que y y x no están relacionadas linealmente, y no tiene sentido un estimado puntual o un intervalo de confianza para x; sin embargo, si se rechaza H_0: $\beta_1 = 0$, con el estadístico F con nivel de significancia α, el estimado puntual de x_0 se calcula con la ecuación (14.13) y el intervalo de confianza de $100\,(1-\alpha)$ por ciento para x_0 es

$$\bar{x} + d_1 \leq x_0 \leq \bar{x} + d_2 \tag{14.15}$$

http://carlos2524.jimdo.com/

en donde d_1 y d_2 son las raíces de

$$d^2\left[\hat{\beta}_1^2 - \frac{t_{\alpha/2,n-2}^2\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right] - 2d\hat{\beta}_1(y_0-\bar{y})$$
$$+\left[(y_0-\bar{y})^2 - t_{\alpha/2,n-2}^2\hat{\sigma}^2\left(1+\frac{1}{n}\right)\right] = 0 \qquad (14.16)$$

Lieberman, Miller y Hamilton [1967] y Miller [1966] presentan otros métodos para estimar intervalos de confianza.

Ejemplo 4.4 *Calibración de un termopar*

Un ingeniero calibra un termopar. Ha escogido 16 valores de temperatura, espaciados uniformemente en el intervalo de 100 a 400 °C. La temperatura real x (medida por un termómetro de precisión conocida), y la indicación observada y del termopar, se muestran en la tabla 14.8, y en la figura 14.4 se ve un diagrama de dispersión. La revisión del diagrama de dispersión indica que la temperatura observada en el termopar se relaciona linealmente con la temperatura real. El modelo rectilíneo es

$$\hat{y} = -6.67 + 0.953x$$

siendo $\sigma^2 = MS_{\text{Res}} = 5.86$. El estadístico F para este modelo es mayor que 20 000, por lo que se rechaza H_0: $\beta_1 = 0$, y se llega a la conclusión que la pendiente de la línea de

TABLA 14.8 Temperatura real y observada

Observación i	Temperatura real x_i (°C)	Temperatura observada y_i (°C)
1	100	88.8
2	120	108.7
3	140	129.8
4	160	146.2
5	180	161.6
6	200	179.9
7	220	202.4
8	240	224.5
9	260	245.1
10	280	257.7
11	300	277.0
12	320	298.1
13	340	318.8
14	360	334.6
15	380	355.2
16	400	377.0

http://carlos2524.jimdo.com/

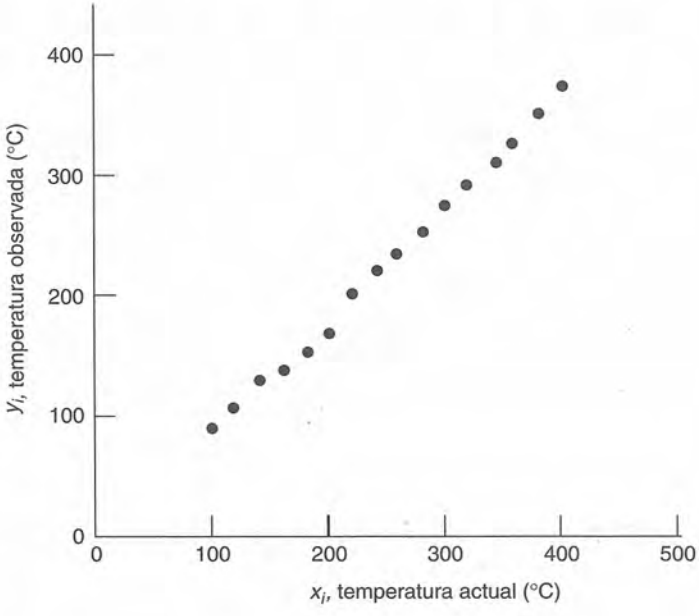


Figura 14.4 Diagrama de dispersión de las temperaturas observadas y reales, ejemplo 14.4.

calibración no es cero. El análisis de residuales no indica algún comportamiento anormal, por lo que se puede usar este modelo para obtener estimados puntuales y de intervalo para la temperatura real, a partir de las indicaciones de temperatura en el termopar.

Supóngase que se obtiene una nueva observación de la temperatura, $y_0 = 200$ °C, con el termopar. Un estimado puntual de la temperatura real, de acuerdo con la línea de calibración, es

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{200 - (-6.67)}{0.953} = 216.86°\text{C}$$

Un intervalo de confianza de 95% para la temperatura real x_0, correspondiente a $y_0 = 200$ °C, se calcula resolviendo la ecuación (14.16):

$$d^2\left[\hat{\beta}_1^2 - \frac{t^2_{\alpha/2,n-2}\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] - 2d\hat{\beta}_1(y_0 - \bar{y})$$

$$+\left[(y_0 - \bar{y})^2 - t^2_{\alpha/2,n-2}\hat{\sigma}^2\left(1 + \frac{1}{n}\right)\right] = 0$$

$$d^2\left[(0.953)^2 - \frac{(2.145)^2(5.86)}{136\,000}\right] - 2d(0.953)(200 - 231.55)$$

$$+\left[(200 - 231.55)^2 - (2.145)^2(5.86)\left(1 + \frac{1}{16}\right)\right] = 0$$

http://carlos2524.jimdo.com/

que se reduce a

$$0.908d^2 + 60.134d + 966.755 = 0$$

Las raíces de esta ecuación cuadrática son $d_1 = -38.752$ y $d_2 = -27.475$. Por consiguiente, de acuerdo con la ecuación (14.15), el intervalo de confianza de 95% para x_0 es

$$\bar{x} + d_1 \leq x_0 \leq \bar{x} + d_2$$

$$250 - 38.752 \leq x_0 \leq 250 - 27.475$$

o sea

$$211.248 \leq x_0 \leq 222.525$$

Otros métodos

El procedimiento clásico descrito en el ejemplo 14.4 no es completamente satisfactorio. Por ejemplo, $\hat{x}_0$ es un estimador sesgado de x_0, y tiene un error cuadrático medio infinito (aunque Williams [1969] demuestra que ningún estimador insesgado de x_0 tiene varianza finita). Hoadley [1970] hace notar que el ancho del intervalo de confianza de x_0 depende de la magnitud del estadístico F; una F grande produce un intervalo corto, y una F pequeña produce uno ancho.

Un método alterno para estimar x sería ajustar el modelo

$$x = \alpha_0 + \alpha_1 y + \varepsilon \tag{14.17}$$

Con esto, se puede hacer la regresión directa de x respecto a y para obtener una ecuación de calibración o de discriminación; a este método con frecuencia se le llama **estimador inverso**; sin embargo, nótese que se supone que x es una variable aleatoria, en la mayor parte de los casos, la variable x es controlada, y **no es** una variable aleatoria, en consecuencia parece inadecuado usar la ecuación (14.17), porque se violan las suposiciones sobre las que se basa el modelo de regresión lineal, lo que nos indica que si x es una variable aleatoria observable, este método sería muy razonable. Desafortunadamente, en muchos problemas prácticos esta premisa no es realista.

A pesar de esas observaciones, Krutchkoff [1967, 1969] ha aconsejado usar el estimador inverso (14.17). Analizó el funcionamiento de ese estimador mediante simulación Monte Carlo, y llegó a la conclusión de que es mejor que el método clásico; no obstante, Berkson [1969], Halperin [1970] y Williams [1969] criticaron el procedimiento inverso con varios argumentos. Hacen notar que no es razonable comparar un estimador con error cuadrático medio infinito (el estimador clásico) con uno con error cuadrático medio finito (el estimador inverso), además, el método inverso sólo es mejor al clásico cuando x_0 está en un intervalo pequeño en torno a $\bar{x}$. Frecuentemente, ese intervalo es muy pequeño. Shukla [1972] demuestra que si es pequeña la cantidad de observaciones que se usan para trazar la línea de calibración, el procedimiento inverso producirá un error cuadrático medio menor para el estimador de la incógnita x que el método clásico; en especial, en casos en los que se requiere extrapolación (hacemos notar que la extrapolación puede ser en extremo peligrosa, y que en general no se recomienda); sin embargo, cuando se usan muestras grandes para trazar la línea de calibración, se prefiere el método clásico. En la práctica puede haber poca diferencia entre los estimadores clásico e inverso, cuando los datos se aproximan mucho a una recta.

http://carlos2524.jimdo.com/

Se han propuesto varios otros estimadores. Graybill [1961, 1976] considera el caso en el que se tienen observaciones repetidas de y en el valor desconocido de x. Desarrolla estimados puntuales y de intervalo para x con el método clásico. La probabilidad de obtener un intervalo de confianza finito para la incógnita x es mayor cuando éstas son observaciones repetidas de y. Hoadley [1970] presenta un tratamiento bayesiano del problema, y deduce un estimador que es un término medio entre los métodos clásico e inverso, también hace notar que el estimador inverso es el estimador de Bayes para la elección particular de la distribución *a priori*. Kalotay [1971], Naszódi [1978], Perng y Tong [1974] y Tucker [1980] propusieron otros estimadores, también tiene interés la publicación de Scheff [1973].

En muchos, si no es que en la mayor parte de los estudios de calibración, el analista puede **diseñar el experimento de recolección de datos**, ya que puede especificar cuáles valores de x se deben observar. Ott y Myers [1968] consideraron la elección de un diseño adecuado para el problema de la estimación inversa, suponiendo que se estima la incógnita x por el método clásico; desarrollan diseños que son óptimos, en el sentido de minimizar el error cuadrático medio integrado; adjuntan figuras para ayudar al analista a seleccionar el diseño.

14.4 BOOTSTRAP EN REGRESIÓN

Cuando se satisfacen las suposiciones del modelo estándar de regresión lineal, existen procedimientos para examinar la precisión de los coeficientes estimados de regresión, así como la precisión del estimado de la media, o la predicción de una observación futura en cualquier punto de interés, esos procedimientos son los acostumbrados errores estándar, los intervalos de confianza y los intervalos de predicción, que se describieron en los capítulos anteriores; sin embargo, hay muchos casos de ajuste de modelos de regresión en los que no hay procedimiento estándar disponible, o donde los resultados disponibles sólo son técnicas aproximadas, porque se basan en la teoría asintótica de muestras grandes. Por ejemplo, para la regresión ridge y para muchos procedimientos de ajuste robusto no se dispone de una teoría para establecer intervalos de confianza o pruebas estadísticas, mientras que en la regresión no lineal y también en los modelos lineales generalizados, las únicas pruebas e intervalos disponibles son los resultados con muestras grandes.

Un procedimiento de cómputo intensivo es el **bootstrap** o **bootstrapping** (su traducción literal al español sería "agujetas" o "amarrarse las agujetas", pero se refiere a indicar algo parecido a una "inicialización" o "tanteo inicial") para permitir la determinación de estimados fiables de los errores estándar en los casos como el que se acaba de escribir. El método de bootstrap lo desarrolló primero Efron [1979, 1982], otras referencias importantes y útiles son la de Davison y Hinkley [1997], Efron [1987], Efron y Tibshirani [1986, 1993] y Wu [1986]. Se explicará e ilustrará el bootsrap en el contexto de determinar el error estándar de un coeficiente de regresión estimado. El mismo procedimiento se aplicaría para obtener errores estándar del estimado de la respuesta media o una observación futura de la respuesta en determinado punto. Después se mostrará cómo obtener por bootstrapping intervalos de confianza aproximados.

Supóngase que se ha ajustado un modelo de regresión, y que interesa determinado coeficiente de la regresión, por ejemplo $\hat{\beta}$. Se desea estimar la precisión de este estimado con el método bootstrap. Ahora bien, este modelo de regresión se ajustó con una muestra de n observaciones. En el método bootstrap se requiere seleccionar una muestra aleatoria de tamaño n, con reemplazo, de esta muestra original, a esto se le llama la **muestra bootstrap**. Como se selecciona con reemplazo, la muestra bootstrap contendrá observaciones de la

http://carlos2524.jimdo.com/

muestra original, algunas de ellas duplicadas y otras omitidas. A continuación se ajusta el modelo a esta muestra bootstrap, con el mismo procedimiento de regresión que para la muestra original. Esto produce un primer estimado bootstrap, por ejemplo $\hat{\beta}_{ii}^*$. Este proceso se repite una gran cantidad de veces. En cada repetición se selecciona una muestra bootstrap, se ajusta el modelo y se obtiene un estimado $\hat{\beta}_{ii}^*$, para $i = 1, 2, \ldots, m$ muestras bootstrap. Como las muestras repetidas se tomaron de la original, al bootstrap también se le llama **procedimiento de remuestreo**. Se representa por $s(\hat{\beta}_{ii}^*)$ la desviación estándar estimada de los m estimados $\hat{\beta}_{ii}^*$ de bootstrap. Esta **desviación estándar bootstrap** $s(\hat{\beta}_{ii}^*)$ es un estimado de la desviación estándar de la distribución de muestreo de $\hat{\beta}$ y, en consecuencia, es una medida de la precisión de la estimación, para el coeficiente β de regresión.

14.4.1 Muestreo bootstrap en regresión

Se describirá cómo aplicar el muestreo bootstrap a un modelo de regresión. Por comodidad, se presentarán los procedimientos en término de un modelo de regresión lineal, pero se podrían aplicar a uno no lineal, o a un modelo lineal generalizado, esencialmente de la misma manera.

Hay dos métodos básicos de bootstrap para estimados de regresión. En el primero, se ajusta el modelo de regresión lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ y se obtienen los n residuales $\mathbf{e}' = [e_1, e_2, \ldots, e_n]$. Se elige una muestra aleatoria de tamaño n con reemplazo, en estos residuales, y se ordenan en un **vector de residuales de bootstrap** $\mathbf{e}^*$. Se adjuntan los residuales bootstrap a los valores predichos $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$ para formar un vector bootstrap de respuestas $\mathbf{y}^*$. Esto es, se calcula

$$\mathbf{y}^* = \mathbf{X}\hat{\beta} + e^* \tag{14.18}$$

A continuación se hace la regresión de las respuestas bootstrap respecto a los regresores originales, con el procedimiento de regresión que se haya usado para ajustar el modelo original. Con esto se obtiene el primer estimado bootstrap del vector de los coeficientes de regresión. Ahora también se podrían obtener estimados bootstrap de cualquier cantidad de interés que sea una **función** de los estimados de parámetro. A este procedimiento se le suele llamar **de residuales bootstrap** o **bootstrap de residuales**.

Otro procedimiento de muestreo bootstrap, llamado generalmente de **casos bootstrap** o de **pares bootstrap**, se usa con frecuencia en casos en los que hay alguna duda sobre la adecuación de la función de regresión que se considera, o cuando la varianza del error no es constante, y/o cuando los regresores no son variables de tipo fijo. En esta variante del muestreo bootstrap, se consideran los n **pares** de muestra $(\mathbf{x}_i, y_i)$ como los datos que se van a remuestrear. Esto es, los n pares originales de la muestra $(\mathbf{x}_i, y_i)$ se muestrean con reemplazo n veces, produciendo una muestra bootstrap $(\mathbf{x}_i^*, y_i^*)$ para $i = 1, 2, \ldots, n$. A continuación se ajusta un modelo de regresión a esta muestra bootstrap, que puede ser

$$\mathbf{y}^* = \mathbf{X}^*\beta + \varepsilon \tag{14.19}$$

dando como resultado el primer estimado bootstrap del vector de los coeficientes de regresión.

Estos procedimientos de muestreo bootstrap se repiten a continuación m veces, generalmente la elección de m depende de la aplicación. A veces se pueden obtener resultados confiables con bootstrap, con una cantidad relativamente pequeña de muestras bootstrap, sin embargo, lo característico es que se empleen de 200 a 100 muestras bootstrap. Una forma de

http://carlos2524.jimdo.com/

seleccionar la m es observar la variabilidad de $s(\hat{\beta}^*)$, la desviación estándar bootstrap, a medida que aumenta m. Cuando se estabilice $s(\hat{\beta}^*)$, se ha llegado al tamaño adecuado de la muestra bootstrap.

14.4.2 Intervalos de confianza bootstrap

Se puede usar la técnica de bootstrap para obtener **intervalos de confianza aproximados** de los coeficientes de regresión y de otras cantidades de interés, como por ejemplo la respuesta media en determinado punto en el espacio x, o un intervalo aproximado de predicción para una observación futura de la respuesta. Como en la sección anterior, se examinarán los coeficientes de regresión, porque la extensión a otras cantidades de regresión es directa.

Un procedimiento sencillo para obtener un intervalo de confianza aproximado de $100(1 - \alpha)$ por ciento mediante bootstrap es el **método de reflexión** [que con cierta frecuencia se llama **método del percentil**]. Este método suele funcionar bien cuando se trabaja con un estimador insesgado. En el método de reflexión para el intervalo de confianza se usan los percentiles inferior de $100(\alpha/2)$ y superior de $100(1 - \alpha/2)$ de la distribución bootstrap de $\hat{\beta}_i^*$. Esos parámetros se representan por $\hat{\beta}^*(\alpha/2)$ y $\hat{\beta}^*(1 - \alpha/2)$, respectivamente. Operacionalmente se obtendrían esos percentiles a partir de la secuencia de los estimados bootstrap que se hayan calculado, $\hat{\beta}_i^*$, $i = 1, 2, \ldots, m$. Se definen como sigue las distancias de esos percentiles a $\hat{\beta}$, el estimado del coeficiente de regresión obtenido de la muestra original:

$$D_1 = \hat{\beta} - \hat{\beta}^*(\alpha/2)$$

$$D_2 = \hat{\beta}^*(1 - \alpha/2) - \hat{\beta} \tag{14.20}$$

Entonces, el **intervalo de confianza bootstrap** aproximado de $100(1 - \alpha/2)$ por ciento para el coeficiente de regresión β es

$$\hat{\beta} - D_2 \leq \beta \leq \hat{\beta} + D_1 \tag{14.21}$$

Antes de presentar ejemplos de este procedimiento, se deben notar dos puntos importantes:

1. Cuando se usa el método de reflexión para establecer intervalos de confianza bootstrap, se aconseja en general usar una mayor cantidad de muestras bootstrap que la que se usaría de ordinario para obtener un error estándar bootstrap. La razón es que se requieren los percentiles pequeños de las colas de la distribución bootstrap, y una muestra grande producirá resultados más confiables. Se recomienda usar cuando menos $m = 500$ muestras bootstrap.
2. La expresión (14.21) del intervalo de confianza relaciona a D_2 con el límite inferior de confianza y a D_1 con el superior, y a primera vista eso parece raro, porque D_1 implica el percentil inferior de la distribución bootstrap, y D_2 el percentil superior. Para visualizar por qué es así, considérese la acostumbrada distribución de muestreo de $\hat{\beta}$, para la cual los percentiles inferior de $100(\alpha/2)$ y superior de $100(1 - \alpha/2)$ por ciento se representan por $\hat{\beta}(\alpha/2)$ y $\hat{\beta}(1 - \alpha/2)$, respectivamente. Ahora se puede afirmar, con una probabilidad de $100(1 - \alpha/2)$ por ciento, que $\hat{\beta}$ quedará dentro del intervalo

$$\hat{\beta}(\alpha/2) \leq \hat{\beta} \leq \hat{\beta}(1 - \alpha/2) \tag{14.22}$$

http://carlos2524.jimdo.com/

Al expresar estos percentiles en función de las distancias a la media de la distribución de muestreo de $\hat{\beta}$, esto es, a $E(\hat{\beta}) = \beta$, se obtiene

$$d_1 = \beta - \hat{\beta}(\alpha/2) \quad \text{y} \quad d_2 = \hat{\beta}(1 - \alpha/2) - \beta$$

Por consiguiente,

$$\begin{aligned} \hat{\beta}(\alpha/2) &= \beta - d_1 \\ \hat{\beta}(1 - \alpha/2) &= \beta + d_2 \end{aligned} \tag{14.23}$$

Al sustituir la ecuación (14.23) en la (14.22) se obtiene

$$\beta - d_1 \leq \hat{\beta} \leq \beta + d_2$$

que se puede escribir como sigue:

$$\begin{aligned} \beta - \beta - \hat{\beta} - d_1 \leq \hat{\beta} - \beta - \hat{\beta} &\leq \beta - \beta - \hat{\beta} + d_2 \\ -\hat{\beta} - d_1 \leq -\beta &\leq -\hat{\beta} + d_2 \\ \hat{\beta} - d_2 \leq \beta &\leq \hat{\beta} + d_1 \end{aligned}$$

esta última ecuación tiene la misma forma que el intervalo de confianza bootstrap, ecuación (14.21), con D_1 y D_2 sustituidos por d_1 y d_2, y usando a $\hat{\beta}$ como un estimado de la media de la distribución de muestreo.

Ahora se presentarán dos ejemplos. En el primero se cuenta con métodos estándar para establecer el intervalo de confianza, y se trata de demostrar que se obtienen resultados semejantes con bootstrap. En el segundo ejemplo interviene la regresión no lineal, y los únicos resultados para el intervalo de confianza que se pueden obtener se basan en la teoría asintótica. Se mostrará cómo se puede usar bootstrap para comprobar la adecuación de los resultados asintóticos.

Ejemplo 14.5 *Los datos del tiempo de entrega*

La versión de estos datos, para regresión múltiple, se presentó por primera vez en el ejemplo 3.1, y se ha usado varias veces en el libro, para ilustrar diversas técnicas de regresión. Se mostrará cómo obtener un intervalo de confianza bootstrap para el coeficiente de regresión del predictor cajas, β_1. De acuerdo con el ejemplo 3.1, el estimado por mínimos cuadrados de β_1 es $\hat{\beta}_1 = 1.61591$. En el ejemplo 3.8 se vio que el error estándar de $\hat{\beta}_1$ es 0.17073, y que el intervalo de 95% de confianza para β_1 es $1.26181 \leq \beta_1 \leq 1.97001$.

Como parece que el modelo se ajusta bien a los datos, y no hay problema de desigualdad de varianza, se hará bootstrap con los residuales para obtener un intervalo bootstrap aproximado de 95% de confianza para β_1. La tabla 3.3 muestra los valores ajustados y los residuales para las 25 observaciones, basándose en el ajuste original por mínimos cuadrados. Para obtener la primera muestra bootstrap, se considera la primera observación. El valor ajustado para esa observación es $\hat{y}_1 = 21.7081$, de acuerdo con la tabla 3.3. Ahora se selecciona al azar un residual de la última columna de esta tabla, por ejemplo $e_5 = -0.4444$.

http://carlos2524.jimdo.com/

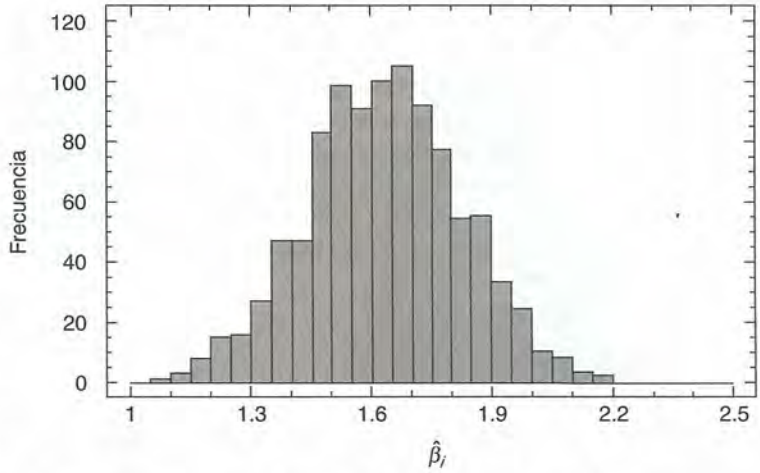


Figura 14.5 Histograma de estimados bootstrap de $\hat{\beta}_1^*$, ejemplo 14.5.

$\bar{\beta}_1^* = 1.63166$, $S(\hat{\beta}_1^*) = 0.18994$, $\hat{\beta}_1^* = (0.025) = 1.24652$, $\hat{\beta}_1^*(0.975) = 1.98970$

Éste se convierte en el primer residual bootstrap, $e_1^* = -0.4444$, entonces, la primera observación bootstrap es $y_1^* = \hat{y}_1 + e_1^* = 21.7081 - 0.4444 = 21.2637$. A continuación se repite este proceso para cada observación siguiente, con los valores ajustados $\hat{y}_i$ y los residuales bootstrap e_i^*, para $i = 2, 3, \ldots, 25$, para formar las observaciones restantes en la muestra bootstrap. Recuérdese que los residuales se muestrean **con reemplazo** de la última columna de la tabla 3.3. Cuando la muestra bootstrap está completa, se ajusta un modelo de regresión lineal a las observaciones (x_{i1}, x_{i2}, y_i^*), $i = 1, 2, \ldots, 25$). El resultado que se obtiene es el primer estimado bootstrap del coeficiente de regresión, $\hat{\beta}_{1,1}^* = 1.64231$. Se repitió este proceso $m = 1\,000$ veces, y se obtuvieron 1 000 estimados bootstrap $\hat{\beta}_{1,u}^*$, $u = 1, 2, \ldots, 1\,000$. La figura 14.5 muestra el histograma de esos estimados bootstrap. Nótese que la forma de este histograma se parece mucho a la distribución normal, esto no es de sorprender, porque la distribución de muestreo de $\hat{\beta}_1^*$ **debería** ser una distribución normal, además, la desviación estándar de los 1 000 estimados bootstrap es $s(\hat{\beta}_1^*) = 0.18994$, razonablemente cercana al error estándar de $\hat{\beta}_1$ acostumbrado, basado en la teoría normal, $\text{se}(\hat{\beta}_1) = 0.17073$.

Para trazar el intervalo aproximado bootstrap de 95% de confianza para $\hat{\beta}_1$ se necesitan los percentiles 2.5 y 97.5 de la distribución de muestreo bootstrap. Esas cantidades son $\hat{\beta}_1^*(0.025) = 1.24652$ y $\beta_1^*(0.975) = 1.98970$, respectivamente (véase la Fig. 14.5). Las distancias D_1 y D_2 se calculan como sigue usando la ecuación (14.20):

$$D_1 = \hat{\beta}_1 - \hat{\beta}_1^*(0.025) = 1.61591 - 1.24652 = 0.36939$$
$$D_2 = \hat{\beta}_1^*(0.975) - \hat{\beta}_1 = 1.98970 - 1.61591 = 0.37379$$

Por último, el intervalo aproximado bootstrap de 95% de confianza se obtiene con la ecuación (14.21):

$$\hat{\beta}_1 - D_2 \le \beta_1 \le \hat{\beta}_1 + D_1$$
$$1.61591 - 0.37379 \le \beta_1 \le 1.61591 + 0.36939$$
$$1.24212 \le \beta_1 \le 1.98530$$

Esto se parece mucho al intervalo de confianza calculado para el ejemplo 3.8: $1.26181 \le \beta_1 \le 1.97001$, calculado con la teoría normal. Cabe esperar que concuerden bastante los

http://carlos2524.jimdo.com/

dos intervalos de confianza, porque en este caso no hay problemas serios con las suposiciones acostumbradas en la regresión.

Las aplicaciones más importantes de bootstrap en la regresión se hacen cuando no hay teoría disponible sobre la cual basar la inferencia estadística, o cuando en los procedimientos se usan resultados para muestras grandes o asintóticos. Por ejemplo, en la regresión no lineal, todas las pruebas estadísticas y los intervalos de confianza son procedimientos con muestras grandes, y sólo se pueden considerar como procedimientos aproximados. En un problema específico, bootstrap se podría usar para examinar la validez de usar esos procedimientos asintóticos.

Ejemplo 14.6 *Datos de puromicina*

En los ejemplos 12.2 y 12.3 se presentaron los datos de puromicina, y se ajustó el modelo de Michaelis-Menten

$$y = \frac{\theta_1 x}{x + \theta_2} + \varepsilon$$

a los datos de la tabla 12.1. Así se obtuvieron los estimados de $\hat{\theta}_1 = 212.7$ y $\hat{\theta}_2 = 0.0641$, respectivamente, también se vio que los errores estándar de muestra grande para esos estimados de parámetro eran $\text{se}(\hat{\theta}_1) = 6.95$, y $\text{se}(\hat{\theta}_2) = 8.28 \times 10^{-3}$, y los intervalos de confianza aproximados de 95% se calcularon en el ejemplo 12.4, resultando

$$197.110 \leq \theta_1 \leq 228.158$$

y

$$0.045 \leq \theta_2 \leq 0.083$$

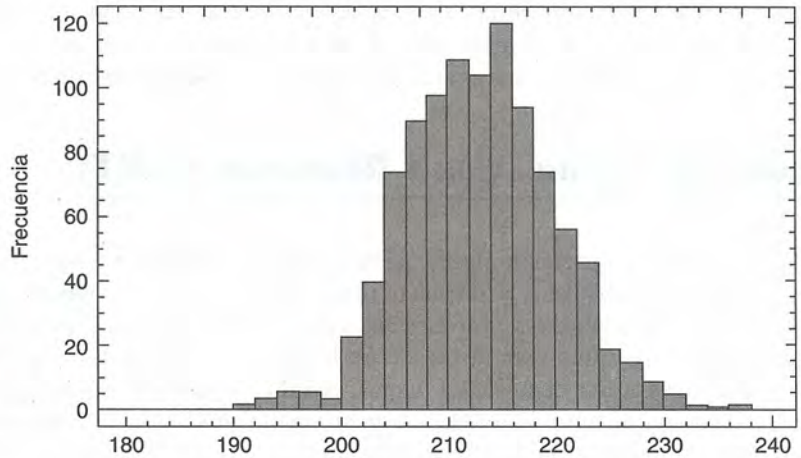
Como los procedimientos de inferencia que se usan aquí se basan en la teoría de muestras grandes, y como el tamaño de muestra usada para ajustar el modelo es relativamente pequeño ($n = 12$), sería bueno comprobar la validez de aplicar los resultados asintóticos calculando desviaciones estándar bootstrap e intervalos de confianza bootstrap para θ_1 y θ_2. Como parece que el modelo de Michaelis-Menten se ajusta bien a los datos, y que no hay problemas importantes con la desigualdad de la varianza, se usó el método de residuales bootstrap para obtener 1 000 muestras bootstrap, cada una de tamaño $n = 12$. En las figuras 14.6 y 14.7 se ven los histogramas que resultaron para los estimados de θ_1 y θ_2, respectivamente. El promedio de muestra, la desviación estándar y los percentiles 2.5 y 97.5 también se indican para cada distribución bootstrap. Obsérvese que los promedios y desviaciones estándar bootstrap se acercan en forma razonable a los valores obtenidos con el ajuste original de mínimos cuadrados no lineales, además, ambos histogramas se parecen razonablemente a la distribución normal, aunque la de $\hat{\theta}_1^*$ podría estar un poco sesgada.

Se pueden calcular los intervalos aproximados de 95% de confianza para θ_1 y θ_2. Primero se verá θ_1. De acuerdo con la ecuación (14.20) y la información de la figura 14.6, se ve que

$$D_1 = \hat{\theta}_1 - \hat{\theta}_1^*(0.025) = 212.7 - 200.386 = 12.314$$

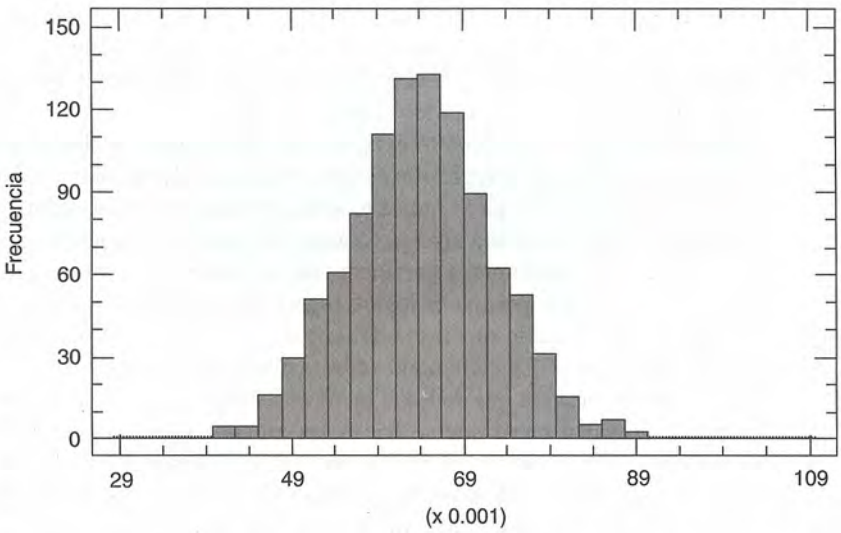
$$D_2 = \hat{\theta}_1^*(0.975) - \hat{\theta}_1 = 226.614 - 212.7 = 13.914$$

http://carlos2524.jimdo.com/



$\bar{\theta}_1^* = 212.88599$, $S(\hat{\theta}_1^*) = 6.87566$, $\hat{\theta}_1^*(0.025) = 200.386$, $\hat{\theta}_1^*(0.975) = 226.614$

Figura 14.6 Histograma de estimados bootstrap de $\hat{\theta}_1^*$, ejemplo 14.6.



$\bar{\theta}_2^* = 0.06388$, $S(\hat{\theta}_2^*) = 0.00845$, $\hat{\theta}_2^*(0.025) = 0.04757$, $\hat{\theta}_2^*(0.975) = 0.08043$

Figura 14.7 Histograma de estimados bootstrap de $\hat{\theta}_2^*$, ejemplo 14.6.

Por consiguiente, el intervalo aproximado de 95% de confianza se calcula como sigue, con la ecuación (14.21):

$$\hat{\theta}_1 - D_2 \leq \theta_1 \leq \hat{\theta}_1 + D_1$$

$$212.7 - 13.914 \leq \theta_1 \leq 212.7 + 12.314$$

$$198.786 \leq \theta_1 \leq 225.014$$

Es muy cercano al intervalo calculado con teoría normal en el problema original. Con un procedimiento parecido se calcula el intervalo aproximado bootstrap de 95% de confianza para θ_2, y el resultado es

$$0.04777 \leq \theta_2 \leq 0.08063$$

http://carlos2524.jimdo.com/

De nuevo, este resultado se parece al intervalo calculado con la teoría normal en el problema original. De este modo se cuenta con algo de seguridad de que se aplican los resultados asintóticos, aun cuando el tamaño de la muestra en este problema sólo es $n = 12$.

14.5 ÁRBOLES DE CLASIFICACIÓN Y REGRESIÓN (CART)

El problema general de clasificación se puede enunciar como sigue: dada una respuesta de interés y ciertos datos taxonómicos (datos de medición o descriptores categóricos) sobre una colección de unidades, usar esos datos para predecir la "clase" a la que pertenece cada unidad. El algoritmo para llevar a cabo lo anterior se puede usar entonces para hacer predicciones acerca de unidades futuras, cuando los datos taxonómicos se conocen pero no la respuesta. Se trata, claro está, de un problema muy general, al que se pueden aplicar muchas herramientas estadísticas diferentes, incluyendo regresión múltiple estándar, regresión logística o los modelos lineales generalizados, análisis de conglomerados, análisis discriminante, etc. En los últimos años, los estadísticos y los científicos de cómputo desarrollaron **algoritmos basados en árboles** para el problema de clasificación. Para conocer más detalles, véanse Breiman, Friedman, Olshen y Stone [1984] y Gunter [1997a, b, 1998].

Cuando la variable de respuesta es discreta, el procedimiento se suele llamar **clasificación**, y cuando es continua, el procedimiento conduce a un **árbol de regresión**. El acrónimo que se usa para indicar los algoritmos que hacen esos procedimientos es **CART** (de *classification and regression trees*) que significa **árboles de clasificación y regresión**. Un árbol de clasificación o regresión es una presentación jerárquica de una serie de preguntas acerca de cada unidad en la muestra. Estas preguntas se relacionan con los valores de los datos taxonómicos de cada unidad. Cuando se contestan esas preguntas, se conoce la "clase" a la que más probablemente pertenece. La presentación acostumbrada para esta información se llama árbol, porque es lógico representar las preguntas como un árbol de cabeza, con una raíz en la parte superior, una serie de ramas que unen o conectan nodos, y las hojas en la parte inferior. En cada nodo se hace una pregunta acerca de una de las variables taxonómicas, y la rama que se tome en el nodo depende de la respuesta. Es importante determinar el orden en el que se hacen las preguntas, porque determina la estructura del árbol. Si bien hay muchas formas de hacerlo, el principio general es hacer la pregunta que maximice la ganancia de **pureza de nodo** en cada oportunidad de ramificación de nodo, y la pureza del nodo mejora al minimizar la variabilidad de los datos de la respuesta en el mismo. Así, si la respuesta es una clasificación discreta, una mayor pureza implicaría menos clases o categorías. Un nodo que contenga una sola clase o categoría de la respuesta sería completamente puro. Si la respuesta es continua, se debe hacer una medida de variabilidad (como una desviación estándar, un error cuadrático medio o una desviación media absoluta de las respuestas en un nodo) tan pequeña como sea posible para maximizar la pureza del nodo.

Para implementar estas ideas muy generales hay muchos algoritmos específicos, y se dispone de muchos programas de cómputo. Las técnicas CART se aplican con frecuencia a conjuntos de datos muy grandes o masivos, por lo que tienden a ser intensivas en cómputo. Hay muchas aplicaciones de técnicas CART, en casos que van desde la interpretación de datos obtenidos con experimentos diseñados, hasta exploración de datos en gran escala (que con frecuencia se llama minería de datos, o descubrimiento de conocimiento en bases de datos).

http://carlos2524.jimdo.com/

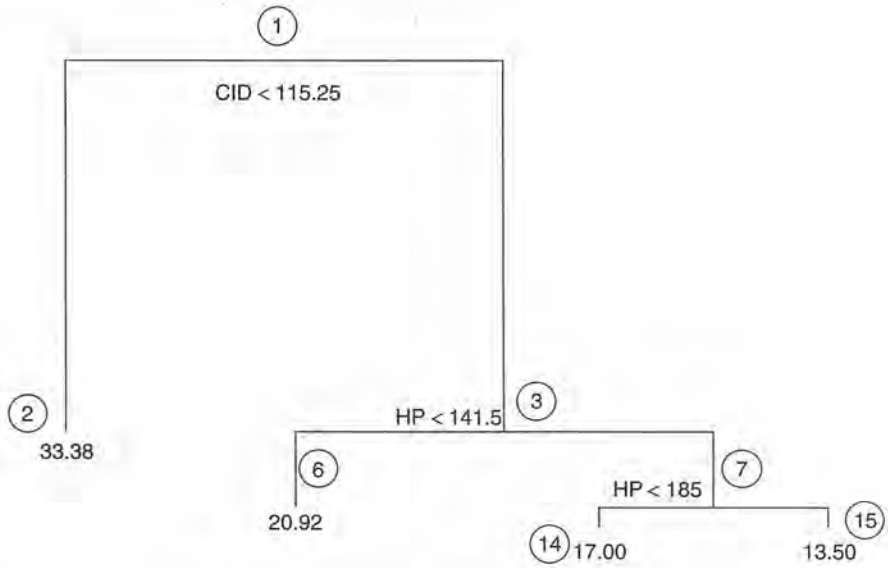


Figura 14.8 Análisis CART con S-PLUS de los datos de rendimiento de gasolina, de la tabla B.3 del apéndice.

```
de), dividido, n, desviación, yval
  * indica nodo terminal

1) root 30 1139.0000 20.04
   2) CID < 115.25 4  22.5500 33.38 *
   3) CID > 115.25  295.600 17.99
      6) HP < 141.5 11  32.4200 20.92 *
      7) HP > 141.5 11  99.0400 15.83
         14) HP < 185 10  50.1600 17.00 *
         15) HP > 185 5  8.1400 13.50 *
```

Ejemplo 14.7 *Datos del rendimiento de la gasolina*

La tabla B.3 del apéndice presenta datos de rendimiento de la gasolina para 32 automóviles, junto con 11 variables taxonómicas. Hay valores faltantes en dos de las observaciones, por lo que se confinará el análisis sólo a los 30 vehículos para los que se dispone de muestras completas. La figura 14.8 presenta un árbol de regresión producido por S-PLUS, aplicado al conjunto de datos. La parte inferior de la figura contiene información descriptiva (también en un formato jerárquico) obtenida con S-PLUS para cada nodo del árbol. La medida de la pureza de nodo, o **desviación** o *deviance* en cada nodo no es más que la suma corregida de cuadrados de las observaciones del mismo, **yval** es el promedio de esas observaciones y n se refiere a la cantidad de observaciones en el nodo.

En el nodo raíz se tienen los 30 coches y en ese lugar la desviación no es más que la suma corregida de cuadrados de los 30. El rendimiento promedio en la muestra es 20.04 mi/gal. La primera rama es sobre la variable CID, las pulgadas cúbicas de cilindrada del motor. Hay cuatro vehículos en el nodo 2 que tienen CID menor que 115.25, su desviación es 22.55, y el rendimiento promedio es 33.38 mi/gal. La desviación en el nodo 3 de la rama derecha en el nodo de la raíz es 295.6, y la suma de las desviaciones de los nodos 2 y 3 es 318.5. No hay otras divisiones posibles en nivel alguno de cualquier variable, para clasificar las observaciones que produzcan una suma de desviaciones menor que 318.5. El nodo 2 es un **nodo terminal**, porque su desviación es un porcentaje de la desviación del nodo

http://carlos2524.jimdo.com/

raíz que es menor que el margen especificado por el usuario, también puede haber nodos terminales si no hay suficientes observaciones (también especificadas por el usuario) para dividir el nodo. Así, en este punto, si se desea identificar los automóviles en el grupo de más alto rendimiento, todo lo que se debe buscar es el desplazamiento del motor.

El nodo 3 contiene 26 coches, y se divide después en el siguiente nodo, por potencia. Once coches con potencia menor que 141.5 forman una rama de este nodo, mientras que la otra la forman 15 coches con potencia mayor que 141.5. La rama de la izquierda da como resultado el nodo terminal 6. La de la derecha entra a otro nodo, el 7, que se vuelve a ramificar por potencia. Esto ilustra una propiedad importante de los árboles de regresión: se puede preguntar lo mismo más de una vez, en distintos nodos del árbol, con lo que se refleja la complejidad de las interrelaciones entre las variables del problema. Los nodos 14 y 15 son terminales, y los automóviles de ambos nodos terminales tienen rendimiento parecido.

El árbol indica que se pueden clasificar los automóviles en alto rendimiento, rendimiento intermedio y bajo rendimiento de gasolina, examinando el desplazamiento y la potencia; son sólo dos de las 11 variables taxonómicas que acompañaban al conjunto de datos originales. Para fines de comparación, la selección de variable hacia adelante, con mi/gal como respuesta, seleccionaría la cilindrada como la única variable importante, y la regresión por etapas o la eliminación hacia atrás seleccionarían la relación de eje trasero, la longitud y el peso; sin embargo, recuérdese que los objetivos de CART y de la regresión múltiple son un poco diferentes: una trata de determinar una estructura óptima o casi óptima de clasificación, mientras que la otra trata de formular una ecuación de predicción.

14.6 REDES NEURALES

Las **redes neurales**, o con más propiedad y las **redes neurales artificiales** se desarrollaron al reconocer que el cerebro humano procesa la información en una forma fundamentalmente distinta de la de una computadora digital típica. La neurona es el elemento estructural y el módulo de procesamiento de información básico del cerebro. Un cerebro humano normal tiene una cantidad enorme de ellas (aproximadamente unos 10 billones de neuronas en la corteza y 60 trillones de sinapsis o conexiones entre ellas), ordenadas en una estructura muy compleja, no lineal y paralela. En consecuencia, el cerebro humano es una estructura muy eficiente para el procesamiento de información, el aprendizaje y el razonamiento.

Una red neural artificial es una estructura diseñada para resolver ciertas clases de problemas, tratando de emular la forma en que lo haría el cerebro humano. La forma general de una red neural es un modelo del tipo "caja negra" que con frecuencia se usa para modelar datos multidimensionales y no lineales. Típicamente, la mayor parte de las redes neurales se usan para resolver problemas de pronóstico de algún sistema, en contraposición a la construcción formal de modelos o al desarrollo de conocimientos básicos de cómo trabaja el sistema. Por ejemplo, una empresa de computadoras podría querer desarrollar un procedimiento para leer en forma automática letra manuscrita, y convertirla a letra de molde. Si el procedimiento lo puede hacer con rapidez y exactitud, podría ser que la empresa tuviera poco interés en el modelo específico que se usa para hacerlo.

Las redes neurales artificiales de varias capas y de avance directo son modelos estadísticos multivariables, usados para relacionar p variables predictoras, $x_1, x_2, \ldots, x_p$, con q variables de respuesta, $y_1, y_2, \ldots, y_q$. El modelo tiene varias **capas**, cada una formada por las variables originales o algunas variables derivadas. En la estructura más común intervienen tres capas: las **entradas**, que son los predictores originales, la **capa oculta**, formada por un conjunto de variables derivadas, y la capa de salida, formada por las respuestas. Cada

http://carlos2524.jimdo.com/

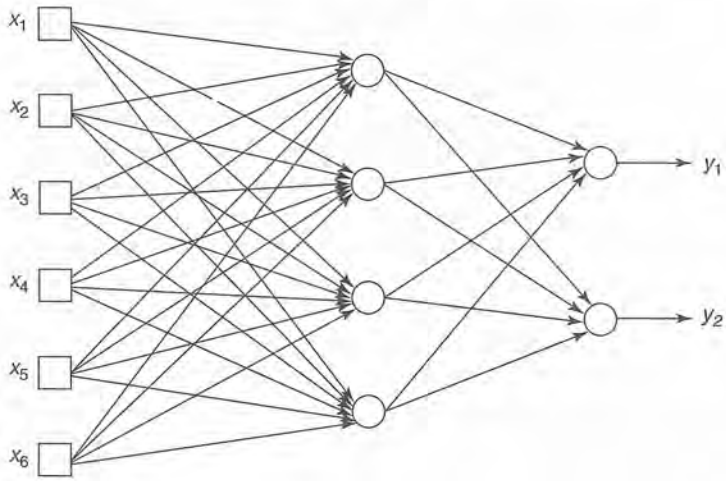


Figura 14.9 Red neural artificial con una capa oculta.

variable en una capa se llama **nodo**. La figura 14.9 muestra una red neural artificial típica, de tres capas.

Un nodo toma como dato o entrada una combinación lineal transformada de las salidas de los nodos en la capa inferior a él. A continuación manda, como salida, una transformación de sí mismo, que se transforma en una de las entradas a uno o más nodos en la capa siguiente. Las funciones de transformación suelen ser sigmoidales (en forma de S) o lineales, y se les llama con frecuencia **funciones de activación** o **funciones de transferencia**. Sea cada uno de los k nodos a_u de la capa oculta una combinación lineal de las variables de entrada:

$$a_u = \sum_{j=1}^{p} w_{1ju} x_j + \theta_u$$

donde las w_{1ju} son parámetros desconocidos que se deben estimar (llamados pesos) y θ_u es un parámetro que juega el papel de ordenada al origen en la regresión lineal (este parámetro se llama, a veces, el nodo de sesgo).

Cada nodo se transforma por la función de activación $g(\ )$. Gran parte de las publicaciones sobre redes neurales representa a esas funciones de activación en la forma $\sigma(\ )$, por su forma sigmoidal (esta opción de notación es desafortunada, en lo que respecta a los estadísticos). Sea $z_u = g(a_u)$ la salida del nodo a_u. Ahora se forma una combinación lineal de esas salidas, por ejemplo $b_v = \Sigma_{u=0}^{k} w_{2uv} z_u$, donde $z_0 = 1$. Por último, la v-ésima respuesta y_v es una transformación de las b_v, por ejemplo $y_v = \tilde{g}(b_v)$, siendo $\tilde{g}(\ )$ la función de activación de la respuesta. Todo lo anterior se puede combinar y resulta

$$y_v = \tilde{g}\left[\sum_{u=1}^{k} w_{2uv} g\left(\sum_{j=1}^{p} w_{1ju} x_j + \theta_{1j}\right) + \theta_{2u}\right] \tag{14.24}$$

La respuesta y_v es una combinación lineal transformada de combinaciones lineales transformadas de los predictores originales. Para la capa oculta, se suele escoger que la función de activación sea la función logística $g(x) = 1/(1 + e^{-x})$ o la función tangente hiperbólica, $g(x) = \tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$. La elección de la función de activación para la capa de salida depende de la naturaleza de la respuesta. Si la respuesta es acotada o dicotómica, la

http://carlos2524.jimdo.com/

función de activación de salida se suele suponer sigmoidal, mientras que si es continua, se usa con frecuencia una función identidad.

El modelo de la ecuación (14.24) es una forma muy flexible, que contiene muchos parámetros, y es esta propiedad la que comunica a una red neural una propiedad casi universal de aproximación. Esto es, se ajusta a muchas funciones que ocurren en la naturaleza; sin embargo, se deben estimar los parámetros de esa ecuación (14.24), y son muchos. El método usual es estimarlos minimizando la suma residual general de cuadrados, calculada con todas las respuestas y todas las observaciones. Se trata de un problema de mínimos cuadrados no lineales, y para resolverlo se pueden usar diversos algoritmos. Con frecuencia se usa un procedimiento llamado **propagación en reversa** (que es una variación del método de descenso más pronunciado), aunque también se han empleado métodos de gradiente, con base en derivadas. Como en cualquier procedimiento no lineal de estimación, se deben especificar los valores iniciales de los parámetros, para usar esos algoritmos. Se acostumbra estandarizar todas las variables de entrada, por lo que se escogen valores pequeños, esencialmente aleatorios, como valores iniciales.

Con tantos parámetros que intervienen en una función no lineal compleja hay un gran riesgo de hacer un **sobreajuste**, lo que nos indica que una red neural permite llegar a un ajuste casi perfecto y un conjunto de datos históricos, o "de adiestramiento", pero con frecuencia pronostica muy mal nuevos datos. El sobreajuste es un problema muy conocido de los estadísticos con experiencia en formación de modelos empíricos. En la comunidad de redes neurales se han desarrollado varios métodos para manejar este problema, como por ejemplo reducir la cantidad de parámetros desconocidos (a lo que se llama "cirugía cerebral óptima"), deteniendo el proceso de estimación de parámetros antes de la convergencia completa, y aplicando validación cruzada para determinar la cantidad de iteraciones que se deben usar, y agregando una función de penalización a la suma de cuadrados de residuales, que aumenta en función de la suma de cuadrados de los estimados de parámetro. También hay muchas estrategias diferentes para escoger la cantidad de capas y la cantidad de neuronas, así como la forma de las funciones de activación, a esto se le llama escoger la **arquitectura de la red**. Se puede usar la validación cruzada para seleccionar la cantidad de nodos en la capa oculta. Unas buenas referencias acerca de las redes neurales artificiales son las de Bishop [1995], Haykin [1994] y Ripley [1994].

Las redes neurales artificiales son un área activa de investigación y aplicaciones, en especial para analizar problemas grandes, complejos y muy no lineales. El asunto del sobreajuste se pasa con frecuencia por alto, por parte de muchos usuarios y partidarios de las redes neurales, y como muchos miembros de esa comunidad no tienen adiestramiento firme en formación empírica de modelos, con frecuencia no aprecian las dificultades que puede causar el sobreajuste, además, muchos programas de cómputo para implementar redes neurales no manejan muy bien el problema de sobreajuste. Nosotros creemos que las redes neurales son un complemento de las herramientas estadísticas usuales del análisis de regresión y del diseño de experimentos, y no un sustituto de ellos, porque una red neural sólo puede presentar un modelo de predicción, y no puede producir una perspectiva fundamental del mecanismo del proceso básico que produjo los datos.

14.7 EXPERIMENTOS DISEÑADOS PARA REGRESIÓN

Como son tantas las propiedades del modelo ajustado de regresión que dependen de los valores de las variables regresoras, en casos en los que se pueden escoger esos valores es natural de considerar el asunto del **diseño de experimentos**, o **diseño experimental**, lo que

http://carlos2524.jimdo.com/

nos lleva a que si se pudieran escoger los valores de las variables regresoras, ¿cuáles se deberían usar? El problema del diseño de experimentos en la regresión consiste en seleccionar los elementos de la matriz **X**. El método general es definir un criterio razonable basado en el modelo que se maneja, para a continuación usarlo para generar un diseño óptimo.

Por ejemplo, supóngase que se ajusta un modelo de regresión lineal simple, $y = \beta_0 + \beta_1 x + \varepsilon$. La varianza del estimado de la pendiente, por mínimos cuadrados, es

$$\operatorname{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Supóngase que el experimento de recolección de datos se está diseñando de tal manera que se deban escoger los valores del regresor x para correr el muestreo. Se desea determinar el conjunto "óptimo" de valores, o diseño óptimo, para este caso, además de que hay un determinado intervalo de valores de x que interesa, sean +1 y −1 los límites superior e inferior de ese intervalo. Ahora, el problema de diseño es escoger n valores para x, como por ejemplo $x_1, x_2, \ldots, x_n$, dentro del intervalo $-1 \leq x \leq 1$, tales que se optimice el criterio de diseño que se esté aplicando.

Supóngase que el criterio de diseño especificado es minimizar la varianza del coeficiente de regresión $\hat{\beta}_1$. Como σ^2 es constante, $\operatorname{Var}(\hat{\beta}_1)$ se minimiza cuando $\Sigma_{i=1}^{n} (x_i - \bar{x})^2$ se maximiza. Es fácil ver que el diseño óptimo es

$$\text{diseño 1 } (n \text{ par}): \quad \frac{n}{2} \text{ corridas en } x = -1, \quad \frac{n}{2} \text{ corridas en } x = 1$$

$$\text{diseño 2 } (n \text{ impar}): \quad \frac{n-1}{2} \text{ corridas en } x = -1, \quad 1 \text{ corrida en } x = 0$$

$$\frac{n-1}{2} \text{ corridas en } x = 1$$

Así, en el diseño óptimo, si se va a hacer una cantidad par de corridas, la mitad de ellas se hacen en cada extremo de la región de interés, mientras que si se debe hacer una cantidad impar de corridas, una de ellas se hace en el centro de la región, y las $n - 1$ corridas restantes se asignan por igual a los extremos. Nótese que con esta estrategia se reparten hacia afuera las x, todo lo lejos posible, de modo que se maximice $\Sigma_{i=1}^{n} (x_1 - \bar{x})^2$.

Aunque este diseño puede ser óptimo con respecto al criterio elegido, no recomendamos usarlo en la práctica. Hablando en general, hay otras consideraciones, además de la varianza de los coeficientes de regresión, que deben tomarse en cuenta en la decisión de diseño. Se desea que la varianza de la pendiente sea pequeña, para que sea correcto el modelo rectilíneo que se haya propuesto; sin embargo, también se desea poder probar la **adecuación** del modelo rectilíneo, lo cual no se puede hacer con mucha eficacia a menos que se usen dos o tres niveles o valores de x, además, puede haber incertidumbre acerca de las regiones de interés para la variable regresora, y para elegir los contornos de esas regiones se necesita tantear mucho; no obstante, el diseño óptimo ubica a todas o a casi todas las corridas en las fronteras de la región. Así, el diseño que es "óptimo" en el contexto de minimizar la varianza de la pendiente, es aquel que requiere que sean absolutamente correctas las hipótesis sobre la estructura del modelo y la región de interés. Desde un punto de vista

http://carlos2524.jimdo.com/

práctico se deben hacer observaciones en varios otros lugares, para dar "seguridad" en caso de que algunas de las hipótesis resulten equivocadas.

Se ha dado mucha atención al problema de diseño óptimo. Se han propuesto varios criterios de optimalidad, y uno de los más frecuentes es maximizar el **determinante** de $\mathbf{X}'\mathbf{X}$. Es un **criterio de estimación de parámetros**, porque si $|\mathbf{X}'\mathbf{X}|$ es máximo, la varianza de los coeficientes individuales de regresión se minimiza, además, también se minimiza el volumen de la **región de confianza conjunta** de los coeficientes de regresión. Un diseño para el cual se maximiza $|\mathbf{X}'\mathbf{X}|$ se llama **diseño óptimo** D, o **diseño** D**-óptimo**.

Otro criterio importante de optimalidad es seleccionar un diseño que minimice la varianza máxima de la respuesta predicha $\hat{y} = \mathbf{x}'\hat{\boldsymbol{\beta}}$ dentro de la región del diseño. Un diseño con esta propiedad se llama **diseño óptimo** G, o **diseño** G**-óptimo**. Es claro que la optimalidad G es un **criterio de estimación de respuesta**. Para tener más información acerca de criterios de optimalidad, como por ejemplo los criterios D y G, véanse Pukelsheim [1995] y Myers y Montgomery [1995].

Desafortunadamente, también surgen muchas de las limitaciones que se vieron en el ejemplo anterior de regresión lineal simple, en los problemas más generales de diseño óptimo. Por ejemplo, los diseños D-óptimos consisten, frecuentemente, en corridas de igual cantidad de puntos como parámetros haya, por lo que no es posible la verificación de la adecuación del modelo; consiguientemente, no hay muchas aplicaciones prácticas de la teoría del diseño óptimo. Un área en la que se ha demostrado la utilidad de la teoría del diseño óptimo es el **aumento** de los datos existentes para ajuste de modelos de regresión, por ejemplo, supóngase que se ajusta un modelo de regresión, y que por lo pronto se dispone de un conjunto de n_1 observaciones. Si se pueden reunir n_2 observaciones adicionales, ¿dónde se deben hacer esas corridas nuevas? Diversos autores (Dykstra [1971] y Gaylor y Merrill [1968]) han descrito métodos para escoger las n_2 corridas adicionales de modo que se maximice el $|\mathbf{X}'\mathbf{X}|$. Este problema se describe con más detalle en Myers y Montgomery [1995]. Otra área en la que la teoría del diseño óptimo ha demostrado su utilidad es el diseño de experimentos para ajustar modelos polinomiales en regiones de forma irregular. Un ejemplo es el problema de diseño de mezclas (véase Cornell [1990], Khuri y Cornell [1996] y Myers y Montgomery [1995]).

Otro problema para el que se han desarrollado diseños útiles es el ajuste de **superficies de respuesta** (véanse Box y Draper [1987], Khuri y Cornell [1996], Montgomery [1997], Myers y Montgomery [1995] y el Cap. 7, Sec. 7.4) en regiones regulares de diseño (cubos y esferas), por lo general polinomios de segundo orden. Una consideración importante en el contexto de la superficie de respuesta es la varianza del valor ajustado $\hat{y}$. Un diseño de superficie de respuesta es **rotable** si la $\mathrm{Var}(\hat{y})$ depende de la **distancia** del punto de interés al centro del diseño, y no depende de la **dirección**; esto implica que las curvas de nivel de $\mathrm{Var}(\hat{y})$ con respecto a $x_1, x_2, \ldots, x_k$ son círculos concéntricos (véase la Sec. 7.4). La razón de que el criterio de rotabilidad sea tan importante es que el objetivo común de un estudio de superficie de respuesta es determinar los valores de $x_1, x_2, \ldots, x_k$ que optimizan la respuesta y, y antes de hacer el experimento el analista no sabe dónde está este punto. Así, todos los puntos equidistantes al centro de diseño tienen igual importancia, independientemente de la dirección. Entre las referencias básicas importantes acerca de diseños de superficie de respuesta están las de Box y Behnken [1960], Box y Draper [1959, 1963] y Box y Hunter [1957]. Para conocer información reciente, véanse Box y Draper [1987], Khuri y Cornell [1996] y Myers y Montgomery [1995].

http://carlos2524.jimdo.com/

PROBLEMAS

14.1 Los datos mostrados abajo muestran la parte del mercado, en porcentaje, de una marca determinada de duraznos en almíbar, durante los últimos 15 meses, y el precio de venta relativo.

t	x_t	y_t
1	100	15.93
2	98	16.26
3	100	15.94
4	89	16.81
5	95	15.67
6	87	16.47
7	93	15.66
8	82	16.94
9	85	16.60
10	83	17.16
11	81	17.77
12	79	18.05
13	90	16.78
14	77	18.17
15	78	17.25

a. Ajustar un modelo de regresión lineal simple a estos datos. Graficar los residuales en función del tiempo. ¿Hay algún indicio de autocorrelación?

b. Aplicar la prueba de Durbin-Watson para determinar si hay autocorrelación positiva de los errores. ¿Cuáles son sus conclusiones?

c. Usar una iteración del procedimiento Cochrane-Orcutt para estimar los coeficientes de regresión. Calcular los errores estándar de esos coeficientes de regresión.

d. ¿Hay autocorrelación positiva todavía, después de la primera iteración? ¿Se podría llegar a la conclusión de que tuvo éxito la estimación iterativa del parámetro?

14.2 Los datos adjuntos muestran las ventas mensuales de un fabricante de cosméticos (y_t) y las ventas mensuales correspondientes de toda la industria (x_t). Las unidades de x_t y y_t son millones de dólares.

a. Ajustar un modelo de regresión lineal simple, que relacione las ventas de la empresa con las ventas de la industria. Graficar los residuales en función del tiempo. ¿Hay indicios de autocorrelación?

b. Usar la prueba de Durbin-Watson para determinar si hay autocorrelación positiva en los errores. ¿Cuáles son sus conclusiones?

c. Usar una iteración del procedimiento Cochrane-Orcutt para estimar los parámetros del modelo. Comparar el error estándar de esos coeficientes de regresión, con el error estándar de los estimados por mínimos cuadrados.

d. Pruebe si hay autocorrelación positiva, después de la primera iteración. ¿Ha tenido éxito el procedimiento iterativo?

http://carlos2524.jimdo.com/

t	x_t	y_t
1	5.00	0.318
2	5.06	0.330
3	5.12	0.356
4	5.10	0.334
5	5.35	0.386
6	5.57	0.455
7	5.61	0.460
8	5.80	0.527
9	6.04	0.598
10	6.16	0.650
11	6.22	0.685
12	6.31	0.713
13	6.38	0.724
14	6.54	0.775
15	6.68	0.782
16	6.73	0.796
17	6.89	0.859
18	6.97	0.883

14.3 Se tiene el modelo de regresión lineal simple $y_t = \beta_0 + \beta_1 x + \varepsilon_t$, en el que se generan los errores con el proceso autorregresivo de segundo orden

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + a_t$$

Describir cómo se podría usar el procedimiento iterativo Cochrane-Orcutt en este caso. ¿Qué transformaciones se usarían con las variables y_t y x_t? ¿Cómo se estimarían los parámetros ρ_1 y ρ_2?

14.4 Para el modelo de regresión con errores autorregresivos de primer orden, definido en la ecuación (14.2), determinar la media, la varianza y la covarianza de los errores de las ecuaciones (14.3a) a (14.3c).

14.5 Usando los datos del problema 14.1, definir un nuevo conjunto de variables transformadas, como la primera diferencia de las variables originales, $x_t' = x_t - x_{t-1}$ y $y_t' = y_t - y_{t-1}$. Hacer la regresión de y_t' con respecto a x_t' que pase por el origen. Comparar el estimado de la pendiente obtenida con este método de primera diferencia, con el obtenido con el método iterativo del problema 14.1.

14.6 Se tiene el modelo de regresión del problema 2.12, que relaciona la presión sanguínea sistólica con el peso. Suponer que se desea predecir el peso de un individuo, dado un valor observado de la presión sanguínea sistólica. ¿Se puede hacer eso con el procedimiento para predecir x dado un valor de y, que se describió en la sección 14.3? En esta aplicación particular, ¿cómo se manejaría la sugerencia de formar un modelo de regresión que relacione el peso con la presión sanguínea sistólica?

14.7 Se tiene el modelo de regresión del problema 2.4, que relaciona el rendimiento de la gasolina con la cilindrada del motor.

a. Si determinado automóvil tiene un rendimiento observado de 17 millas por galón, calcular un estimado puntual para la cilindrada correspondiente del motor.

b. Determinar un intervalo de confianza de 95% para la cilindrada del motor.

http://carlos2524.jimdo.com/

14.8 Se forma un modelo de regresión que relaciona el flujo total de calor con la deflexión radial, para los datos de energía solar de la tabla B.2 del apéndice.

a. Suponer que el flujo total de calor observado es 250 kW. Determinar un estimado puntual para la deflexión radial correspondiente.

b. Determinar un intervalo de confianza de 90% para la deflexión radial.

14.9 Se tienen los datos del tiempo de entrega de bebidas gaseosas del ejemplo 3.1. Determinar un intervalo de confianza bootstrap aproximado para el coeficiente de regresión de la distancia, con $m = 1\,000$ muestras bootstrap. Compararlo con el intervalo de confianza obtenido con la teoría acostumbrada.

14.10 Se tienen los datos del tiempo de entrega de bebidas gaseosas del ejemplo 3.1. Determinar el estimado bootstrap de la desviación estándar de $\hat{\beta}_1$, con las siguientes cantidades de muestras bootstrap: $m = 100$, $m = 200$, $m = 300$, $m = 400$ y $m = 500$. ¿Se pueden sacar algunas conclusiones acerca de cuántas muestras bootstrap se necesitan para obtener un estimado confiable de la precisión del estimado de $\hat{\beta}_1$?

14.11 Describir cómo se obtendría un estimado bootstrap de la desviación estándar del estimado de la respuesta media en determinado punto, por ejemplo, $\mathbf{x}_0$.

14.12 Describir cómo se puede determinar un intervalo de confianza bootstrap aproximado para la respuesta media en determinado punto, por ejemplo, $\mathbf{x}_0$.

14.13 Se tiene el modelo de ajuste con regresión no lineal para los datos del problema 12.11. Determinar los errores estándar bootstrap de los coeficientes de regresión $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_3$, usando $m = 1\,000$ muestras bootstrap. Con base en los resultados obtenidos, comentar acerca de cómo parece aplicarse la teoría asintótica a este problema.

14.14 Para el modelo de ajuste con regresión lineal para los datos del problema 12.11, determinar intervalos aproximados de 95% de confianza bootstrap para los coeficientes de regresión $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_3$, usando $m = 1\,000$ muestras bootstrap. Comparar esos intervalos con los basados en los resultados para muestra grande. Con base en la comparación de esos intervalos, comentar la forma en que la teoría asintótica parece aplicarse a este problema.

14.15 Para los datos de desempeño de equipos de la Liga Nacional de Futbol (NFL) en la tabla B.1 del apéndice, elaborar un árbol de regresión.

14.16 **Diseño de experimento para la regresión lineal.** Se desea ajustar un modelo de regresión lineal simple para la región $-1 \leq x \leq 1$, con $n = 10$ observaciones. Se consideran cuatro diseños de experimento: i) 5 observaciones, en $x = -1$ y 5 observaciones en $x = +1$; ii) 4 observaciones en $x = -1$, 2 observaciones en $x = 0$ y 4 observaciones en $x = +1$; iii) 2 observaciones en $x = -1, -1/2, 0, +1/2$ y $+1$, y iv) 1 observación en $x = -1, -0.8, -0.6, -0.4, -0.2, +0.2, +0.4, +0.6, +0.8$ y $+1$. Para cada uno de esos diseños, determinar la cantidad disponible de grados de libertad para evaluar el error puro y para probar la falta de ajuste, el error estándar de la pendiente (hasta una σ constante) y el valor del determinante de $\mathbf{X}'\mathbf{X}$. Con base en estos análisis, ¿cuál diseño se seleccionaría?

14.17 Un analista ajusta un modelo de regresión simple, con objeto de obtener un estimado de varianza mínima de la ordenada al origen β_0. ¿Cómo se debe diseñar el experimento de obtención de datos?

14.18 Suponer que se trata de ajustar un modelo de regresión lineal simple que se usará para predecir la respuesta media en determinado punto, por ejemplo en x_0. ¿Cómo se debe diseñar el experimento de obtención de datos para obtener un estimado de varianza mínima de la media de y en x_0?

http://carlos2524.jimdo.com/

14.19 Se tiene el modelo de regresión lineal $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, en donde se codificaron los regresores de tal modo que

$$\sum_{i=1}^{n} x_{i1} = \sum_{i=1}^{n} x_{i2} = 0 \quad \text{y} \quad \sum_{i=1}^{n} x_{i1}^2 = \sum_{i=1}^{n} x_{i2}^2 = n$$

a. Demostrar que un diseño ortogonal ($\mathbf{X}'\mathbf{X}$ diagonal) minimiza la varianza de $\hat{\beta}_1$ y $\hat{\beta}_2$.

b. Demostrar que cualquier diseño para ajustar este modelo de primer orden que sea ortogonal, también es rotable.

http://carlos2524.jimdo.com/

CAPÍTULO 15

VALIDACIÓN DE LOS MODELOS DE REGRESIÓN

15.1 INTRODUCCIÓN

Los modelos de regresión se usan en forma extensa para predicción o estimación, descripción de datos, estimación de parámetros y control; con frecuencia, el usuario del modelo de regresión es una persona distinta de quien lo desarrolló. Antes de que el modelo pase al usuario se debe hacer cierta evaluación de su **validez**. Se hará la diferencia entre la **comprobación de adecuación del modelo** y la **validación del modelo**. La comprobación de la adecuación del modelo incluye el análisis de residuales, la prueba de falta de ajuste, la búsqueda de observaciones atípicas o demasiado influyentes y demás análisis internos que investigan el ajuste del modelo de regresión a los datos disponibles; sin embargo, la validación del modelo se concentra en la determinación de si el modelo funcionará bien en su ambiente pretendido de operación.

Como el ajuste del modelo a los datos disponibles forma la base de muchas de las técnicas que se usaron en el proceso de desarrollo del modelo (como por ejemplo, la selección de las variables), es atractiva la conclusión de que un modelo que se ajusta bien a los datos también funcionará bien en la aplicación final; eso no es necesariamente cierto; por ejemplo, un modelo puede haber sido desarrollado principalmente para predecir observaciones nuevas. No hay la seguridad de que la ecuación que proporciona el mejor ajuste a los datos observados sea un buen predictor. Los factores influyentes que se desconocían durante la etapa de formación del modelo pueden afectar en forma importante las nuevas observaciones, haciendo casi inútiles las predicciones, además, la estructura correlativa entre los regresores puede ser distinta en la formación del modelo y en la predicción de datos; ello puede producir mal desempeño predictivo del modelo. La validación adecuada de un modelo desarrollado para pronosticar nuevas observaciones debe implicar la prueba del modelo en su ambiente, antes de ser entregado al usuario.

Otra razón crítica de la validación es que el desarrollador del modelo generalmente tiene poco o ningún control sobre el uso final de éste. Por ejemplo, Snee [1977] observa que aunque se haya desarrollado un modelo como ecuación de interpolación, cuando el usuario descubre que es bueno en este respecto, también lo extrapolará, si se presenta la necesidad, a pesar de cualquier advertencia o nota precautoria del desarrollador, además, si esta extrapolación funciona mal, casi siempre es al **desarrollador** y no al **usuario** a quien se le reclama por la falla. Los usuarios de modelos de regresión también llegarán a conclusiones acerca del proceso que se estudia a través de los signos y las magnitudes de los coeficientes de su modelo, aunque se les hayan descrito los riesgos de interpretar los coeficientes parciales de la regresión. La validación del modelo permite contar con una medida de protección, tanto al desarrollador como al usuario del modelo.

http://carlos2524.jimdo.com/

La validación adecuada de un modelo de regresión debe incluir un **estudio de los coeficientes**, para determinar si sus signos y sus magnitudes son razonables, que nos lleva a preguntarnos ¿se puede interpretar, en forma razonable, que $\hat{\beta}_j$ sea un estimado del efecto de x_j? También se debe investigar la **estabilidad** de los coeficientes de regresión, es decir, las $\hat{\beta}_j$ obtenidas con una nueva muestra ¿probablemente serán parecidas a los coeficientes actuales? Por último, una validación requiere investigar el **desempeño del modelo en predicción**. Se deben tener en cuenta los modos de interpolación y extrapolación.

En este capítulo se describirán e ilustrarán varias de las técnicas útiles para validar modelos de regresión. Algunas referencias acerca del tema general de la validación son las de Brown, Durbin y Evans [1975], Geisser [1975], McCarthy [1976], Snee [1977] y Stone [1974]; se recomienda en especial el trabajo de Snee.

15.2 TÉCNICAS DE VALIDACIÓN

Para validar un modelo de regresión existen tres clases de procedimientos:

1. Análisis de los coeficientes del modelo y los valores predichos, incluyendo comparaciones con la experiencia, la teoría física y otros modelos analíticos o resultados de simulación.
2. Recolección de nuevos datos (o datos frescos) con los cuales investigar el desempeño del modelo en predicción.
3. División de datos, es decir, apartar algunos de los datos originales y usar esas observaciones para investigar el desempeño del modelo en predicción.

El **uso final pretendido** del modelo indica, frecuentemente, cuál es la metodología adecuada de validación, así, la validación de un modelo para usarse como ecuación de predicción se debe concentrar en la determinación de la exactitud del modelo. Sin embargo, como el desarrollador consiguientemente no controla el uso del modelo, se recomienda que, siempre que sea posible, se usen todas las técnicas de validación que se acaban de mencionar. Para conocer algunos ejemplos adicionales, véase Snee [1977].

15.2.1 Análisis de los coeficientes y valores predichos por el modelo

Se deben estudiar los coeficientes del modelo final de regresión, para determinar si son **estables** y si son razonables sus **signos y magnitudes**. La experiencia, las consideraciones teóricas o un modelo analítico pueden proporcionar, frecuentemente, información acerca de la dirección y el tamaño relativo de los efectos de los regresores. Se deben comparar con esta información los coeficientes en el modelo estimado. Los coeficientes que tengan signos inesperados, o que tengan valores absolutos demasiado grandes, suelen indicar ya sea que el modelo es inadecuado (regresores faltantes o mal especificados) o malos estimados de los efectos de los regresores individuales. Los factores de inflación de varianza (y otros diagnósticos de multicolinealidad del Cap. 10) también son una guía importante para la validez del modelo. Si cualquier factor de inflación de varianza es mayor que 5 o 10, el coeficiente correspondiente está mal estimado o es inestable debido a dependencias casi lineales entre los regresores. Cuando los datos se obtienen en función del tiempo, se puede examinar la estabilidad de los coeficientes ajustando el modelo en intervalos de

http://carlos2524.jimdo.com/

tiempo menores; por ejemplo, si se contara con varios años de datos mensuales, se podría formar un modelo para cada año; es de esperar que los coeficientes para cada año sean similares.

Los valores predichos de la respuesta $\hat{y}$ también pueden proporcionar una medida de la validez del modelo. Los valores predichos no realistas, como por ejemplo predicciones negativas de una cantidad positiva, o predicciones que salen del rango real de la respuesta, indican que los coeficientes están mal estimados, o que la forma del modelo es incorrecta. Los valores predichos dentro y en la frontera del dominio de la variable regresora proporcionan una medida de la eficiencia del modelo en interpolación. Los valores predichos que salen de esa región son una medida de la eficiencia en extrapolación.

Ejemplo 15.1 *Datos del acetileno*

Se examinarán los datos del acetileno, presentados en el ejemplo 10.1. Al comprobar la validez del modelo de mínimos cuadrados con nueve términos, para esos datos, estudiando los coeficientes y los valores predichos se revelan varios problemas. Los coeficientes normalizados de regresión, correspondientes a x_1^2, x_3^2 y x_1x_3 tienen valor absoluto grande. No es normal que esos coeficientes, en un modelo cuadrático, sean tan grandes, además, los factores de inflación de varianza para los coeficientes donde intervienen x_1 y x_3 son muy grandes (véase la Tabla 10.5), e indican una alta correlación entre x_1 y x_3. Los valores predichos de la conversión porcentual (Fig. 10.3) también indican que hay problemas potenciales con el modelo de mínimos cuadrados de nueve términos. Los valores predichos dentro de la región de los datos originales son satisfactorios, pero con una extrapolación relativamente pequeña se obtienen varios estimados negativos del porcentaje de conversión, cantidad que debe ser positiva. Hay evidencias de que, aunque el modelo de mínimos cuadrados de nueve términos puede ser una ecuación válida para interpolación, sus extrapolaciones son malas.

En el capítulo 10 se desarrollaron varios modelos alternos para los datos del acetileno, aplicando técnicas de selección de variables y de estimación sesgada, para combatir la multicolinealidad de los datos. Tanto el modelo de nueve variables, de regresión ordinaria de cresta (Ej. 10.2) como el modelo de mínimos cuadrados con cinco variables (Ej. 10.4) son ecuaciones más razonables, porque reducen la magnitud de los coeficientes de x_1^2, x_3^2 y x_1x_3, a valores más adecuados, y producen valores predichos más realistas, tanto en interpolación como en extrapolación.

15.2.2 Recolección de nuevos datos - Corridas de confirmación

El método más eficaz para validar un modelo de regresión con respecto a su eficiencia de predicción es reunir nuevos datos y comparar en forma directa las predicciones del modelo con esos datos. Si el modelo produce predicciones exactas para nuevos datos, el usuario tendrá más confianza tanto en el modelo como en el proceso de su obtención. A veces, a esas nuevas observaciones se les llama **corridas de confirmación**. Es preferible contar con al menos de 15 a 20 observaciones nuevas, para tener una evaluación fiable de la eficiencia de predicción del modelo. En casos en donde se hayan desarrollado, con los mismos datos, dos o más modelos de regresión alternativos, la comparación de la eficiencia de predicción de ellos con nuevos datos puede ser la base de la selección del modelo final.

http://carlos2524.jimdo.com/

TABLA 15.1 Conjunto de datos de predicción para el ejemplo de datos del tiempo de entrega

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					Ajuste por mínimos cuadrados		Ajuste robusto	
Observación	Ciudad	Cajas x_1	Distancia x_2	Tiempo observado, y	$\hat{y}$	$y - \hat{y}$	$\hat{y}$	$y - \hat{y}$
26	San Diego	22	905	51.00	50.9230	0.0770	48.8072	2.1928
27	San Diego	7	520	16.80	21.1405	−4.3405	21.2599	−4.4599
28	Boston	15	290	26.16	30.7557	−4.5957	30.0648	−3.9048
29	Boston	5	500	19.90	17.6207	2.2793	18.0106	1.8894
30	Boston	6	1000	24.00	26.4366	−2.4366	26.2615	−2.2615
31	Boston	6	225	18.55	15.2766	3.2734	15.7812	2.7688
32	Boston	10	775	31.93	29.6602	2.2698	29.1764	2.7536
33	Boston	4	212	16.95	11.8576	5.0924	12.6266	4.3234
34	Austin	1	144	7.00	6.0307	0.9693	7.2388	−0.2388
35	Austin	3	126	14.00	9.0033	4.9967	9.9742	4.0258
36	Austin	12	655	37.03	31.1640	5.8660	30.5325	6.4975
37	Louisville	10	420	18.62	24.5482	−5.9282	24.3758	−5.7558
38	Louisville	7	150	16.10	15.8125	0.2875	16.2563	−0.1563
39	Louisville	8	360	24.38	20.4524	3.9276	20.5856	3.7944
40	Louisville	32	1530	64.75	76.0820	−11.3320	72.1531	−7.4031

Ejemplo 15.2 *Datos del tiempo de entrega*

Se revisarán los datos del tiempo de entrega que se presentaron en el ejemplo 3.1. Antes ya se había desarrollado el ajuste por mínimos cuadrados (Ej. 3.1) y también un ajuste robusto con la función E_a de Ramsay (Ej. 11.5) con esos datos. El objetivo de ajustar el modelo de regresión es predecir nuevas observaciones. Se investigará la validez de esos modelos como predictores, y se compararán los ajustes por mínimos cuadrados y robusto calculando el tiempo de entrega con nuevos datos.

Recuérdese que las 25 observaciones originales provenían de cuatro ciudades: Austin, San Diego, Boston y Minneapolis. En la tabla 15.1 se muestran 15 nuevas observaciones de Austin, Boston, San Diego y una quinta y nueva ciudad: Louisville, también se ven allí los tiempos predichos correspondientes, y los errores de predicción del ajuste con mínimos cuadrados $\hat{y} = 2.3412 + 1.6159x_1 + 0.0144x_2$ (columnas 5 y 6), y del ajuste robusto $\hat{y} = 3.8021 + 1.4894x_1 + 0.0135x_2$ (columnas 7 y 8). Nótese que este conjunto de datos para predicción consiste en 11 observaciones procedentes de ciudades que se estudiaron en el proceso original de recolección de datos, y que cuatro observaciones son de una nueva ciudad. Esta mezcla de ciudades anteriores y nuevas puede aportar alguna información acerca de lo bien que predicen los dos modelos en lugares donde se reunieron los datos originales y en nuevos sitios.

La columna 6 de la tabla 15.1 muestra los errores de predicción para el modelo de mínimos cuadrados. El error promedio de predicción es 0.4060, casi cero, por lo que el modelo parece producir aproximadamente predicciones insesgadas; sólo hay un error de predicción relativamente grande, asociado con la última observación de Louisville. Al revisar los datos originales se ve que esa observación es un punto de extrapolación, además, ese punto se parece bastante al punto 9, que ya se sabe que es influyente. Desde una perspectiva general, esos errores de predicción aumentan la confianza para la utilidad del modelo.

http://carlos2524.jimdo.com/

Nótese que los errores de predicción son, en general, mayores que los residuales del ajuste por mínimos cuadrados. Esto se visualiza comparando el cuadrado medio de residuales

$$MS_{\text{Res}} = 10.6239$$

obtenido con el modelo ajustado, con el error de predicción promedio al cuadrado

$$\frac{\sum_{i=26}^{40} (y_i - \hat{y}_i)^2}{15} = \frac{332.2809}{15} = 22.1521$$

obtenido con los nuevos datos de predicción. Como MS_{Res} (que se puede considerar como la varianza promedio de los residuales del ajuste) es menor que el error promedio de predicción al cuadrado, el modelo de mínimos cuadrados no predice los nuevos datos tan bien como se ajusta a los datos existentes; sin embargo, la degradación de la eficiencia no es grave, por lo que se puede concluir que es probable que el modelo de mínimos cuadrados sea buen predictor. Nótese también que, además del único punto de extrapolación, los errores de predicción para Louisville no son notablemente distintos de los que se obtienen en las ciudades donde se reunieron los datos originales. Si bien la muestra es pequeña, se cuenta con un indicio de que el modelo puede ser "portátil". Sería bueno contar con datos más extensos de otros lugares, para verificar esta conclusión.

También es instructivo comparar R^2 del ajuste con mínimos cuadrados (0.9596) con el porcentaje de variabilidad que explica el modelo con los nuevos datos, es decir

$$R^2_{\text{Predicción}} = 1 - \frac{\sum_{i=26}^{40} (y_i - \hat{y}_i)^2}{\sum_{i=26}^{40} (y_i - \bar{y})^2} = 1 - \frac{332.2809}{3206.2338} = 0.8964$$

De nuevo se ve que el modelo de mínimos cuadrados no predice las observaciones tan bien como se ajusta a los datos originales; no obstante, la "pérdida" en R^2 para predicción es pequeña.

Las columnas 7 y 8 de la tabla 15.1 muestran los valores predichos y los errores de predicción para el ajuste robusto. El error promedio de predicción es 0.2710, por lo que el procedimiento de regresión robusta produce también predicciones aproximadamente insesgadas; sin embargo, parece que los errores de predicción con el ajuste robusto son más pequeños en general que sus contrapartes por mínimos cuadrados. Eso se confirma calculando el error promedio de predicción al cuadrado del ajuste robusto, por ejemplo

$$\frac{\sum_{i=26}^{40} (y_i - \hat{y}_i)^2}{15} = \frac{243.4093}{15} = 16.2273$$

que es menor que la cantidad correspondiente para los errores de predicción por mínimos cuadrados. En forma parecida, la R^2 para predicción con el ajuste robusto es

http://carlos2524.jimdo.com/

$$R^2_{\text{Predicción}} = 1 - \frac{\sum_{i=26}^{40} (y_i - \hat{y}_i)^2}{\sum_{i=26}^{40} (y_i - \bar{y})^2} = 1 - \frac{243.4093}{3\,206.2338} = 0.9241$$

que es un poco mayor que el valor correspondiente del modelo de mínimos cuadrados. Nótese que el ajuste robusto no sobreestima el tiempo de entrega para el punto de extrapolación tan mal como lo hacen los mínimos cuadrados.

Recabando nuevos datos se ha visto que los ajustes por mínimos cuadrados como el robusto, a los datos del tiempo de entrega, producen ecuaciones razonablemente buenas de predicción. Hay cierta evidencia que, en este ejemplo, el ajuste robusto es un poco mejor que los mínimos cuadrados.

15.2.3 División de los datos

En muchos casos no es posible reunir nuevos datos para fines de validación. Puede haberse agotado el presupuesto de recolección, la planta puede haber sido modificada para producir otros artículos, o el equipo y recursos necesarios para reunir datos pueden ser inasequibles. Cuando se presentan estos casos, un procedimiento razonable es dividir los datos disponibles en dos partes, a las que Snee [1977] llama **datos de estimación** y **datos de predicción**. Los datos de estimación se usan para formar el modelo de regresión y a continuación los datos de predicción se usan para estudiar la capacidad predictiva del modelo, a veces, a la división de datos se le llama **validación cruzada** (véanse Mosteller y Tukey [1968] y Stone [1974]).

La división de datos se puede hacer de diversas maneras. Por ejemplo, se puede considerar el estadístico PRESS

$$\text{PRESS} = \sum_{i=1}^{n} \left[y_i - \hat{y}_{(i)} \right]^2 = \sum_{i=1}^{n} \left(\frac{e_i}{1 - h_{ii}} \right)^2 \tag{15.1}$$

como una forma de división de datos. Recuérdese que se puede aplicar PRESS para calcular un estadístico parecido a R^2:

$$R^2_{\text{Predicción}} = 1 - \frac{\text{PRESS}}{SS_{\text{T}}}$$

que mide, en forma aproximada, cuánto cabe esperar que el modelo explique la variabilidad de las nuevas observaciones. Para dar un ejemplo, recuérdese que en el capítulo 4 (Ej. 4.6) se calculó PRESS para el ajuste del modelo a las 25 observaciones originales del tiempo de entrega, y se vio que PRESS = 457.4000. Por consiguiente,

$$\begin{aligned} R^2_{\text{Predicción}} &= 1 - \frac{\text{PRESS}}{SS_{\text{T}}} \\ &= 1 - \frac{457.4000}{5\,784.5426} \\ &= 0.9209 \end{aligned}$$

http://carlos2524.jimdo.com/

Ahora la R^2 para el ajuste por mínimos cuadrados es $R^2 = 0.9596$, por lo que PRESS indicaría que es probable que el modelo sea un muy buen predictor de observaciones nuevas. Nótese que la R^2 para predicción, basada en PRESS, se parece mucho a la eficiencia real de predicción que se observa en este modelo con los nuevos datos, en el ejemplo 15.2.

Si los datos se reunieron en una **secuencia de tiempo**, entonces se puede usar el **tiempo** como base de la división de los datos. Esto es, se identifica determinado periodo, y todas las observaciones reunidas antes de ese periodo se usan para formar el conjunto de datos de estimación, mientras que las obtenidas después de ese tiempo forman el conjunto de datos de predicción. Un procedimiento razonable de validación, para determinar la forma probable en la que funcione un modelo en el futuro es ajustar el modelo a los datos de estimación y examinar su exactitud de predicción con los datos de predicción. Este tipo de procedimiento de validación es de uso relativamente frecuente en el análisis de series de tiempo, para investigar la eficiencia potencial de un modelo de pronóstico (véanse algunos ejemplos en Montgomery, Johnson y Gardiner [1990]). Para conocer ejemplos donde intervienen modelos de regresión, véanse Cady y Allen [1972) y Draper y Smith [1998].

Además del tiempo, con frecuencia se pueden usar otras características de los datos para dividirlos. Por ejemplo, para los datos del tiempo de entrega en el ejemplo 3.1, se puede suponer que se cuenta también con las 15 observaciones adicionales de la tabla 15.1. Como hay cinco ciudades representadas en la muestra, se podrían usar (por ejemplo), las observaciones de San Diego, Boston y Minneapolis como datos de estimación, y las de Austin y Louisville como datos de predicción, así se tendrían 29 observaciones para estimación y 11 para validación. En otras situaciones puede ser que los operadores, lotes de materia prima, unidades de equipo de prueba, laboratorios, etc., se pueden usar para formar los conjuntos de datos de estimación y de predicción. En los casos en los que no hay base lógica para dividir los datos, se pueden asignar al azar las observaciones a los conjuntos de datos de estimación y de predicción.

Una desventaja potencial de esos métodos de división de datos, que son algo **arbitrarios**, es que frecuentemente no hay la seguridad de que el conjunto de datos de predicción "pruebe" la severidad del modelo lo suficiente. Por ejemplo, una **división aleatoria** de los datos no necesariamente asegura que algunos puntos del conjunto de datos de predicción sean de extrapolación, y el esfuerzo de validación no proporcionaría información sobre lo bien que probablemente se extrapole en el modelo. En ausencia de una base obvia para dividir los datos, sería útil contar con un procedimiento formal para elegir los conjuntos de datos de estimación y de predicción.

Snee [1977] describe el algoritmo DUPLEX para dividir datos. Acredita a R. W. Kennard el desarrollo de ese procedimiento, y hace notar que se parece al algoritmo CADEX que propusieron Kennard y Stone [1969] para construcción de diseños. El algoritmo comienza con una lista de las n observaciones, en la que los k regresores están estandarizados a una longitud unitaria, esto es

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_{jj}^{1/2}}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k,$$

donde $S_{jj} = \Sigma_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ es la suma corregida de cuadrados del j-ésimo regresor. A continuación, se **ortonormalizan** los regresores estandarizados. Esto se hace factorizando la matriz $\mathbf{Z}'\mathbf{Z}$ como sigue:

$$\mathbf{Z}'\mathbf{Z} = \mathbf{T}'\mathbf{T} \tag{15.2}$$

http://carlos2524.jimdo.com/

siendo **T** una matriz triangular superior única, de $k \times k$. Los elementos de **T** se pueden determinar con el método de Cholesky, de raíz cuadrada (véase Graybill [1976, pp. 231-236]). En seguida se hace la transformación

$$\mathbf{W} = \mathbf{Z}\mathbf{T}^{-1} \tag{15.3}$$

y se obtiene un nuevo conjunto de variables (las w) que son ortogonales y tienen varianza unitaria. Esta transformación hace más esférico el espacio de los factores.

Se calcula la distancia euclidiana entre todos los $\binom{n}{2}$ pares de puntos usando los puntos ortonormalizados. El par de puntos más alejado se asigna al conjunto de datos de estimación. Este par de puntos se saca de la lista , y el siguiente par de puntos más alejado se asigna al conjunto de datos de predicción. A su vez, se quita de la lista este par de puntos y siguiente punto más alejado del par de puntos en el conjunto de datos de estimación, se incluye en ese mismo conjunto. En el siguiente paso, el siguiente punto no asignado que está más alejado de los dos puntos en el conjunto de datos de predicción se agrega a los datos de predicción. El algoritmo continúa colocando los puntos restantes en el conjunto de predicción o en el de estimación, hasta que se hayan asignado todas las n observaciones.

Snee [1977] sugiere medir las propiedades estadísticas de los conjuntos de datos de estimación y de predicción, comparando la p-ésima raíz de los determinantes de las matrices $\mathbf{X}'\mathbf{X}$ para esos dos conjuntos de datos, siendo p la cantidad de parámetros en el modelo. El determinante de $\mathbf{X}'\mathbf{X}$ se relaciona con el volumen de la región abarcada por los puntos, así, si $\mathbf{X}_E$ y $\mathbf{X}_P$ representan las matrices **X** de los puntos en los conjuntos de datos de estimación y de predicción, respectivamente, entonces

$$\left(\frac{|\mathbf{X}'_E\mathbf{X}_E|}{|\mathbf{X}'_P\mathbf{X}_P|}\right)^{1/p}$$

es una medida de los volúmenes relativos de las regiones abarcadas por esos dos conjuntos de datos. En el caso ideal, esa relación debería aproximarse a la unidad. También puede ser de utilidad examinar los factores de inflación de varianza para los dos conjuntos de datos, y el espectro de eigenvalores de $\mathbf{X}'_E\mathbf{X}_E$ y $\mathbf{X}'_P\mathbf{X}_P$, para medir la correlación relativa entre los regresores.

Al usar el algoritmo DUPLEX se deben tener en cuenta los siguientes puntos:

1. Algunos conjuntos de datos pueden ser demasiado pequeños para aplicar bien la división de datos. Snee [1977] sugiere que se requiere un mínimo de $n \geq 2p + 25$ observaciones si los conjuntos de datos de estimación y de predicción son del mismo tamaño, siendo p la mayor cantidad de parámetros que probablemente se requieran en el modelo. Este requisito de tamaño de muestra asegura que haya una cantidad razonable de grados de libertad de error para el modelo.

2. Aunque los conjuntos de datos de estimación y de predicción son, con frecuencia, del mismo tamaño, DUPLEX puede dividir los datos en cualquier relación que se prefiera. En forma típica, el conjunto de datos de estimación puede ser mayor que el de datos de predicción. Esas divisiones se determinan aplicando el algoritmo, hasta que el conjunto de datos de predicción contenga la cantidad requerida de puntos, para después poner los puntos restantes no asignados en el conjunto de datos de estimación. Recuérdese que el conjunto de datos de predicción debe contener al menos 15 puntos, para obtener una evaluación razonable de la eficacia del modelo.

http://carlos2524.jimdo.com/

3. Se deben eliminar las réplicas o los puntos que sean vecinos cercanos en el espacio x, antes de dividir los datos con DUPLEX. A menos que esas réplicas se eliminen, los conjuntos de datos de predicción y de estimación se parecerían mucho, y así no necesariamente se probaría el modelo con la suficiente severidad. En un caso extremo, en el que cada punto tiene réplica, DUPLEX formaría un conjunto de datos de estimación con una réplica y un conjunto de datos de predicción con la otra. Snee [1977] sugiere usar análisis de grupo para identificar los vecinos cercanos, y el promedio de las coordenadas x de esos puntos se debe usar en el procedimiento de división de datos.

4. Una **desventaja** potencial de la división de datos es que **reduce la precisión** con la que se estiman los coeficientes de regresión, lo que indica que los errores estándar de los coeficientes de regresión, obtenidos con el conjunto de datos de estimación serán mayores que lo que serían si se hubieran usado todos los datos para estimar los coeficientes. En conjuntos grandes de datos, los errores estándar pueden ser lo suficientemente pequeños como para que no sea importante esa pérdida de precisión, sin embargo, el aumento porcentual en los errores estándar puede ser grande. Si el modelo obtenido con el conjunto de datos de estimación es un predictor satisfactorio, una forma de mejorar la precisión de la estimación es volver a estimar los coeficientes usando todo el conjunto de datos. Los estimados de los coeficientes en los dos análisis deberían ser muy parecidos, si el modelo es un predictor adecuado del conjunto de datos de predicción.

5. La **doble validación cruzada** puede ayudar en algunos problemas, ya que es un procedimiento en el que primero se dividen los datos en conjuntos de estimación y de predicción, se desarrolla un modelo con los datos de estimación y se investiga su eficiencia con los datos de predicción, a continuación se **invierten** los papeles de los dos conjuntos de datos; se desarrolla un modelo con los **datos originales de predicción** y se usa para predecir los **datos originales de estimación**. La ventaja de este procedimiento es que proporciona dos evaluaciones de la eficiencia del modelo. La desventaja es que hay **tres** modelos de dónde escoger: los dos desarrollados con la división de datos y el modelo ajustado a todos los datos. Si el modelo es buen predictor, será poca la diferencia en usar cualquiera de ellos, excepto que los errores estándar de los coeficientes en el modelo ajustado con el conjunto total de datos será menor. Si hay grandes diferencias en la eficiencia de predicción, los estimados de coeficientes o la forma funcional de esos modelos, es necesario más análisis para determinar las razones de esas diferencias.

Ejemplo 15.3 *Datos del tiempo de entrega*

En la tabla 15.2 se muestran las 40 observaciones de los datos del tiempo de entrega en los ejemplos 3.1 y 15.2. Se supondrá que los 40 puntos se recolectaron de una sola vez, y se usarán los datos para ilustrar la división de datos con el algoritmo DUPLEX. Ya que el modelo tendrá dos regresores, una división igual de los datos dará como resultado 17 grados de libertad para los datos de estimación. Es adecuado, por lo que se puede aplicar DUPLEX para generar los conjuntos de datos de estimación y predicción. En la figura 15.1 se muestra una gráfica $x_1 - x_2$; al examinar los datos se ve que hay dos pares de puntos que son vecinos cercanos en el espacio x: las observaciones 15 y 23, y las observaciones 16 y 32. Esos dos grupos de puntos se encierran con círculo en la figura 15.1. Las coordenadas $x_1 - x_2$ de esos grupos de puntos se promedian, y la lista de puntos para usar en el algoritmo DUPLEX se ve en las columnas 1 y 2 de la tabla 15.3.

Los datos estandarizados y ortonormalizados se ven en las columnas 3 y 4 de la tabla 15.3, y se grafican en la figura 15.2. Nótese que la región cubierta es más esférica que la

http://carlos2524.jimdo.com/

TABLA 15.2 Datos del tiempo de entrega

Observación i	Cajas x_1	Distancia x_2	Tiempo de entrega y	Conjunto de datos de estimación (E) o de predicción (P)
1	7	560	16.68	P
2	3	220	11.50	P
3	3	340	12.03	P
4	4	80	14.88	E
5	6	150	13.75	E
6	7	330	18.11	E
7	2	110	8.00	E
8	7	210	17.83	E
9	30	1460	79.24	E
10	5	605	21.50	E
11	16	688	40.33	P
12	10	215	21.00	P
13	4	255	13.50	E
14	6	462	19.75	P
15	9	448	24.00	E
16	10	776	29.00	P
17	6	200	15.35	P
18	7	132	19.00	E
19	3	36	9.50	P
20	17	770	35.10	E
21	10	140	17.90	E
22	26	810	52.32	E
23	9	450	18.75	E
24	8	635	19.83	E
25	4	150	10.75	E
26	22	905	51.00	P
27	7	520	16.80	E
28	15	290	26.16	P
29	5	500	19.90	E
30	6	1000	24.00	E
31	6	225	18.55	E
32	10	775	31.93	P
33	4	212	16.95	P
34	1	144	7.00	P
35	3	126	14.00	P
36	12	655	37.03	P
37	10	420	18.62	P
38	7	150	16.10	P
39	8	360	24.38	P
40	32	1530	64.75	P

http://carlos2524.jimdo.com/

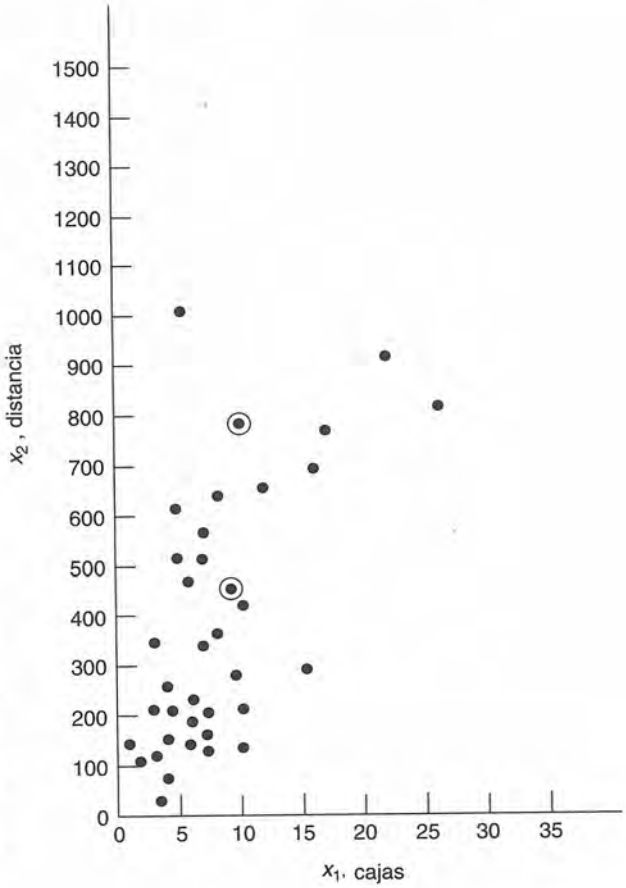


Figura 15.1 Diagrama de dispersión del volumen entregado, x_1, en función de la distancia, x_2; ejemplo 15.3.

de la figura 15.1. La figura 15.2 y las tablas 15.2 y 15.3 también muestran cómo divide DUPLEX los puntos originales en datos de estimación y de predicción. Los contornos convexos de los dos conjuntos de datos se ven en la figura 15.2. Esto indica que el conjunto de datos de predicción contiene puntos tanto de interpolación como de extrapolación. Para esos dos conjuntos de datos se ve que $|\mathbf{X}'_E\mathbf{X}_E| = 0.44696$ y que $|\mathbf{X}'_P\mathbf{X}_P| = 0.22441$. Entonces,

$$\left(\frac{|\mathbf{X}'_E\mathbf{X}_E|}{|\mathbf{X}'_P\mathbf{X}_P|}\right)^{1/3} = \left(\frac{0.44696}{0.22441}\right)^{1/3} = 1.26$$

que indica que los volúmenes de las dos regiones son muy parecidos. Los factores de inflación de varianza para los datos de estimación y de predicción son 2.22 y 4.43, respectivamente, por lo que no hay una evidencia fuerte de multicolinealidad, y ambos conjuntos de datos tienen una estructura correlativa semejante.

La parte A de la tabla 15.4 es un resumen del ajuste a los datos de estimación, por mínimos cuadrados. En este modelo, los estimados de los parámetros tienen signos y magnitudes razonables, y los factores de inflación de varianza son aceptablemente pequeños. El análisis de los residuales (que no se muestra) indica que no hay inadecuaciones graves de modelo, pero la gráfica de probabilidad normal indica que la distribución de los errores tiene colas

http://carlos2524.jimdo.com/

TABLA 15.3 Datos del tiempo de entrega promediando los puntos vecinos cercanos

Observación i	(1) Variables originales x_1 (Cajas)	(2) Variables originales x_2 (Distancia)	(3) Datos estandarizados y ortonormalizados w_1	(4) Datos estandarizados y ortonormalizados w_2	Conjunto de datos de estimación (E) o de predicción (P)
1	7	560	−.047671	.158431	P
2	3	220	−.136037	.013739	P
3	3	340	.136037	.108082	P
4	4	80	−.113945	−.126981	E
5	6	150	−.069762	−.133254	E
6	7	330	−.047671	−.022393	E
7	2	110	−.158128	−.042089	E
8	7	210	−.047671	−.116736	E
9	30	1460	.460432	.160977	E
10	5	605	−.091854	.255116	E
11	16	688	.151152	−.016816	P
12	10	215	.018603	−.204765	P
13	4	255	−.113945	.010603	E
14	6	462	−.069762	.112038	P
15, 23	9	449	−.003488	.009857	E
16, 32	10	775.5	.018603	.235895	P
17	6	200	−.069762	−.093945	P
18	7	132	−.047671	−.178059	E
19	3	36	−.136037	−.130920	P
20	17	770	.173243	.016998	E
21	10	140	.018603	−.263729	E
22	26	810	.372066	−.227434	E
24	8	635	−.025580	.186742	E
25	4	150	−.113945	−.071948	E
26	22	905	.283700	−.030133	P
27	7	520	−.047671	.126983	E
28	15	290	.129060	−.299067	P
29	5	500	−.091854	.172566	E
30	6	1000	−.069762	.535009	E
31	6	225	−.069762	−.074290	E
33	4	212	−.113945	−.023204	P
34	1	144	−.180219	.015295	P
35	3	126	−.136037	−.060163	P
36	12	655	.062786	.079853	P
37	10	420	.018603	−.043596	P
38	7	150	−.047671	−.163907	P
39	8	360	−.025580	−.029461	P
40	32	1530	.504614	.154704	P

http://carlos2524.jimdo.com/

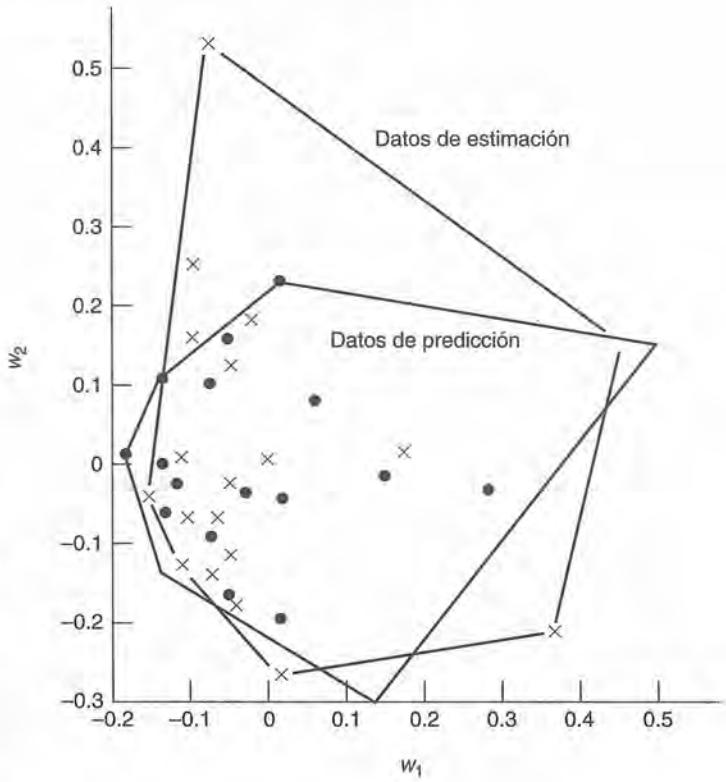


Figura 15.2 Datos de estimación (×) y datos de predicción (•) usando regresores ortonormalizados.

TABLA 15.4 Resumen del ajuste a los datos del tiempo de entrega por mínimos cuadrados

A. Análisis con datos de estimación				B. Análisis con todos los datos			
Variable	Estimado del coeficiente	Error estándar	t_0	Variable	Estimado del coeficiente	Error estándar	t_0
Ordenada al origen	2.4123	1.4165	1.70	Ordenada al origen	3.9840	0.9861	4.04
x_1	1.6392	0.1769	9.27	x_1	1.4877	0.1376	10.81
x_2	0.0136	0.0036	3.78	x_2	0.0134	0.0028	4.72
$MS_{\text{Res}} = 13.9145$, $R^2 = 0.952$				$MS_{\text{Res}} = 13.6841$, $R^2 = 0.944$			

más gruesas que la distribución normal. Al revisar la tabla 15.2 se ve que el punto 9, que ya se vio antes que es influyente, está en el conjunto de datos de estimación, además de la preocupación acerca de la suposición de normalidad y de la influencia del punto 9, se llega a la conclusión que el ajuste de los datos de estimación por mínimos cuadrados es razonable.

Las columnas 2 y 3 de la tabla 15.5 muestran los resultados de calcular las observaciones del conjunto de datos de predicción, mediante el modelo de mínimos cuadrados desarrollado con los datos de estimación. Se ve que en general los valores predichos corresponden bien a los valores observados. El único error de predicción anormalmente grande es para el punto 40, que tiene el máximo tiempo observado en los datos de predicción. Este punto también tiene los valores máximos de x_1 (32 cajas) y x_2 (1 530 pies) en el

http://carlos2524.jimdo.com/

TABLA 15.5 Eficiencia de predicción de modelos desarrollados con los datos de estimación

	(1)	(2)	(3)	(4)	(5)
		Ajuste por mínimos cuadrados		Ajuste robusto ($E_{0.3}$ de Ramsay)	
Observación i	Observado y_i	Predicho $\hat{y}_i$	Error de predicción $e_i = y_i - \hat{y}_i$	Predicho $\hat{y}_i$	Error de predicción $e_i = y_i - \hat{y}_i$
1	16.68	21.4976	– 4.8176	21.5240	– 4.8440
2	11.50	10.3199	1.1801	10.7482	0.7518
3	12.03	11.9508	0.0792	12.3055	– 0.2755
11	40.33	37.9901	2.3399	37.5032	2.8268
12	21.00	21.7264	– 0.7264	21.8197	– 0.8197
14	19.75	18.5265	1.2235	18.6614	1.0886
16	29.00	29.3509	– 0.3509	29.0998	– 0.0098
17	15.35	14.9657	0.3843	15.2614	0.0886
19	9.50	7.8192	1.6808	8.3605	1.1395
26	51.00	50.7746	0.2254	49.8646	1.1354
28	26.16	30.9417	– 4.7817	30.7474	– 4.5874
32	31.93	29.3373	2.5927	29.0868	2.8432
33	16.95	11.8504	5.0996	12.2353	4.7147
34	7.00	6.0086	0.9914	6.5802	0.4198
35	14.00	9.0424	4.9576	9.5284	4.4716
36	37.03	30.9848	6.0452	30.7113	6.3187
37	18.62	24.5125	– 5.8925	24.4799	– 5.8599
38	16.10	15.9254	0.1746	16.2034	– 0.1034
39	24.38	20.4187	3.9613	20.5195	3.8605
40	64.75	75.6609	– 10.9109	73.8842	– 9.1342

conjunto total de datos. Se parece mucho al punto 9 en los datos de estimación ($x_1 = 30$, $x_2 = 1\,460$), pero representa una extrapolación del modelo ajustado a los datos de estimación. La suma de cuadrados de los errores de predicción es $\Sigma e_i^2 = 322.4452$, y la R^2 aproximada para predicción es

$$R_{\text{Perdicción}} = 1 - \frac{\Sigma e_i^2}{SS_{\text{T}}} = 1 - \frac{322.4452}{4\,113.5442} = 0.922$$

en donde $SS_{\text{T}} = 4\,113.5442$ es la suma corregida de cuadrados de las respuestas en el conjunto de datos de predicción. Así, cabe esperar que este modelo "explique" un 92.2% de la variabilidad en los nuevos datos, en comparación del 95.2% explicado por el ajuste a los datos de estimación por mínimos cuadrados. Esta pérdida de R^2 es pequeña, por lo que hay indicios razonablemente fuertes de que el modelo de mínimos cuadrados será un predictor satisfactorio.

La parte B de la tabla 15.4 muestra los resultados de un ajuste por mínimos cuadrados a todas las 40 observaciones. Nótese que los estimados de la ordenada al origen y de β_1 son algo distintos de los obtenidos con el ajuste a los datos de estimación. Además, como era de esperarse, los errores estándar de los coeficientes de regresión del ajuste que usa todos los datos son menores que sus contrapartes, cuando sólo se usaron los datos de estimación. El conjunto de datos de estimación contiene al punto 9, que ya se sabe que es influyente. Si se

http://carlos2524.jimdo.com/

elimina el punto 9, la ordenada al origen será mayor y β_1 será menor, como sucede en el ajuste por mínimos cuadrados para el conjunto de todos los datos. Un ajuste robusto a los datos de estimación produciría un efecto parecido. El conjunto de todos los datos incluye también al punto 40 (parecido al punto 9). Es posible que el mayor conjunto de datos diluya el efecto de esos puntos influyentes.

Como hay cierta evidencia de que el ajuste por mínimos cuadrados para los datos de estimación puede estar influido por el punto 9, sería interesante examinar la eficiencia predictiva de un ajuste robusto a esos datos. Ese ajuste robusto, mediante la función $E_{0.3}$ de Ramsay es

$$\hat{y} = 3.1206 + 1.5909x_1 + 0.0130$$

Las columnas 4 y 5 de la tabla 15.5 muestran los resultados de aplicar este modelo a los datos de predicción. La suma de cuadrados de los errores de predicción es $\Sigma e_i^2 = 280.6992$, menor que la del ajuste por mínimos cuadrados. La R^2 aproximada de predicción para el ajuste robusto es

$$R^2_{\text{Predicción}} = 1 - \frac{\Sigma e_i^2}{SS_{\text{T}}} = 1 - \frac{280.6992}{4\,113.5442} = 0.932$$

que sólo es un poco mayor que la obtenida con el modelo de mínimos cuadrados. También, es probable que el ajuste robusto sea un predictor satisfactorio.

¿Cuál modelo se debe recomendar para usarlo en la práctica? En términos de la eficiencia predictiva, hay poca diferencia entre los ajustes por mínimos cuadrados y robusto, a los datos de estimación. Como el ajuste robusto no se afecta en forma tan grave debido a las observaciones remotas, en comparación con los mínimos cuadrados, algunos analistas lo considerarán un poco más preferible. En esencia, éstas son las mismas conclusiones a las que se llegó con otros estudios de validación, para este conjunto de datos, en la sección 15.2.2.

15.3 DATOS OBTENIDOS CON EXPERIMENTOS PLANEADOS

En la mayor parte de las técnicas de validación que se describieron en este capítulo se supone que el modelo se desarrolla a partir de datos no planeados. Si bien esas técnicas también se podrían aplicar en casos en los que se usó un experimento planeado para reunir los datos, por lo general es más fácil la validación de un modelo desarrollado con datos planeados. Muchos diseños de experimentos dan como resultado coeficientes de regresión casi no correlacionados, por lo que la multicolinealidad no suele ser problema. Un aspecto importante del diseño de experimentos es la selección de los factores por estudiar, y la identificación de los intervalos dentro de los que se van a variar. Si esto se hace bien, ayuda a asegurar que todos los regresores importantes estén incluidos en los datos, y que se obtenga un intervalo adecuado de valores para cada regresor; además, en los experimentos diseñados se suele dar mucha importancia al proceso mismo de obtención de datos. Eso ayuda a minimizar problemas con observaciones "raras" o dudosas, y produce datos con errores de medición relativamente pequeños.

Cuando se usan experimentos planeados para reunir datos, se suele aconsejar hacer ensayos adicionales para usarlos en la prueba de la eficiencia predictiva del modelo. En las publicaciones sobre diseño de experimentos, a esas pruebas adicionales se les llama **corridas**

http://carlos2524.jimdo.com/

de confirmación. Un método que se usa mucho es incluir los puntos que permitirían ajustar un modelo en un grado mayor que el que se esté empleando; así, si se espera ajustar un modelo de primer orden, el diseño debe incluir suficientes puntos para ajustar al menos algunos de los términos de un modelo de segundo orden.

PROBLEMAS

15.1 Se tiene el modelo de regresión desarrollado para los datos de la Liga Nacional de Futbol en el problema 3.1.

a. Calcular el estadístico PRESS para este modelo. ¿Qué comentarios se pueden hacer acerca del probable desempeño predictivo de este modelo?

b. Eliminar la mitad de las observaciones (seleccionadas al azar), y volver a ajustar el modelo de regresión. ¿Cambiaron mucho los coeficientes de regresión? ¿Qué tan bien predice este modelo la cantidad de juegos ganados para las observaciones eliminadas?

c. Eliminar la observación para Dallas, Los Ángeles, Houston, San Francisco, Chicago y Atlanta, y volver a ajustar el modelo. ¿Qué tan bien predice este modelo la cantidad de juegos ganados por esos equipos?

15.2 Dividir los datos de la Liga Nacional de Futbol del problema 3.1, en conjuntos de datos de estimación y de predicción. Evaluar las propiedades estadísticas de esos dos conjuntos de datos. Desarrollar un modelo a partir de los datos de estimación y evaluar su desempeño con los datos de predicción. Comentar el desempeño predictivo de este modelo.

15.3 Calcular el estadístico PRESS para el modelo desarrollado con los datos de estimación en el problema 15.2. ¿Es probable que el modelo pueda predecir bien? Comparar esta indicación de la eficiencia predictiva con la eficiencia real, observada en el problema 15.2.

15.4 Para los datos de tiempo de entrega mencionados en el ejemplo 15.3, calcular el estadístico PRESS para el modelo desarrollado con los datos de estimación. ¿Qué tan bien es probable que funcione el modelo como predictor? Compararlo con la eficiencia observada en predicción.

15.5 Se tienen los datos del tiempo de entrega del ejemplo 15.3.

a. Desarrollar un modelo de regresión con el conjunto de datos de predicción.

b. ¿Cómo se comparan los estimados de parámetro en este modelo, con los del modelo desarrollado con los datos de estimación? ¿Qué implica eso acerca de la validez del modelo?

c. Usar el modelo desarrollado en la parte a para predecir los tiempos de entrega para las observaciones en los datos originales de estimación. ¿Son consistentes los resultados con los obtenidos en el ejemplo 15.3?

15.6 En el problema 3.5 se desarrolló un modelo de regresión para los datos de rendimiento de gasolina, usando el regresor x_1, cilindrada del motor, y el x_6, cantidad de gargantas del carburador. Calcular el estadístico PRESS para este modelo. ¿A qué conclusión se puede llegar acerca del probable desempeño predictivo del modelo?

15.7 En el problema 3.6 se desarrolló un modelo de regresión para los datos de rendimiento de gasolina, con los regresores x_8, longitud del vehículo, y x_{10}, peso del vehículo. Calcular el estadístico PRESS para este modelo. ¿Qué conclusiones se pueden sacar acerca de la eficiencia potencial de este modelo como predictor?

http://carlos2524.jimdo.com/

15.8 Se calcularon estadísticos PRESS con dos modelos distintos para los datos del rendimiento de la gasolina, en los problemas 15.6 y 15.7. Con base en esos estadísticos, ¿cuál modelo es probable que sea el mejor predictor?

15.9 Se tienen los datos del rendimiento de la gasolina en la tabla B.3 del apéndice. Eliminar ocho observaciones (eligiéndolas al azar), y desarrollar un modelo adecuado de regresión. Usar este modelo para predecir las ocho observaciones que se separaron. ¿Qué evaluación se podría hacer del desempeño predictivo de este modelo?

15.10 Para los datos del rendimiento de la gasolina en la tabla B.3 del apéndice, dividirlos en conjuntos de estimación y de predicción.

a. Evaluar las propiedades estadísticas de esos datos.

b. Ajustar un modelo donde intervengan x_1 y x_6 para estimar los datos. ¿Parecen razonables los coeficientes y los valores ajustados, obtenidos con este modelo?

c. Usar este modelo para predecir las observaciones del conjunto de datos de predicción. ¿Cómo se evaluaría el desempeño predictivo de este modelo?

15.11 Véase el problema 15.2. ¿Cuáles son los errores estándar de los coeficientes de regresión del modelo desarrollado con los datos de estimación? ¿Cómo se comparan con los errores estándar del modelo del problema 3.5, usando todos los datos?

15.12 Véase el problema 15.2. Desarrollar un modelo para los datos de la Liga Nacional de Futbol, usando el conjunto de datos de predicción.

a. ¿Cómo se comparan los coeficientes y los valores estimados con las cantidades correspondientes de los modelos desarrollados con los datos de estimación?

b. ¿Qué tan bien predice este modelo las observaciones del conjunto original de datos de estimación?

15.13 ¿Qué dificultades es probable encontrar al desarrollar un programa de cómputo que implemente el algoritmo DUPLEX? Por ejemplo, ¿cuán eficiente resulta, probablemente, el procedimiento para tamaños grandes de muestra? ¿Qué modificaciones del procedimiento se pueden sugerir para superar esas dificultades?

15.14 Si $\mathbf{Z}$ es la matriz $n \times k$ de regresores estandarizados, y $\mathbf{T}$ es la matriz triangular superior de $k \times k$ en la ecuación (15.3), demostrar que los regresores transformados $\mathbf{W} = \mathbf{Z}\mathbf{T}^{-1}$ son ortogonales y tienen varianza unitaria.

15.15 Demostrar que el estimado de β por mínimos cuadrados (digamos que es $\hat{\beta}_{(i)}$) omitiendo la i-ésima observación, se puede expresar como sigue, en función del estimado basado en todos los n puntos:

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{e_i}{1 - h_{ii}}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i.$$

15.16 Acerca de los datos de tratamiento térmico en la tabla B.12 del apéndice, dividirlos en conjuntos de datos de predicción y de estimación.

a. Ajustar un modelo al conjunto de datos de estimación, usando todas las regresiones posibles. Seleccionar el modelo con C_p mínimo.

b. Usar el modelo de la parte a para predecir las respuestas de cada observación del conjunto de datos de predicción. Calcular la R^2 de predicción. Comentar acerca de la adecuación del modelo.

15.17 Examinar los datos de empuje de la turbina de reacción, en la tabla B.13 del apéndice. Dividirlos en conjunto de datos de predicción y de estimación.

http://carlos2524.jimdo.com/

a. Ajustar un modelo a los datos de estimación, usando todas las regresiones posibles. Seleccionar el modelo con C_p mínimo.

b. Con el modelo de la parte a, predecir cada observación del conjunto de datos de predicción. Calcular R^2 de predicción. Comentar la adecuación del modelo.

15.18 Para los datos del inversor electrónico, en la tabla B.14 del apéndice, eliminar la segunda observación del conjunto y dividir las restantes en conjuntos de datos de predicción y de estimación.

a. Calcular la ecuación de C_p mínimo para el conjunto de datos de estimación.

b. Usar el modelo de la parte a para predecir cada observación del conjunto de datos de predicción. Calcular la R^2 de predicción y comentar la adecuación del modelo.

15.19 En la tabla B.11 del apéndice se presentan 38 observaciones de calidad de vino.

a. Seleccionar cuatro observaciones al azar de este conjunto, separarlas y ajustar un modelo donde sólo intervenga el aroma y las variables indicadoras de información de región, a las observaciones restantes. Usar este modelo para predecir las observaciones separadas y calcular R^2 para predicción.

b. Repetir 100 veces la parte a, y calcular el promedio de las R^2 para predicción para las 100 repeticiones.

c. Ajustar el modelo con las 38 observaciones y calcular R^2 para predicción, con base en PRESS.

d. Comentar los tres métodos, de las partes a a c anteriores, como medidas de la validez del modelo.

15.20 Examinar las 40 observaciones del tiempo de entrega. Separar 10% (es decir 4) de las observaciones, al azar. Ajustar un modelo a las 36 observaciones restantes, predecir los cuatro valores separados y calcular R^2 de predicción. Repetir 100 veces estos cálculos y calcular el promedio de las R^2 para predicción. ¿Qué información produce lo anterior acerca de la capacidad predictiva del modelo? ¿Cómo se compara el promedio de las 100 R^2 para predicción de valores, con la R^2 de predicción basada en PRESS para las 40 observaciones?

http://carlos2524.jimdo.com/

APÉNDICE A

TABLAS ESTADÍSTICAS

Tabla A.1 Distribución normal estándar acumulada
Tabla A.2 Puntos porcentuales de la distribución χ^2
Tabla A.3 Puntos porcentuales de la distribución t
Tabla A.4 Puntos porcentuales de la distribución F
Tabla A.5 Polinomios ortogonales
Tabla A.6 Valores críticos del estadístico Durbin-Watson

http://carlos2524.jimdo.com/

TABLA A.1 Distribución normal estándar acumulada

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$

z	.00	.01	.02	.03	.04	z
.0	.50000	.50399	.50798	.51197	.51595	.0
.1	.53983	.54379	.54776	.55172	.55567	.1
.2	.57926	.58317	.58706	.59095	.59483	.2
.3	.61791	.62172	.62551	.62930	.63307	.3
.4	.65542	.65910	.66276	.66640	.67003	.4
.5	.69146	.69497	.69847	.70194	.70540	.5
.6	.72575	.72907	.73237	.73565	.73891	.6
.7	.75803	.76115	.76424	.76730	.77035	.7
.8	.78814	.79103	.79389	.79673	.79954	.8
.9	.81594	.81859	.82121	.82381	.82639	.9
1.0	.84134	.84375	.84613	.84849	.85083	1.0
1.1	.86433	.86650	.86864	.87076	.87285	1.1
1.2	.88493	.88686	.88877	.89065	.89251	1.2
1.3	.90320	.90490	.90658	.90824	.90988	1.3
1.4	.91924	.92073	.92219	.92364	.92506	1.4
1.5	.93319	.93448	.93574	.93699	.93822	1.5
1.6	.94520	.94630	.94738	.94845	.94950	1.6
1.7	.95543	.95637	.95728	.95818	.95907	1.7
1.8	.96407	.96485	.96562	.96637	.96711	1.8
1.9	.97128	.97193	.97257	.97320	.97381	1.9
2.0	.97725	.97778	.97831	.97882	.97932	2.0
2.1	.98214	.98257	.98300	.98341	.98382	2.1
2.2	.98610	.98645	.98679	.98713	.98745	2.2
2.3	.98928	.98956	.98983	.99010	.99036	2.3
2.4	.99180	.99202	.99224	.99245	.99266	2.4
2.5	.99379	.99396	.99413	.99430	.99446	2.5
2.6	.99534	.99547	.99560	.99573	.99585	2.6
2.7	.99653	.99664	.99674	.99683	.99693	2.7
2.8	.99744	.99752	.99760	.99767	.99774	2.8
2.9	.99813	.99819	.99825	.99831	.99836	2.9
3.0	.99865	.99869	.99874	.99878	.99882	3.0
3.1	.99903	.99906	.99910	.99913	.99916	3.1
3.2	.99931	.99934	.99936	.99938	.99940	3.2
3.3	.99952	.99953	.99955	.99957	.99958	3.3
3.4	.99966	.99968	.99969	.99970	.99971	3.4
3.5	.99977	.99978	.99978	.99979	.99980	3.5
3.6	.99984	.99985	.99985	.99986	.99986	3.6
3.7	.99989	.99990	.99990	.99990	.99991	3.7
3.8	.99993	.99993	.99993	.99994	.99994	3.8
3.9	.99995	.99995	.99996	.99996	.99996	3.9

http://carlos2524.jimdo.com/

TABLA A.1 (*Continuación*)

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$

z	.05	.06	.07	.08	.08	z
.0	.51994	.52392	.52790	.53188	.53586	.0
.1	.55962	.56356	.56749	.57142	.57534	.1
.2	.59871	.60257	.60642	.61026	.61409	.2
.3	.63683	.64058	.64431	.64803	.65173	.3
.4	.67364	.67724	.68082	.68438	.68793	.4
.5	.70884	.71226	.71566	.71904	.72240	.5
.6	.74215	.74537	.74857	.75175	.75490	.6
.7	.77337	.77637	.77935	.78230	.78523	.7
.8	.80234	.80510	.80785	.81057	.81327	.8
.9	.82894	.83147	.83397	.83646	.83891	.9
1.0	.85314	.85543	.85769	.85993	.86214	1.0
1.1	.87493	.87697	.87900	.88100	.88297	1.1
1.2	.89435	.89616	.89796	.89973	.90147	1.2
1.3	.91149	.91308	.91465	.91621	.91773	1.3
1.4	.92647	.92785	.92922	.93056	.93189	1.4
1.5	.93943	.94062	.94179	.94295	.94408	1.5
1.6	.95053	.95154	.95254	.95352	.95448	1.6
1.7	.95994	.96080	.96164	.96246	.96327	1.7
1.8	.96784	.96856	.96926	.96995	.97062	1.8
1.9	.97441	.97500	.97558	.97615	.97660	1.9
2.0	.97982	.98030	.98077	.98124	.98169	2.0
2.1	.98422	.98461	.98500	.98537	.98574	2.1
2.2	.98778	.98809	.98840	.98870	.98899	2.2
2.3	.99061	.99086	.99111	.99134	.99158	2.3
2.4	.99286	.99305	.99324	.99343	.99361	2.4
2.5	.99461	.99477	.99492	.99506	.99520	2.5
2.6	.99598	.99609	.99621	.99632	.99643	2.6
2.7	.99702	.99711	.99720	.99728	.99736	2.7
2.8	.99781	.99788	.99795	.99801	.99807	2.8
2.9	.99841	.99846	.99851	.99856	.99861	2.9
3.0	.99886	.99889	.99893	.99897	.99900	3.0
3.1	.99918	.99921	.99924	.99926	.99929	3.1
3.2	.99942	.99944	.99946	.99948	.99950	3.2
3.3	.99960	.99961	.99962	.99964	.99965	3.3
3.4	.99972	.99973	.99974	.99975	.99976	3.4
3.5	.99981	.99981	.99982	.99983	.99983	3.5
3.6	.99987	.99987	.99988	.99988	.99989	3.6
3.7	.99991	.99992	.99992	.99992	.99992	3.7
3.8	.99994	.99994	.99995	.99995	.99995	3.8
3.9	.99996	.99996	.99996	.99997	.99997	3.9

Fuente: Reproducida con autorización de *Probability and Statistics in Engineering and Management Science*, tercera edición, 1990, por W. W. Hines y D. C. Montgomery, Wiley, New York.

http://carlos2524.jimdo.com/

TABLA A.2 Puntos porcentuales de la distribución χ^2

	α										
ν	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.48	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	2.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.45
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	139.56	135.81	140.17

n = grados de libertad.

Fuente: Reproducida con autorización de *Probability and Statistics in Engineering and Management Science*, tercera edición, 1990, por W. W. Hines y D. C. Montgomery, Wiley, New York.

http://carlos2524.jimdo.com/

TABLA A.3 Puntos porcentuales de la distribución t

	α									
v	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	2.833	4.504	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	2.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	8.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	2.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.308	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.648	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Fuente: Adaptada con autorización de *Biometrika Tables for Statisticians*, vol. 1, tercera edición, 1966, por E. S. Pearson y H. O. Hartley, Cambridge University Press, Cambridge.

http://carlos2524.jimdo.com/

TABLA A.4 Puntos porcentuales de la distribución F

$F_{.25,\,\nu_1,\,\nu_2}$

Grados de libertad para el denominador (ν_2)	Grados de libertad para el numerador (ν_1)								
ν_2	1	2	3	4	5	6	7	8	9
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.70
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27

http://carlos2524.jimdo.com/

TABLA A.4 (*Continuación*)

$F_{.25, \nu_1, \nu_2}$

Grados de libertad para el numerador (ν_1)									
10	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48
1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42
1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38
1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
1.44	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30
1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29
1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28
1.39	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.28
1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	1.27
1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26
1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25
1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25
1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24
1.36	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.25	1.24
1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23
1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23
1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19
1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10
1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00

http://carlos2524.jimdo.com/

TABLA A.4 (*Continuación*)

$F_{.10,\,\nu_1,\,\nu_2}$

Grados de libertad para el denominador (ν_2)	Grados de libertad para el numerador (ν_1)								
ν_2	1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.95	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.03	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.17	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

http://carlos2524.jimdo.com/

TABLA A.4 (*Continuación*)

$F_{.10, \nu_1, \nu_2}$

Grados de libertad para el numerador (ν_1)									
10	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	2.76
3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

http://carlos2524.jimdo.com/

TABLA A.4 (*Continuación*)

$F_{.05,\,\nu_1,\,\nu_2}$

Grados de libertad para el denominador (ν_2)	Grados de libertad para el numerador (ν_1)								
ν_2	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.48	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

http://carlos2524.jimdo.com/

TABLA A.4 (*Continuación*)

$F_{.05,\,\nu_1,\,\nu_2}$

Grados de libertad para el numerador (ν_1)									
10	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
1.91	1.83	1.75	1.66	1.61	1.55	1.55	1.43	1.35	1.25
1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

http://carlos2524.jimdo.com/

TABLA A.4 (*Continuación*)

$F_{.025, v_1, v_2}$

Grados de libertad para el denominador (v_2)	Grados de libertad para el numerador (v_1)								
v_2	1	2	3	4	5	6	7	8	9
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

http://carlos2524.jimdo.com/

TABLA A.4 (*Continuación*)

$F_{.025,\,\nu_1,\,\nu_2}$

Grados de libertad para el numerador (ν_1)									
10	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
39.40	39.41	39.43	39.45	39.46	39.46	39.47	29.48	39.49	39.50
14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

http://carlos2524.jimdo.com/

TABLA A.4 (*Continuación*)

$F_{.01, \nu_1, \nu_2}$

Grados de libertad para el denominador (ν_2)	Grados de libertad para el numerador (ν_1)								
ν_2	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.36	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

http://carlos2524.jimdo.com/

TABLA A.4 (*Continuación*)

$F_{.01, \nu_1, \nu_2}$

Grados de libertad para el numerador (ν_1)									
10	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
27.23	27.05	26.87	26.69	26.00	26.50	26.41	26.32	26.22	26.13
14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
4.85	4.71	4.56	4.41	4.33	5.25	4.17	4.08	4.00	3.91
4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.59
3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Fuente: Adaptada con autorización de *Biometrika Tables for Statisticians*, vol. 1, tercera edición, 1996, por E. S. Pearson y H. O. Hartley, Cambridge University Press, Cambridge.

http://carlos2524.jimdo.com/

TABLA A.5 Polinomios ortogonales

	$n = 3$		$n = 4$			$n = 5$			
X_j	P_1	P_2	P_1	P_2	P_3	P_1	P_2	P_3	P_3
1	–1	1	–3	1	–1	–2	2	–1	1
2	0	–2	–1	–1	3	–1	–1	2	–4
3	1	1	1	–1	–3	0	–2	0	6
4			3	1	1	1	–1	–2	–4
5						2	2	1	1
6									
7									
$\sum_{j=1}^{n} \{P_i(X_j)\}^2$	2	6	20	4	20	10	14	10	70
λ	1	3	2	1	$\frac{10}{3}$	1	1	$\frac{5}{6}$	$\frac{35}{12}$

	$n = 8$						$n = 9$	
X_j	P_1	P_2	P_3	P_4	P_5	P_6	PP_1	P_2
1	–7	7	–7	7	–7	1	–4	28
2	–5	1	5	–13	23	–5	–3	7
3	–5	–3	7	–3	–17	9	–2	–8
4	–1	–5	3	9	–15	–5	–1	–17
5	1	–5	–3	9	15	–5	0	–20
6	3	–3	–7	–3	17	9	1	–17
7	5	1	–5	–13	–23	–5	2	–8
8	7	7	7	7	7	1	3	7
9							4	28
10								
$\sum_{j=1}^{n} \{P_i(X_j)\}^2$	168	168	264	616	2 184	264	60	2 772
λ	2	1	$\frac{2}{3}$	$\frac{7}{12}$	$\frac{7}{10}$	$\frac{11}{60}$	1	3

http://carlos2524.jimdo.com/

TABLA A.5 (*Continuación*)

$n = 6$					$n = 7$					
P_1	P_2	P_3	P_4	P_5	P_1	P_2	P_3	P_4	P_5	P_6
–5	5	–5	1	–1	–3	5	–1	3	–1	1
–3	–1	7	–3	5	–2	0	1	–7	4	–6
–1	–4	4	2	–10	–1	–3	1	1	–5	15
1	–4	–4	2	10	0	–4	0	6	0	–20
3	–1	–7	–3	–5	1	–3	–1	1	5	5
5	5	5	1	1	2	0	–1	–7	–4	–6
					3	5	1	3	1	1
70	84	180	28	252	28	84	6	154	84	924
2	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{7}{12}$	$\frac{21}{10}$	1	1	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{7}{20}$	$\frac{77}{60}$

$n = 9$				$n = 10$					
P_3	P_4	P_5	P_6	P_1	P_2	P_3	P_4	P_5	P_6
–14	14	–4	4	–9	6	–42	18	–6	3
7	–21	11	–17	–7	2	14	–22	14	–11
13	–11	–4	22	–5	–1	35	–17	–1	10
9	9	–9	1	–3	–3	31	3	–11	6
0	18	0	–20	–1	–4	12	18	–6	–8
–9	9	9	1	1	–4	–12	18	6	–8
–13	–11	4	22	3	–3	–31	3	11	6
–7	–21	–11	–17	5	–1	–35	–17	1	10
14	14	4	4	7	2	–14	–22	–14	–11
				9	6	42	18	6	3
990	2 002	468	1 980	330	132	8 580	2 860	780	660
$\frac{5}{6}$	$\frac{7}{12}$	$\frac{3}{20}$	$\frac{11}{60}$	2	$\frac{1}{2}$	$\frac{5}{3}$	$\frac{5}{12}$	$\frac{1}{10}$	$\frac{11}{240}$

Fuente: Adaptada con autorización de *Biometrika Tables for Statisticians*, vol. 1, tercera edición, 1996, por E. S. Pearson y H. O. Hartley, Cambridge University Press, Cambridge.

http://carlos2524.jimdo.com/

TABLA A.6 Valores críticos del estadístico Durbin-Watson

Tamaño de muestra	Probabilidad en la cola inferior (nivel de significancia = α)	k = Cantidad de regresores (excepto ordenada al origen)									
		1		2		3		4		5	
		d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
	.01	.81	1.07	.70	1.25	.59	1.46	.49	1.70	.39	1.96
15	.025	.95	1.23	.83	1.40	.71	1.61	.59	1.84	.48	2.09
	.05	1.08	1.36	.95	1.54	.82	1.75	.69	1.97	.56	2.21
	.01	.95	1.15	.86	1.27	.77	1.41	.63	1.57	.60	1.74
20	.025	1.08	1.28	.99	1.41	.89	1.55	.79	1.70	.70	1.87
	.05	1.20	1.41	1.10	1.54	1.00	1.68	.90	1.83	.79	1.99
	.01	1.05	1.21	.98	1.30	.90	1.41	.83	1.52	.75	1.65
25	.025	1.13	1.34	1.10	1.43	1.02	1.54	.94	1.65	.86	1.77
	.05	1.20	1.45	1.21	1.55	1.12	1.66	1.04	1.77	.95	1.89
	.01	1.13	1.26	1.07	1.34	1.01	1.42	.94	1.51	.88	1.61
30	.025	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	.98	1.73
	.05	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
	.01	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
40	.025	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
	.05	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
	.01	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
50	.025	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
	.05	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
	.01	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
60	.025	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
	.05	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
	.01	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
80	.025	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
	.05	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
	.01	1.52	1.56	1.50	1.58	1.48	1.60	1.45	1.63	1.44	1.65
100	.025	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72
	.05	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Fuente: Adaptada de "Testing for Serial Correlation in Least Squares Regression II", por J. Durbin y G. S. Watson, *Biometrika,* vol. 38, 1951, con autorización del editor.

http://carlos2524.jimdo.com/

APÉNDICE B

CONJUNTOS DE DATOS PARA LOS EJERCICIOS

Tabla B.1 Desempeño de los equipos de la Liga Nacional de Futbol en 1976
Tabla B.2 Datos de pruebas de energía térmica solar
Tabla B.3 Rendimiento de gasolina para 32 automóviles
Tabla B.4 Datos de avalúo de propiedades
Tabla B.5 Corridas de licuefacción en Belle Ayr
Tabla B.6 Datos de reactor de flujo tubular
Tabla B.7 Datos de extracción de aceite de cacahuate
Tabla B.8 Datos de formación de catratos
Tabla B.9 Datos de caída de presión
Tabla B.10 Datos de viscosidad cinemática
Tabla B.11 Datos de calidad de vino
Tabla B.12 Datos de tratamiento térmico
Tabla B.13 Datos de empuje de motor de turbina de reacción
Tabla B.14 Datos de inversor electrónico

http://carlos2524.jimdo.com/

TABLA B.1 Desempeño de los equipos de la Liga Nacional de Futbol en 1976

Equipo	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Washington	10	2113	1985	38.9	64.7	+4	868	59.7	2205	1917
Minnesota	11	2003	2855	38.8	61.3	+3	615	55.0	2096	1575
New England	11	2957	1737	40.1	60.0	+14	914	65.6	1847	2175
Oakland	13	2285	2905	41.6	45.3	–4	957	61.4	1903	2476
Pittsburgh	10	2971	1666	39.2	53.8	+15	836	66.1	1457	1866
Baltimore	11	2309	2927	39.7	74.1	+8	786	61.0	1848	2339
Los Ángeles	10	2528	2341	38.1	65.4	+12	754	66.1	1564	2092
Dallas	11	2147	2737	37.0	78.3	–1	761	58.0	1821	1909
Atlanta	4	1689	1414	42.1	47.6	–3	714	57.0	2577	2001
Buffalo	2	2566	1838	42.3	54.2	–1	797	58.9	2476	2254
Chicago	7	2363	1480	37.3	48.0	+19	984	67.5	1984	2217
Cincinnati	10	2109	2191	39.5	51.9	+6	700	57.2	1917	1758
Cleveland	9	2295	2229	37.4	53.6	–5	1037	58.8	1761	2032
Denver	9	1932	2204	35.1	71.4	+3	986	58.6	1709	2025
Detroit	6	2213	2140	38.8	58.3	+6	819	59.2	1901	1686
Green Bay	5	1722	1730	36.6	52.6	–19	791	54.4	2288	1835
Houston	5	1498	2072	35.3	59.3	–5	776	49.6	2072	1914
Kansas City	5	1873	2929	41.1	55.3	+10	789	54.3	2861	2496
Miami	6	2118	2268	38.2	69.6	+6	582	58.7	2411	2670
Nueva Orleans	4	1775	1983	39.3	78.3	+7	901	51.7	2289	2202
Nueva York Giants	3	1904	1792	39.7	38.1	–9	734	61.9	2203	1988
Nueva York Jets	3	1929	1606	39.7	68.8	–21	627	52.7	2592	2324
Philadelphia	4	2080	1492	35.5	68.8	–8	722	57.8	2053	2550
St. Louis	10	2301	2835	35.3	74.1	+2	683	59.7	1979	2110
San Diego	6	2040	2416	38.7	50.0	0	576	54.9	2048	2628
San Francisco	8	2447	1638	39.9	57.1	–8	848	65.3	1786	1776
Seattle	2	1416	2649	37.4	56.3	–22	684	43.8	2876	2524
Tampa Bay	0	1503	1503	39.3	47.0	–9	875	53.5	2560	2241

y: Juegos ganados (por temporada de 14 juegos)
x_1: Yardas por tierra (temporada)
x_2: Yardas por aire (temporada)
x_3: Promedio de pateo (yardas/patada)
x_4: Porcentaje de goles de campo (GC hechos/GC intentados, temporada)
x_5: Diferencia de pérdidas de balón (pérdidas ganadas/pérdidas perdidas)
x_6: Yardas de castigo (temporada)
x_7: Porcentaje de carreras (jugadas por tierra/jugadas totales)
x_8: Yardas por tierra del contrario (temporada)
x_9: Yardas por aire del contrario (temporada)

http://carlos2524.jimdo.com/

TABLA B.2 Datos de pruebas de energía solar térmica

y	x_1	x_2	x_3	x_4	x_5
271.8	783.35	33.53	40.55	16.66	13.20
264.0	748.45	36.50	36.19	16.46	14.11
238.8	684.45	34.66	37.31	17.66	15.68
230.7	827.80	33.13	32.52	17.50	10.53
251.6	860.45	35.75	33.71	16.40	11.00
257.9	875.15	34.46	34.14	16.28	11.31
263.9	909.45	34.60	34.85	16.06	11.96
266.5	905.55	35.38	35.89	15.93	12.58
229.1	756.00	35.85	33.53	16.60	10.66
239.3	769.35	35.68	33.79	16.41	10.85
258.0	793.50	35.35	34.72	16.17	11.41
257.6	801.65	35.04	35.22	15.92	11.91
267.3	819.65	34.07	36.50	16.04	12.85
267.0	808.55	32.20	37.60	16.19	13.58
259.6	774.95	34.32	37.89	16.62	14.21
240.4	711.85	31.08	37.71	17.37	15.56
227.2	694.85	35.73	37.00	18.12	15.83
196.0	638.10	34.11	36.76	18.53	16.41
278.7	774.55	34.79	34.62	15.54	13.10
272.3	757.90	35.77	35.40	15.70	13.63
267.4	753.35	36.44	35.96	16.45	14.51
254.5	704.70	37.82	36.26	17.62	15.38
224.7	666.80	35.07	36.34	18.12	16.10
181.5	568.55	35.26	35.90	19.05	16.73
227.5	653.10	35.56	31.84	16.51	10.58
253.6	704.05	35.73	33.16	16.02	11.28
263.0	709.60	36.46	33.83	15.89	11.91
265.8	726.90	36.26	34.89	15.83	12.65
263.8	697.15	37.20	36.27	16.71	14.06

y: Flujo total de calor (kwatts)
x_1: Insolación (watts/m^2)
x_2: Posición del foco en dirección este (pulgadas)
x_3: Posición del foco en dirección sur (pulgadas)
x_4: Posición del foco en dirección norte (pulgadas)
x_5: Hora del día

http://carlos2524.jimdo.com/

TABLA B.3 Rendimiento de la gasolina para 32 automóviles

Automóvil	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
Apollo	18.90	350	165	260	8.0:1	2.56:1	4	3	200.3	69.9	3910	A
Omega	17.00	350	170	275	8.5:1	2.56:1	4	3	199.6	72.9	2860	A
Nova	20.00	250	105	185	8.25:1	2.73:1	1	3	196.7	72.2	3510	A
Monarch	18.25	351	143	255	8.0:1	3.00:1	2	3	199.9	74.0	3890	A
Duster	20.07	225	95	170	8.4:1	2.76:1	1	3	194.1	71.8	3365	M
Jenson Conv.	11.2	440	215	330	8.2:1	2.88:1	4	3	184.5	69	4215	A
Skyhawk	22.12	231	110	175	8.0:1	2.56:1	2	3	179.3	65.4	3020	A
Monza	21.47	262	110	200	8.5:1	2.56:1	2	3	179.3	65.4	3180	A
Scirocco	34.70	89.7	70	81	8.2:1	3.90:1	2	4	155.7	64	1905	M
Corolla SR-5	30.40	96.9	75	83	9.0:1	4.30:1	2	5	165.2	65	2320	M
Camaro	16.50	350	155	250	8.5:1	3.08:1	4	3	195.4	74.4	3885	A
Datsun B210	36.50	85.3	80	83	8.5:1	3.89:1	2	4	160.6	62.2	2009	M
Capri II	21.50	171	109	146	8.2:1	3.22:1	2	4	170.4	66.9	2655	M
Pacer	19.70	258	110	195	8.0:1	3.08:1	1	3	171.5	77	3375	A
Babcat	20.30	140	83	109	8.4:1	3.40:1	2	4	168.8	69.4	2700	M
Granada	17.80	302	129	220	8.0:1	3.0:1	2	3	199.9	74	3890	A
Eldorado	14.39	500	190	360	8.5:1	2.73:1	4	3	224.1	79.8	5290	A
Imperial	14.89	440	215	330	8.2:1	2.71:1	4	3	231.0	79.7	5185	A
Nova LN	17.80	350	155	250	8.5:1	3.08:1	4	3	196.7	72.2	3910	A
Valiant	16.41	318	145	255	8.5:1	2.45:1	2	3	197.6	71	3660	A
Starfire	23.54	231	110	175	8.0:1	2.56:1	2	3	179.3	65.4	3050	A
Cordoba	21.47	360	180	290	8.4:1	2.45:1	2	3	214.2	76.3	4250	A
Trans AM	16.59	400	185	NA	7.6:1	3.08:1	4	3	196	73	3850	A
Corolla E-5	31.90	96.9	75	83	9.0:1	4.30:1	2	5	165.2	61.8	2275	M
Astre	29.40	140	86	NA	8.0:1	2.92:1	2	4	176.4	65.4	2150	M
Mark IV	13.27	460	223	366	8.0:1	3.00:1	4	3	228	79.8	5430	A
Celica GT	23.90	133.6	96	120	8.4:1	3.91:1	2	5	171.5	63.4	2535	M
Charger SE	19.73	318	140	255	8.5:1	2.71:1	2	3	215.3	76.3	4370	A
Cougar	13.90	351	148	243	8.0:1	3.25:1	2	3	215.5	78.5	4540	A
Elite	13.27	351	148	243	8.0:1	3.26:1	2	3	216.1	78.5	4715	A
Matador	13.77	360	195	295	8.25:1	3.15:1	4	3	209.3	77.4	4215	A
Corvette	16.50	350	165	255	8.5:1	2.73:1	4	3	185.2	69	3660	A

y: Millas/galón
x_1: Cilindrada (pulgadas cúbicas)
x_2: Potencia (Hp)
x_3: Par de torsión (pies-lb)
x_4: Relación de compresión
x_5: Relación de eje trasero
x_6: Carburador (gargantas)
x_7: Número de velocidades en la transmisión
x_8: Longitud total (pulgadas)
x_9: Ancho (pulgadas)
x_{10}: Peso (lb)
x_{11}: Tipo de transmisión (A = automática, M = manual)

Fuente: Motor Trend, 1975

http://carlos2524.jimdo.com/

TABLA B.4 Datos de avalúo de propiedades

y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
25.9	4.9176	1.0	3.4720	0.9980	1.0	7	4	42	0
29.5	5.0208	1.0	3.5310	1.5000	2.0	7	4	62	0
27.9	4.5429	1.0	2.2750	1.1750	1.0	6	3	40	0
25.9	4.5573	1.0	4.0500	1.2320	1.0	6	3	54	0
29.9	5.0597	1.0	4.4550	1.1210	1.0	6	3	42	0
29.9	3.8910	1.0	4.4550	0.9880	1.0	6	3	56	0
30.9	5.8980	1.0	5.8500	1.2400	1.0	7	3	51	1
28.9	5.6039	1.0	9.5200	1.5010	0.0	6	3	32	0
35.9	5.8282	1.0	6.4350	1.2250	2.0	6	3	32	0
31.5	5.3003	1.0	4.9883	1.5520	1.0	6	3	30	0
31.0	6.2712	1.0	5.5200	0.9750	1.0	5	2	30	0
30.9	5.9592	1.0	6.6660	1.1210	2.0	6	3	32	0
30.0	5.0500	1.0	5.0000	1.0200	0.0	5	2	46	1
36.9	8.2464	1.5	5.1500	1.6640	2.0	8	4	50	0
41.9	6.6969	1.5	6.9020	1.4880	1.5	7	3	22	1
40.5	7.7841	1.5	7.1020	1.3760	1.0	6	3	17	0
43.9	9.0384	1.0	7.8000	1.5000	1.5	7	3	23	0
37.5	5.9894	1.0	5.5200	1.2560	2.0	6	3	40	1
37.9	7.5422	1.5	5.0000	1.6900	1.0	6	3	22	0
44.5	8.7951	1.5	9.8900	1.8200	2.0	8	4	50	1
37.9	6.0831	1.5	6.7265	1.6520	1.0	6	3	44	0
38.9	8.3607	1.5	9.1500	1.7770	2.0	8	4	48	1
36.9	8.1400	1.0	8.0000	1.5040	2.0	7	3	3	0
45.8	9.1416	1.5	7.3262	1.8310	1.5	8	4	31	0

y: Precio de venta de la casa/1000
x_1: Impuestos (locales, escuela, municipal)/1000
x_2: Cantidad de baños
x_3: Tamaño del terreno (pies cuadrados $\times$ 1000)
x_4: Superficie construida (pies cuadrados $\times$ 1000)
x_5: Cantidad de cajones en cochera
x_6: Cantidad de habitaciones
x_7: Cantidad de recámaras
x_8: Edad de la casa (años)
x_9: Cantidad de chimeneas

Fuente: "Prediction, Linear Regression and Minimum Sum of Relative Errors", por S. C. Narula y J. F. Wellington, *Technometrics*, **19**, 1977. También véase "Carta al Editor", *Technometrics*, **22**, 1980.

http://carlos2524.jimdo.com/

TABLA B.5 Corridas de licuefacción en Belle Ayr

Corrida núm.	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	36.98	5.1	400	51.37	4.24	1484.83	2227.25	2.06
2	13.74	26.4	400	72.33	30.87	289.94	434.90	1.33
3	10.08	23.8	400	71.44	33.01	320.79	481.19	0.97
4	8.53	46.4	400	79.15	44.61	164.76	247.14	0.62
5	36.42	7.0	450	80.47	33.84	1097.26	1645.89	0.22
6	26.59	12.6	450	89.90	41.26	605.06	907.59	0.76
7	19.07	18.9	450	91.48	41.88	405.37	608.05	1.71
8	5.96	30.2	450	98.6	70.79	253.70	380.55	3.93
9	15.52	53.8	450	98.05	66.82	142.27	213.40	1.97
10	56.61	5.6	400	55.69	8.92	1362.24	2043.36	5.08
11	26.72	15.1	400	66.29	17.98	507.65	761.48	0.60
12	20.80	20.3	400	58.94	17.79	377.60	566.40	0.90
13	6.99	48.4	400	74.74	33.94	158.05	237.08	0.63
14	45.93	5.8	425	63.71	11.95	130.66	1961.49	2.04
15	43.09	11.2	425	67.14	14.73	682.59	1023.89	1.57
16	15.79	27.9	425	77.65	34.49	274.20	411.30	2.38
17	21.60	5.1	450	67.22	14.48	1496.51	2244.77	0.32
18	35.19	11.7	450	81.48	29.69	652.43	978.64	0.44
19	26.14	16.7	450	83.88	26.33	458.42	687.62	8.82
20	8.60	24.8	450	89.38	37.98	312.25	468.28	0.02
21	11.63	24.9	450	79.77	25.66	307.08	460.62	1.72
22	9.59	39.5	450	87.93	22.36	193.61	290.42	1.88
23	4.42	29.0	450	79.50	21.52	155.96	233.95	1.43
24	38.89	5.5	460	72.73	17.86	1392.08	2088.12	1.35
25	11.19	11.5	450	77.88	25.20	663.09	994.63	1.61
26	75.62	5.2	470	75.50	8.66	1464.11	2196.17	4.78
27	36.03	10.6	470	83.15	22.39	720.07	1080.11	5.88

y: CO_2
x_1: Tiempo de expansión, min
x_2: Temperatura, °C
x_3: Por ciento de solvatación
x_4: Rendimiento de aceite (g/100 g MAF)
x_5: Carbón total
x_6: Solvente total
x_7: Consumo de hidrógeno

Fuente: "Belle Ayr Liquefaction Runs with Solvent", *Industrial Chemical Process Design Development*, **17**, núm. 3, 1978.

http://carlos2524.jimdo.com/

TABLA B.6 Datos de reactor de flujo tubular

Corrida núm.	y	x_1	x_2	x_3	x_4
1	0.000450	0.0105	90.9	0.0164	0.0177
2	0.000450	0.0110	84.6	0.0165	0.0172
3	0.000473	0.0106	88.9	0.0164	0.0157
4	0.000507	0.0116	488.7	0.0187	0.0082
5	0.000457	0.0121	454.4	0.0187	0.0070
6	0.000452	0.0123	439.2	0.0187	0.0065
7	0.000453	0.0122	447.1	0.0186	0.0071
8	0.000426	0.0122	451.6	0.0187	0.0062
9	0.001215	0.0123	487.8	0.0192	0.0153
10	0.001256	0.0122	467.6	0.0192	0.0129
11	0.001145	0.0094	95.4	0.0163	0.0354
12	0.001085	0.0100	87.1	0.0162	0.0342
13	0.001066	0.0101	82.7	0.0162	0.0323
14	0.001111	0.0099	87.0	0.0163	0.0337
15	0.001364	0.0110	516.4	0.0190	0.0161
16	0.001254	0.0117	488.0	0.0189	0.0149
17	0.001396	0.0110	534.5	0.0189	0.0163
18	0.001575	0.0104	542.3	0.0189	0.0164
19	0.001615	0.0067	98.8	0.0163	0.0379
20	0.001733	0.0066	84.8	0.0162	0.0360
21	0.002753	0.0044	69.6	0.0163	0.0327
22	0.003186	0.0073	436.9	0.0189	0.0263
23	0.003227	0.0078	406.3	0.0192	0.0200
24	0.003469	0.0067	447.9	0.0192	0.0197
25	0.001911	0.0091	58.5	0.0164	0.0331
26	0.002588	0.0079	394.3	0.0177	0.0674
27	0.002635	0.0068	461.0	0.0174	0.0770
28	0.002725	0.0065	469.2	0.0173	0.0780

y: Concentración de $NbOCl_3$ (g-mol/l)
x_1: Concentración de $COCl_2$ (g-mol/l)
x_2: Tiempo de expansión (seg)
x_3: Densidad molar (g-mol/l)
x_4: Fracción mol de CO_2

Fuente: "Kinetics of Chlornation of Niobium Oxychloride by Phosgene in a Tube-Flow Reactor", *Industrial and Engineering Chemistry, Process Design Development*, **11**, núm. 2, 1972.

http://carlos2524.jimdo.com/

TABLA B.7 Datos de extracción de aceite de cacahuate

Presión (bar)	Temp. (°C)	Humedad (% en peso)	Flujo (L/min)	Tamaño de partícula (mm)	Rendimiento
415	25	5	40	1.28	63
550	25	5	40	4.05	21
415	95	5	40	4.05	36
550	95	5	40	1.28	99
415	25	15	40	4.05	24
550	25	15	40	1.28	66
415	95	15	40	1.28	71
550	95	15	40	4.05	54
415	25	5	60	4.05	23
550	25	5	60	1.28	74
415	95	5	60	1.28	80
550	95	5	60	4.05	33
415	25	15	60	1.28	63
550	25	15	60	4.05	21
415	95	15	60	4.05	44
550	95	15	60	1.28	96

Fuente: "An Application of Fractional Experimental Design", por M. B. Kilgo, *Quality Engineering*, **1**, pp. 19-23.

TABLA B.8 Datos de formación de catratos

x_1	x_2	y	x_1	x_2	y
0	10	7.5	0.02	30	19
0	50	15	0.02	60	26.4
0	85	22	0.02	90	28.5
0	110	28.6	0.02	120	29
0	140	31.6	0.02	210	35
0	170	34	0.02	30	15.1
0	200	35	0.02	60	26.4
0	230	35.5	0.02	120	27
0	260	36.5	0.02	150	29
0	290	38.5	0.05	20	21
0	10	12.3	0.05	40	27.3
0	30	18	0.05	130	48.5
0	62	20.8	0.05	190	50.4
0	90	25.7	0.05	250	52.5
0	150	32.5	0.05	60	34.4
0	210	34	0.05	90	46.5
0	270	35	0.05	120	50
0.02	10	14.4	0.05	150	51.9

y: Formación de catrato (% en masa)
x_1: Cantidad de surfactante (% en masa)
x_2: Tiempo (minutos)

Fuente: "Study on a Cool Storage System Using HCFC (Hydro-Chloro-Fluoro-Carbon)-14 lb (1, 1-dichloro-1-fluoro-ethane) Clathrate", por T. Tanii, M. Minemoto, K. Nakazawa y Y. Ando. *Canadian Journal of Chemical Engineering*, **75**, pp. 353-360.

http://carlos2524.jimdo.com/

TABLA B.9 Datos de caída de presión

x_1	x_2	x_3	x_4	y
2.14	10	0.34	1	28.9
4.14	10	0.34	1	31
8.15	10	0.34	1	26.4
2.14	10	0.34	0.246	27.2
4.14	10	0.34	0.379	26.1
8.15	10	0.34	0.474	23.2
2.14	10	0.34	0.141	19.7
4.14	10	0.34	0.234	22.1
8.15	10	0.34	0.311	22.8
2.14	10	0.34	0.076	29.2
4.14	10	0.34	0.132	23.6
8.15	10	0.34	0.184	23.6
2.14	2.63	0.34	0.679	24.2
4.14	2.63	0.34	0.804	22.1
8.15	2.63	0.34	0.89	20.9
2.14	2.63	0.34	0.514	17.6
4.14	2.63	0.34	0.672	15.7
8.15	2.63	0.34	0.801	15.8
2.14	2.63	0.34	0.346	14
4.14	2.63	0.34	0.506	17.1
8.15	2.63	0.34	0.669	18.3
2.14	2.63	0.34	1	33.8
4.14	2.63	0.34	1	31.7
8.15	2.63	0.34	1	28.1
5.6	1.25	0.34	0.848	18.1
5.6	1.25	0.34	0.737	16.5
5.6	1.25	0.34	0.651	15.4
5.6	1.25	0.34	0.554	15
4.3	2.63	0.34	0.748	19.1
4.3	2.63	0.34	0.682	16.2
4.3	2.63	0.34	0.524	16.3
4.3	2.63	0.34	0.472	15.8
4.3	2.63	0.34	0.398	15.4
5.6	10.1	0.25	0.789	19.2
5.6	10.1	0.25	0.677	8.4
5.6	10.1	0.25	0.59	15
5.6	10.1	0.25	0.523	12
5.6	10.1	0.34	0.789	21.9

http://carlos2524.jimdo.com/

TABLA B.9 (*Continuación*)

x_1	x_2	x_3	x_4	y
5.6	10.1	0.34	0.677	21.3
5.6	10.1	0.34	0.59	21.6
5.6	10.1	0.34	0.523	19.8
4.3	10.1	0.34	0.741	21.6
4.3	10.1	0.34	0.617	17.3
4.3	10.1	0.34	0.524	20
4.3	10.1	0.34	0.457	18.6
2.4	10.1	0.34	0.615	22.1
2.4	10.1	0.34	0.473	14.7
2.4	10.1	0.34	0.381	15.8
2.4	10.1	0.34	0.32	13.2
5.6	10.1	0.55	0.789	30.8
5.6	10.1	0.55	0.677	27.5
5.6	10.1	0.55	0.59	25.2
5.6	10.1	0.55	0.523	22.8
2.14	112	0.34	0.68	41.7
4.14	112	0.34	0.803	33.7
8.15	112	0.34	0.889	29.7
2.14	112	0.34	0.514	41.8
4.14	112	0.34	0.672	37.1
8.15	112	0.34	0.801	40.1
2.14	112	0.34	0.306	42.7
4.14	112	0.34	0.506	48.6
8.15	112	0.34	0.668	42.4

y: Factor adimensional para la caída de presión a través de un plato de burbujeo
x_1: Velocidad superficial del gas (cm/s)
x_2: Viscosidad cinemática
x_3: Abertura de la malla (cm)
x_4: Número adimensional que relaciona la velocidad superficial del gas a la velocidad superficial del líquido.

Fuente: "A Correlation of Two-Phase Pressure Drops in Screen-plate Bubble Column", por C. H. Liu, M. Kan y B. H. Chen, *Canadian Journal of Chemical Engineering*, **71**, pp. 460-463.

http://carlos2524.jimdo.com/

TABLA B.10 Datos de viscosidad cinemática

x_1	x_2	y
0.9189	–10	3.128
0.9189	0	2.427
0.9189	10	1.94
0.9189	20	1.586
0.9189	30	1.325
0.9189	40	1.126
0.9189	50	0.9694
0.9189	60	0.8473
0.9189	70	0.7481
0.9189	80	0.6671
0.7547	–10	2.27
0.7547	0	1.819
0.7547	10	1.489
0.7547	20	1.246
0.7547	30	1.062
0.7547	40	0.916
0.7547	50	0.8005
0.7547	60	0.7091
0.7547	70	0.6345
0.7547	80	0.5715
0.5685	–10	1.593
0.5685	0	1.324
0.5685	10	1.118
0.5685	20	0.9576
0.5685	30	0.8302
0.5685	40	0.7282
0.5685	50	0.647
0.5685	60	0.5784
0.5685	70	0.5219
0.5685	80	0.4735
0.361	–10	1.161
0.361	0	0.9925
0.361	10	0.8601
0.361	20	0.7523
0.361	30	0.6663
0.361	40	0.594
0.361	50	0.5338
0.361	60	0.4804
0.361	70	0.4361
0.361	80	0.4016

y: Viscosidad cinemática (10^{-6} m^2/s)
x_1: Relación de 2-metoxietanol a 1,2-dimetoxietano (adimensional)
x_2: Temperatura (°C)

Fuente: "Viscosimetric Studies on 2-Methoxyethanol + 1, 2-Dimethoxyethane Binary Mixtures from –10 to 80 °C", *Canadian Journal of Chemical Engineering*, **75**, pp. 494-501.

http://carlos2524.jimdo.com/

TABLA B.11 Datos de calidad de vino (determinados con MINITAB)

x_1 = claridad	x_2 = aroma	x_3 = cuerpo	x_4 = sabor	x_5 = fuerza	y = calidad	Región
1	3.3	2.8	3.1	4.1	9.8	1
1	4.4	4.9	3.5	3.9	12.6	1
1	3.9	5.3	4.8	4.7	11.9	1
1	3.9	2.6	3.1	3.6	11.1	1
1	5.6	5.1	5.5	5.1	13.3	1
1	4.6	4.7	5	4.1	12.8	1
1	4.8	4.8	4.8	3.3	12.8	1
1	5.3	4.5	4.3	5.2	12	1
1	4.3	4.3	3.9	2.9	13.6	3
1	4.3	3.9	4.7	3.9	13.9	1
1	5.1	4.3	4.5	3.6	14.4	3
0.5	3.3	5.4	4.3	3.6	12.3	2
0.8	5.9	5.7	7	4.1	16.1	3
0.7	7.7	6.6	6.7	3.7	16.1	3
1	7.1	4.4	5.8	4.1	15.5	3
0.9	5.5	5.6	5.6	4.4	15.5	3
1	6.3	5.4	4.8	4.6	13.8	3
1	5	5.5	5.5	4.1	13.8	3
1	4.6	4.1	4.3	3.1	11.3	1
0.9	3.4	5	3.4	3.4	7.9	2
0.9	6.4	5.4	6.6	4.8	15.1	3
1	5.5	5.3	5.3	3.8	13.5	3
0.7	4.7	4.1	5	3.7	10.8	2
0.7	4.1	4	4.1	4	9.5	2
1	6	5.4	5.7	4.7	12.7	3
1	4.3	4.6	4.7	4.9	11.6	2
1	3.9	4	5.1	5.1	11.7	1
1	5.1	4.9	5	5.1	11.9	2
1	3.9	4.4	5	4.4	10.8	2
1	4.5	3.7	2.9	3.9	8.5	2
1	5.2	4.3	5	6	10.7	2
0.8	4.2	3.8	3	4.7	9.1	1
1	3.3	3.5	4.3	4.5	12.1	1
1	6.8	5	6	5.2	14.9	3
0.8	5	5.7	5.5	4.8	13.5	1
0.8	3.5	4.7	4.2	3.3	12.2	1
0.8	4.3	5.5	3.5	5.8	10.3	1
0.8	5.2	4.8	5.7	3.5	13.2	1

La clase de este vino es Pinot Noir. Región se refiere a regiones geográficas distintas.

http://carlos2524.jimdo.com/

TABLA B.12 Datos de tratamiento térmico

TEMP	SOAKTIME	SOAKPCT	DIFFTIME	DIFFPCT	PITCH
1650	0.58	1.10	0.25	0.90	0.013
1650	0.66	1.10	0.33	0.90	0.016
1650	0.66	1.10	0.33	0.90	0.015
1650	0.66	1.10	0.33	0.95	0.016
1600	0.66	1.15	0.33	1.00	0.015
1600	0.66	1.15	0.33	1.00	0.016
1650	1.00	1.10	0.50	0.80	0.014
1650	1.17	1.10	0.58	0.80	0.021
1650	1.17	1.10	0.58	0.80	0.018
1650	1.17	1.10	0.58	0.80	0.019
1650	1.17	1.10	0.58	0.90	0.021
1650	1.17	1.10	0.58	0.90	0.019
1650	1.17	1.15	0.58	0.90	0.021
1650	1.20	1.15	1.10	0.80	0.025
1650	2.00	1.15	1.00	0.80	0.025
1650	2.00	1.10	1.10	0.80	0.026
1650	2.20	1.10	1.10	0.80	0.024
1650	2.20	1.10	1.10	0.80	0.025
1650	2.20	1.15	1.10	0.80	0.024
1650	2.20	1.10	1.10	0.90	0.025
1650	2.20	1.10	1.10	0.90	0.027
1650	2.20	1.10	1.50	0.90	0.026
1650	3.00	1.15	1.50	0.80	0.029
1650	3.00	1.10	1.50	0.70	0.030
1650	3.00	1.10	1.50	0.75	0.028
1650	3.00	1.15	1.66	0.85	0.032
1650	3.33	1.10	1.50	0.80	0.033
1700	4.00	1.10	1.50	0.70	0.039
1650	4.00	1.10	1.50	0.70	0.040
1650	4.00	1.15	1.50	0.85	0.035
1700	12.50	1.00	1.50	0.70	0.056
1700	18.50	1.00	1.50	0.70	0.068

y = PITCH: Resultados de la prueba de análisis de residuo de carbón
TEMP: Temperatura del horno
SOAKTIME: Duración del ciclo de cementación
SOAKPCT: Concentración de carbón
DIFFTIME: Duración del ciclo de difusión
DIFFPCT: Concentración de carbón en el ciclo de difusión

http://carlos2524.jimdo.com/

TABLA B.13 Datos de empuje de un motor de turbina de reacción

Observación número	y	x_1	x_2	x_3	x_4	x_5	x_6
1	4540	2140	20640	30250	205	1732	99
2	4315	2016	20280	30010	195	1697	100
3	4095	1905	19860	29780	184	1662	97
4	3650	1675	18980	29330	164	1598	97
5	3200	1474	18100	28960	144	1541	97
6	4833	2239	20740	30083	216	1709	87
7	4617	2120	20305	29831	206	1669	87
8	4340	1990	19961	29604	196	1640	87
9	3820	1702	18916	29088	171	1572	85
10	3368	1487	18012	28675	149	1522	85
11	4445	2107	20520	30120	195	1740	101
12	4188	1973	20130	29920	190	1711	100
13	3981	1864	19780	29720	180	1682	100
14	3622	1674	19020	29370	161	1630	100
15	3125	1440	18030	28940	139	1572	101
16	4560	2165	20680	30160	208	1704	98
17	4340	2048	20340	29960	199	1679	96
18	4115	1916	19860	29710	187	1642	94
19	3630	1658	18950	29250	164	1576	94
20	3210	1489	18700	28890	145	1528	94
21	4330	2062	20500	30190	193	1748	101
22	4119	1929	20050	29960	183	1713	100
23	3891	1815	19680	29770	173	1684	100
24	3467	1595	18890	29360	153	1624	99
25	3045	1400	17870	28960	134	1569	100
26	4411	2047	20540	30160	193	1746	99
27	4203	1935	20160	29940	184	1714	99
28	3968	1807	19750	29760	173	1679	99
29	3531	1591	18890	29350	153	1621	99
30	3074	1388	17870	28910	133	1561	99
31	4350	2071	20460	30180	198	1729	102
32	4128	1944	20010	29940	186	1692	101
33	3940	1831	19640	29750	178	1667	101
34	3480	1612	18710	29360	156	1609	101
35	3064	1410	17780	28900	136	1552	101
36	4402	2066	20520	30170	197	1758	100
37	4180	1954	20150	29950	188	1729	99
38	3973	1835	19750	29740	178	1690	99
39	3530	1616	18850	29320	156	1616	99
40	3080	1407	17910	28910	137	1569	100

y: Empuje
x_1: Velocidad primaria de rotación
x_2: Velocidad secundaria de rotación
x_3: Gasto del combustible
x_4: Presión
x_5: Temperatura de escape
x_6: Temperatura ambiente durante la prueba

http://carlos2524.jimdo.com/

TABLA B.14 Datos de inversor electrónico

Observación número	x_1	x_2	x_3	x_4	x_5	y
1	3	3	3	3	0	0.787
2	8	30	8	8	0	0.293
3	3	6	6	6	0	1.710
4	4	4	4	12	0	0.203
5	8	7	6	5	0	0.806
6	10	20	5	5	0	4.713
7	8	6	3	3	25	0.607
8	6	24	4	4	25	9.107
9	4	10	12	4	25	9.210
10	16	12	8	4	25	1.365
11	3	10	8	8	25	4.554
12	8	3	3	3	25	0.293
13	3	6	3	3	50	2.252
14	3	8	8	3	50	9.167
15	4	8	4	8	50	0.694
16	5	2	2	2	50	0.379
17	2	2	2	3	50	0.485
18	10	15	3	3	50	3.345
19	15	6	2	3	50	0.208
20	15	6	2	3	75	0.201
21	10	4	3	3	75	0.329
22	3	8	2	2	75	4.966
23	6	6	6	4	75	1.362
24	2	3	8	6	75	1.515
25	3	3	8	8	75	0.751

y: Punto transitorio (volts) de inversores PMOS-NMOS
x_1: Ancho del dispositivo NMOS
x_2: Longitud del dispositivo NMOS
x_3: Ancho del dispositivo PMOS
x_4: Longitud del dispositivo PMOS

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

APÉNDICE C

MATERIAL SUPLEMENTARIO

C.1 Resumen de pruebas básicas en estadística
C.2 Resumen de la teoría de modelos lineales
C.3 Resultados importantes acerca de SS_R y SS_{Res}
C.4 El teorema de Gauss-Markov, $\text{Var}(\varepsilon) = \sigma^2 \mathbf{I}$
C.5 Aspectos computacionales de la regresión múltiple
C.6 Un resultado acerca de la inversa de una matriz
C.7 Desarrollo del estadístico PRESS
C.8 Desarrollo de $S^2_{(i)}$
C.9 Una prueba para valores atípicos basada en R de Student
C.10 El teorema de Gauss-Markov, $\text{Var}(\varepsilon) = \mathbf{V}$
C.11 El sesgo en MS_{Res} cuando el modelo está subespecificado
C.12 Cálculo de diagnósticos de influencia
C.13 Modelos lineales generalizados

C.1 RESUMEN DE PRUEBAS BÁSICAS EN ESTADÍSTICA

Se indicará que Y es una variable aleatoria que sigue una distribución normal con media μ y varianza σ^2, con la notación

$$Y \sim N(\mu, \sigma^2)$$

C.1.1 Distribuciones centrales

1. Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes normalmente distribuidas, tales que $E(Y_i) = \mu_i$ y $\text{Var}(Y_i) = \sigma_i^2$. Sean $a_1, a_2, \ldots, a_n$ constantes conocidas. Si se define la combinación lineal de las Y_i con

$$U = \sum_{i=1}^{n} a_i Y_i$$

entonces

$$U \sim N\left(\sum_{i=1}^{n} a_i \mu, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

El asunto fundamental es que las combinaciones lineales de variables aleatorias normalmente distribuidas también siguen distribuciones normales.

http://carlos2524.jimdo.com/

2. Si $Y \sim N(\mu, \sigma^2)$, entonces

$$Z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

Z se llama variable aleatoria normal estándar.

3. Sea $Z = (Y - \mu)/\sigma$. Si $Y \sim N(\mu, \sigma^2)$, entonces Z^2 sigue una distribución χ^2, que se representará como sigue:

$$Z^2 \sim \chi_1^2$$

El punto clave es que el cuadrado de una variable aleatoria normal estándar es una variable aleatoria χ^2 con un grado de libertad.

4. Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes normalmente distribuidas, tales que $E(Y_i) = \mu_i$ y $\text{Var}(Y_i) = \sigma_i^2$, y sean

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i}$$

Entonces

$$\sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

Los puntos clave son: 1) la suma de los cuadrados de n variables normales estándar tiene una distribución χ^2 con n grados de libertad, y 2) la suma de variables aleatorias χ^2 también tiene una distribución χ^2.

5. ***El teorema del límite central*** Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes e idénticamente distribuidas, tales que $E(Y_i) = \mu$ y $\text{Var}(Y_i) = \sigma^2 < \infty$, entonces

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

converge en su distribución hacia una distribución normal estándar, cuando $n \to \infty$. El punto clave es que si n es suficientemente grande, entonces $\bar{Y}$ tiene aproximadamente una distribución normal. Lo que es "suficientemente grande" depende de la distribución básica de las Y_i.

6. Si $Z \sim N(0, 1)$, $V \sim \chi_\nu^2$ y Z y V son independientes, entonces

$$\frac{Z}{\sqrt{V/\nu}} \sim t_\nu$$

en donde t_ν es la distribución t con ν grados de libertad.

7. Sean $V \sim \chi_\nu^2$ y $W \sim \chi_\eta^2$. Si V y W son independientes, entonces

$$\frac{V/\nu}{W/\eta} \sim F_{\nu,\eta}$$

http://carlos2524.jimdo.com/

en donde $F_{\nu,\eta}$ es la distribución F con ν y η grados de libertad. Esto quiere decir que la razón de dos variables independientes χ^2, dividida cada una entre sus respectivos grados de libertad, sigue una distribución F.

C.1.2 Distribuciones no centrales

1. Sean $X \sim N(\delta, 1)$ y $V \sim \chi^2_\nu$. Si X y V son independientes, entonces

$$\frac{X}{\sqrt{V/\nu}} \sim t'_{\nu,\delta}$$

donde $t'_{\nu,\delta}$ es la distribución t no central con ν grados de libertad y parámetro δ de no centralidad.

2. Si $X \sim N(\delta, 1)$, entonces

$$X^2 \sim \chi^{2\prime}_{1,\delta^2}$$

donde $\chi^{2\prime}_{1,\delta^2}$ es la distribución χ^2 no central con un grado de libertad y parámetro δ^2 de no centralidad.

3. Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes y normalmente distribuidas, tales que $E(X_i) = \delta_i$ y $\mathrm{Var}(X_i) = 1$, entonces

$$\sum_{i=1}^{n} X_i^2 \sim \chi^{2\prime}_{n,\lambda}$$

en donde el parámetro λ de no centralidad es

$$\lambda = \sum_{i=1}^{n} \delta_i^2$$

4. Sean $V \sim \chi^{2\prime}_{\nu,\lambda}$ y $W \sim \chi^2_\nu$. Si V y W son independientes, entonces

$$\frac{V/\nu}{W/\eta} \sim F'_{\nu,\eta,\lambda}$$

en donde $F'_{\nu,\eta,\lambda}$ es una distribución F no central con ν y η grados de libertad, y parámetro de no centralidad λ.

C.2 RESUMEN DE LA TEORÍA DE MODELOS LINEALES

C.2.1 Definiciones básicas

1. ***Rango de una matriz*** El rango de una matriz **A** es la cantidad de columnas linealmente independientes. También es la cantidad de renglones linealmente independientes.

http://carlos2524.jimdo.com/

2. ***Matriz identidad*** La matriz identidad de orden k se representa por $\mathbf{I}$ o por $\mathbf{I}_k$, y es una matriz cuadrada de $k \times k$, cuyos elementos diagonales son 1, y cuyos elementos no diagonales son ceros. Es decir,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \cdots & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

3. ***Inversa de una matriz*** Sea $\mathbf{A}$ una matriz de $k \times k$. La inversa de $\mathbf{A}$ se representa por $\mathbf{A}^{-1}$ y es otra matriz de $k \times k$ tal que

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Si la inversa existe, es única.

4. ***Transpuesta de una matriz*** Sea $\mathbf{A}$ una matriz de $n \times k$. La transpuesta de $\mathbf{A}$ se representa por $\mathbf{A}'$ o por $\mathbf{A}^T$, y es una matriz de $k \times n$ cuyas columnas son los renglones de $\mathbf{A}$; así, si

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix}, \quad \text{cuando } \mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & & & \\ a_{1k} & a_{2k} & \cdots & a_{nk} \end{bmatrix}$$

Nota: Si $\mathbf{A}$ es una matriz de $n \times m$ y $\mathbf{B}$ es una matriz de $m \times p$, entonces

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

5. ***Matriz simétrica*** Sea $\mathbf{A}$ una matriz de $k \times k$. Se dice que $\mathbf{A}$ es simétrica si $\mathbf{A} = \mathbf{A}'$.
6. ***Matriz idempotente*** Sea $\mathbf{A}$ una matriz de $k \times k$. Se dice que $\mathbf{A}$ es idempotente si

$$\mathbf{A} = \mathbf{AA}$$

Si además $\mathbf{A}$ es simétrica, entonces se dice que es simétrica idempotente. Si $\mathbf{A}$ es simétrica idempotente, entonces $\mathbf{I} - \mathbf{A}$ también es simétrica idempotente.

7. ***Matriz ortogonal*** Sea $\mathbf{A}$ una matriz de $k \times k$. Si $\mathbf{A}$ es una matriz ortogonal, entonces $\mathbf{A}'\mathbf{A} = \mathbf{I}$. En consecuencia, si $\mathbf{A}$ es una matriz ortogonal, $\mathbf{A}^{-1} = \mathbf{A}'$.
8. ***Forma cuadrática*** Sea $\mathbf{y}$ un vector de $k \times 1$, y sea $\mathbf{A}$ una matriz de $k \times k$. La función

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} y_i y_j$$

se llama forma cuadrática. $\mathbf{A}$ es la matriz de la forma cuadrática.

http://carlos2524.jimdo.com/

9. ***Matrices positiva definida y positiva semidefinida*** Sea **A** una matriz de $k \times k$. Se dice que **A** es positiva definida si se cumplen las siguientes condiciones:

   a) $\mathbf{A} = \mathbf{A}'$ (**A** es simétrica), y

   b) $\mathbf{y}'\mathbf{A}\mathbf{y} > 0 \ \forall \ \mathbf{y} \in \Re^k, \mathbf{y} \neq \mathbf{0}$

   Se dice que **A** es positiva semidefinida si se cumple la siguiente condición:

   c) $\mathbf{y}'\mathbf{A}\mathbf{y} = 0$ para cierta $\mathbf{y} \neq \mathbf{0}$.

10. ***Traza de una matriz*** Sea **A** una matriz de $k \times k$. La traza de **A**, representada por traza(**A**) o tr(**A**), es la suma de los elementos diagonales de **A**, es decir,

$$\text{traza}(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}$$

   Nota:

   a) Si **A** es una matriz de $m \times n$ y **B** es una matriz de $n \times m$, entonces

$$\text{traza}(\mathbf{AB}) = \text{traza}\ (\mathbf{BA})$$

   b) Si las matrices son adecuadamente conformables, entonces

$$\text{traza}(\mathbf{ABC}) = \text{traza}(\mathbf{CAB})$$

   c) Si **A** y **B** son matrices de $k \times k$, y a y b son escalares, entonces

$$\text{traza}(a\mathbf{A} + b\mathbf{B}) = a\ \text{traza}(\mathbf{A}) + b\ \text{traza}(\mathbf{B})$$

11. ***Rango de una matriz idempotente*** Sea **A** una matriz idempotente. El rango de **A** es su traza.

12. ***Identidad importante para una matriz particionada*** Sea **X** una matriz de $n \times p$, particionada de tal modo que

$$\mathbf{X} = [\mathbf{X}_1\mathbf{X}_2]$$

   Se observa que

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}$$

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{X}_1\ \mathbf{X}_2] = \mathbf{X}$$

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{X}_1\ \mathbf{X}_2] = [\mathbf{X}_1\ \mathbf{X}_2]$$

   En consecuencia,

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1 = \mathbf{X}_1 \text{ y}$$

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_2 = \mathbf{X}_2$$

   Similarmente,

$$\mathbf{X}'_1\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}'_1 \text{ y}$$

$$\mathbf{X}'_2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_2 = \mathbf{X}'_2$$

http://carlos2524.jimdo.com/

C.2.2 Derivadas de matrices

Sean **A** una matriz de $k \times k$ de constantes, **a** un vector de $k \times 1$ de constantes y **y** un vector de $k \times 1$ de variables.

1. Si $\mathbf{z} = \mathbf{a}'\mathbf{y}$, entonces

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = \frac{\partial \mathbf{a}'\mathbf{y}}{\partial \mathbf{y}} = \mathbf{a}$$

2. Si $\mathbf{z} = \mathbf{y}'\mathbf{y}$, entonces

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = \frac{\partial \mathbf{y}'\mathbf{y}}{\partial \mathbf{y}} = 2\mathbf{y}$$

3. Si $\mathbf{z} = \mathbf{a}'\mathbf{A}\mathbf{y}$, entonces

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = \frac{\partial \mathbf{a}'\mathbf{A}\mathbf{y}}{\partial \mathbf{y}} = \mathbf{A}'\mathbf{a}$$

4. Si $\mathbf{z} = \mathbf{y}'\mathbf{A}\mathbf{y}$, entonces

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = \frac{\partial \mathbf{y}'\mathbf{A}\mathbf{y}}{\partial \mathbf{y}} = \mathbf{A}\mathbf{y} + \mathbf{A}'\mathbf{y}$$

Si **A** es simétrica, entonces

$$\frac{\partial \mathbf{y}'\mathbf{A}\mathbf{y}}{\partial \mathbf{y}} = 2\mathbf{A}\mathbf{y}$$

C.2.3 Valores esperados

Sean **A** una matriz de $k \times k$ de constantes, **a** un vector de $k \times 1$ de constantes, y **y** un vector aleatorio de $k \times 1$, con media $\boldsymbol{\mu}$ y matriz no singular **V** de varianza-covarianza.

1. $E(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\boldsymbol{\mu}$.
2. $E(\mathbf{A}\mathbf{y}) = \mathbf{A}\boldsymbol{\mu}$.
3. $\mathrm{Var}(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\mathbf{V}\mathbf{a}$.
4. $\mathrm{Var}(\mathbf{A}\mathbf{y}) = \mathbf{A}\mathbf{V}\mathbf{A}'$.
   Nota: si $\mathbf{V} = \sigma^2\mathbf{I}$, entonces $\mathrm{Var}(\mathbf{A}\mathbf{y}) = \sigma^2\mathbf{A}\mathbf{A}'$.
5. $E(\mathbf{y}'\mathbf{A}\mathbf{y}) = \mathrm{traza}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.
   Nota: Si $\mathbf{V} = \sigma^2\mathbf{I}$, entonces $E(\mathbf{y}'\mathbf{A}\mathbf{y}) = \sigma^2\,\mathrm{traza}(\mathbf{A}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.

C.2.4 Teoría de la distribución

Sea **A** una matriz de $k \times k$ de constantes, y sea **y** un vector aleatorio normal multivariado de $k \times 1$, con media $\boldsymbol{\mu}$ y matriz no singular **V** de varianza-covarianza; entonces

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{V})$$

http://carlos2524.jimdo.com/

Sea U la forma cuadrática definida por $U = \mathbf{y}'\mathbf{A}\mathbf{y}$.

1. Si $\mathbf{AV}$ o $\mathbf{VA}$ es una matriz idempotente de rango p, entonces

$$U \sim \chi^{2\prime}_{p,\lambda}$$

donde $\lambda = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.

2. Sea $\mathbf{V} = \sigma^2\mathbf{I}$, lo cual es una suposición típica. Si $\mathbf{A}$ es idempotente y de rango p, entonces

$$\frac{U}{\sigma^2} \sim \chi^{2\prime}_{p,\lambda}$$

donde $\lambda = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}/\sigma^2$.

3. Sea $\mathbf{B}$ una matriz de $q \times k$ y $\mathbf{W}$ la forma lineal definida por $\mathbf{W} = \mathbf{By}$. La forma cuadrática $U = \mathbf{y}'\mathbf{A}\mathbf{y}$ y $\mathbf{W}$ son independientes si

$$\mathbf{BVA} = \mathbf{0}$$

Nota: Si $\mathbf{V} = \sigma^2\mathbf{I}$, entonces U y $\mathbf{W}$ son independientes si $\mathbf{BA} = \mathbf{0}$.

4. Sea $\mathbf{B}$ una matriz de $k \times k$. Sea $V = \mathbf{y}'\mathbf{B}\mathbf{y}$. Las dos formas cuadráticas, U y V son independientes si

$$\mathbf{AVB} = \mathbf{0}$$

Nota: si $\mathbf{V} = \sigma^2\mathbf{I}$, entonces U y V son independientes si $\mathbf{AB} = \mathbf{0}$.

C.3 RESULTADOS IMPORTANTES ACERCA DE SS_{R} Y SS_{Res}

C.3.1 SS_{R}

Por definición,

$$SS_{\mathrm{R}} = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2$$

Se observa que $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, y que

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\mathbf{1}'\mathbf{y}$$

donde $\mathbf{1}$ es un vector de $n \times 1$, con todos sus elementos iguales a uno. Además, $n = \mathbf{1}'\mathbf{1}$, y en consecuencia $\bar{y} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y}$. Por consiguiente, se puede escribir SS_{R} en la forma

$$SS_{\mathrm{R}} = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2$$

http://carlos2524.jimdo.com/

$$
\begin{aligned}
&= [\hat{\mathbf{y}} - \mathbf{1}\bar{y}]'[\hat{\mathbf{y}} - \mathbf{1}\bar{y}] \\
&= \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y}\right]'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y}\right] \\
&= \mathbf{y}'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\right]'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\right]\mathbf{y}
\end{aligned}
$$

Nótese que $\mathbf{X} = [\mathbf{1}\ \mathbf{X}_R]$, siendo $\mathbf{X}_R$ la matriz formada por los valores reales de los regresores, por lo que SS_R es un caso especial de una matriz dividida. Por lo anterior, se puede usar la identidad especial para matrices divididas, para demostrar que

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{1} = \mathbf{1}, \qquad \text{y} \qquad \mathbf{1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{1}'$$

Por consiguiente, se puede demostrar que $[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']$ es idempotente. De acuerdo con la premisa que $\text{Var}(\varepsilon) = \sigma^2\mathbf{I}$,

$$\frac{SS_R}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{y}'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\right]\mathbf{y}$$

sigue una distribución χ^2 no central, con parámetro λ de no centralidad, y con grados de libertad iguales al rango de $[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']$. Como esta matriz es idempotente, su rango es igual a su traza. Se observa que

$$
\begin{aligned}
\text{Traza}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\right] &= \text{traza}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] - \text{traza}\left[\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\right] \\
&= \text{traza}\left[\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] - \text{traza}\left[\mathbf{1}'(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}\right] \\
&= \text{traza}\ (\mathbf{I}_p) - \text{traza}\ (1) \\
&= p - 1 = k
\end{aligned}
$$

Suponiendo que el modelo es correcto,

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} = [\mathbf{1}\ \boldsymbol{\beta}_R]\begin{bmatrix}\beta_0 \\ \boldsymbol{\beta}_R\end{bmatrix} = \beta_0\mathbf{1} + \mathbf{X}_R\boldsymbol{\beta}_R$$

Así, el parámetro de no centralidad es

$$
\begin{aligned}
\lambda &= \frac{1}{\sigma^2}E(\mathbf{y})'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\right]E(\mathbf{y}) \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\right]\mathbf{X}\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\begin{bmatrix}\beta_0 & \boldsymbol{\beta}_R'\end{bmatrix}\begin{bmatrix}\mathbf{1}' \\ \mathbf{X}_R'\end{bmatrix}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\right]\begin{bmatrix}\mathbf{1} & \mathbf{X}_R\end{bmatrix}\begin{bmatrix}\beta_0 \\ \boldsymbol{\beta}_R\end{bmatrix} \\
&= \frac{1}{\sigma^2}\begin{bmatrix}\beta_0 & \boldsymbol{\beta}_R'\end{bmatrix}\begin{bmatrix}\mathbf{1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' \\ \mathbf{X}_R'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}_R'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\end{bmatrix}\begin{bmatrix}\mathbf{1} & \mathbf{X}_R\end{bmatrix}\begin{bmatrix}\beta_0 \\ \boldsymbol{\beta}_R\end{bmatrix} \\
&= \frac{1}{\sigma^2}\begin{bmatrix}\beta_0 & \boldsymbol{\beta}_R'\end{bmatrix}\begin{bmatrix}\mathbf{0}' \\ \mathbf{X}_R' - \mathbf{X}_R'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\end{bmatrix}\begin{bmatrix}\mathbf{1} & \mathbf{X}_R\end{bmatrix}\begin{bmatrix}\beta_0 \\ \boldsymbol{\beta}_R\end{bmatrix}
\end{aligned}
$$

http://carlos2524.jimdo.com/

$$= \frac{1}{\sigma^2}\begin{bmatrix}\beta_0 & \boldsymbol{\beta}'_{\mathrm{R}}\end{bmatrix}\begin{bmatrix}0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{X}'_{\mathrm{R}}\mathbf{X}_{\mathrm{R}} - \mathbf{X}'_{\mathrm{R}}\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{X}_{\mathrm{R}}\end{bmatrix}\begin{bmatrix}\beta_0 \\ \boldsymbol{\beta}_{\mathrm{R}}\end{bmatrix}$$

$$= \frac{1}{\sigma^2}\boldsymbol{\beta}'_{\mathrm{R}}\left[\mathbf{X}'_{\mathrm{R}}\mathbf{X}_{\mathrm{R}} - \mathbf{X}'_{\mathrm{R}}\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{X}_{\mathrm{R}}\right]\boldsymbol{\beta}_{\mathrm{R}}$$

Si se define como sigue la matriz de los valores de los regresores **centrados** $\mathbf{X}_{\mathrm{C}}$:

$$\mathbf{X}_{\mathrm{C}} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1k} - \bar{x}_k \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2k} - \bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{nk} - \bar{x}_k \end{bmatrix}$$

siendo $\bar{x}_1$ el valor promedio del primer regresor, $\bar{x}_2$ el valor promedio del segundo regresor, y así sucesivamente, se puede demostrar con facilidad que el parámetro de no centralidad se expresa como sigue:

$$\lambda = \frac{1}{\sigma^2}\boldsymbol{\beta}'_{\mathrm{R}}[\mathbf{X}'_{\mathrm{C}}\mathbf{X}_{\mathrm{C}}]\boldsymbol{\beta}_{\mathrm{R}}$$

El valor esperado de SS_{R} es

$$\begin{aligned} E(SS_{\mathrm{R}}) &= E\left(\mathbf{y}'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1}\right]\mathbf{y}\right) \\ &= \text{traza}\left(\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1}\right]\sigma^2\mathbf{I}\right) \\ &\quad + E(\mathbf{y})'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1}\right]E(\mathbf{y}) \\ &= k\sigma^2 + \boldsymbol{\beta}'_{\mathrm{R}}\mathbf{X}'_{\mathrm{C}}\mathbf{X}_{\mathrm{C}}\boldsymbol{\beta}_{\mathrm{R}} \end{aligned}$$

Por consiguiente,

$$E(MS_{\mathrm{R}}) = E\left(\frac{SS_{\mathrm{R}}}{k}\right) = \sigma^2 + \frac{\boldsymbol{\beta}'_{\mathrm{R}}\mathbf{X}'_{\mathrm{C}}\mathbf{X}_{\mathrm{C}}\boldsymbol{\beta}_{\mathrm{R}}}{k}$$

C.3.2 SS_{Res}

Por definición,

$$SS_{\mathrm{Res}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Se ve que SS_{Res} se puede escribir en la forma siguiente:

$$\begin{aligned} SS_{\mathrm{Res}} &= (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) \\ &= \left[\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right]'\left[\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right] \\ &= \mathbf{y}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y} \end{aligned}$$

http://carlos2524.jimdo.com/

La demostración que $[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$ es simétrica e idempotente es trivial; en consecuencia

$$\frac{SS_{\text{Res}}}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{y}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y}$$

sigue una distribución χ^2. Los grados de libertad se deben al rango de $[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$, que a su vez es su traza. La demostración que la traza es $n - p$ es directa. Bajo la premisa que el modelo es correcto,

$$E(\mathbf{y}) = \mathbf{X}\beta$$

Así, el parámetro de no centralidad es

$$\begin{aligned}
&\frac{1}{\sigma^2}E(\mathbf{y})'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]E(\mathbf{y}) \\
&= \frac{1}{\sigma^2}\beta'\mathbf{X}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{X}\beta \\
&= \frac{1}{\sigma^2}\beta'\left[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\right]\beta = 0
\end{aligned}$$

Y como resultado,

$$\frac{SS_{\text{Res}}}{\sigma^2} \sim \chi^2_{n-p}$$

El valor esperado de SS_{Res} es

$$\begin{aligned}
E(SS_{\text{Res}}) &= E\left(\mathbf{y}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y}\right) \\
&= \text{traza}\left(\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\sigma^2\mathbf{I}\right) + E(\mathbf{y})'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]E(\mathbf{y}) \\
&= (n - p)\sigma^2
\end{aligned}$$

Y como resultado,

$$E(MS_{\text{Res}}) = E\left(\frac{SS_{\text{Res}}}{n - p}\right) = \sigma^2$$

C.3.3 La prueba *F* global o general

Un estadístico F es la razón de dos variables independientes χ^2, cada una dividida entre sus respectivos grados de libertad. Aquí se ha demostrado que tanto SS_{R}/σ^2 como SS_{Res}/σ^2 siguen distribuciones χ^2; lo fundamental ahora es demostrar que son independientes. De acuerdo con la teoría básica de los modelos lineales, SS_{R} y SS_{Res} son independientes, suponiendo que $\text{Var}(\varepsilon) = \sigma^2\mathbf{I}$, si

$$\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1}\right]\sigma^2\mathbf{I}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})'\mathbf{X}] = \mathbf{0}$$

http://carlos2524.jimdo.com/

Se observa que

$$\begin{aligned}
&\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1}\right]\sigma^2\mathbf{I}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})'\mathbf{X}] \\
&\quad = \sigma^2\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1}\right][\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})'\mathbf{X}] \\
&\quad = \sigma^2\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right. \\
&\qquad\qquad \left. + \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \\
&\quad = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1}' + \mathbf{1}(\mathbf{1}'\mathbf{1})'\mathbf{1}' = \mathbf{0}
\end{aligned}$$

Por consiguiente, SS_R y SS_{Res} son independientes. A continuación se observa que

$$\frac{SS_R}{k\sigma^2} = \frac{MS_R}{\sigma^2} \quad \text{y} \quad \frac{SS_{Res}}{(n-p)\sigma^2} = \frac{MS_{Res}}{\sigma^2}$$

son variables aleatorias χ^2, cada una dividida entre su grado de libertad respectivo. Entonces

$$\frac{MS_R}{MS_{Res}} \sim F'_{k,n-p,\lambda}$$

siendo

$$\lambda = \frac{1}{\sigma^2}\boldsymbol{\beta}'_R\mathbf{X}'_C\mathbf{X}_C\boldsymbol{\beta}_R$$

En el caso especial de las regresiones lineales simples sólo se tiene un regresor; entonces $\beta_R = \beta_1$ y

$$\mathbf{X}'_C\mathbf{X}_C = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Como resultado, para la regresión lineal simple,

$$\frac{MS_R}{MS_{Res}} \sim F_{1,n-2,\lambda}$$

siendo $\lambda = \beta_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2$.

C.3.4 El principio de la suma extra de cuadrados

Los SS_R son un caso especial del principio de la suma extra de cuadrados. Se tiene el modelo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

en el que $\mathbf{X}_1$ es la matriz de $p_1 \times 1$ del modelo, asociada con β_1, $\mathbf{X}_2$ es la matriz de $p_2 \times 1$ del modelo, asociada con β_2 y $p_1 + p_2 = p$. Una medida común de la contribución de β_2, dada la presencia de β_1 en el modelo, es

$$R(\beta_2 \mid \beta_1) = \mathbf{y}'[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1]\mathbf{y}$$

http://carlos2524.jimdo.com/

En el caso de SS_R, $\mathbf{X}_1 = \mathbf{1}$ y $\mathbf{X}_2 = \mathbf{X}_R$. En cierto sentido, $SS_R = R(\beta_R|\beta_0)$. Se puede demostrar que $[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1]$ es simétrica e idempotente. Las claves de esta deducción son que $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1 = \mathbf{X}_1$ y que $\mathbf{X}'_1\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}'_1$. Después, es directa la demostración que

$$\frac{R(\beta_2|\beta_1)}{\sigma^2} \sim \chi'^2_{p_2,\lambda}$$

siendo $\lambda = (1/\sigma^2)\beta_2'\mathbf{X}_2'[\mathbf{I} - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1]\mathbf{X}_2\beta_2$. Entonces se ve que

$$\frac{R(\beta_2|\beta_2)}{p_2 MS_{\text{Res}}} \sim F'_{p_2,n-p,\lambda}$$

C.4 EL TEOREMA DE GAUSS-MARKOV, $\text{Var}(\varepsilon) = \sigma^2\mathbf{I}$

El teorema de Gauss-Markov establece que el estimador de mínimos cuadrados ordinarios (OLS, de *ordinary least-squares*) de β, que es $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, es BLUE, el **mejor estimador lineal insesgado** (de *best linear unbiased estimator*). En este caso, mejor quiere decir que $\hat{\beta}$ tiene la mínima varianza, en cierto sentido significativo, entre la clase de todos los estimadores insesgados que son combinaciones lineales de los datos. Un problema es que $\hat{\beta}$ es un vector, y en consecuencia, su varianza es en realidad una matriz. Por ello se tratará de demostrar que $\hat{\beta}$ minimiza la varianza para cualquier combinación lineal de los coeficientes estimados, $\ell'\hat{\beta}$. Se observa que

$$\begin{aligned}\text{Var}(\ell'\hat{\beta}) &= \ell'\,\text{Var}(\hat{\beta})\,\ell \\ &= \ell'\left[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right]\ell \\ &= \sigma^2\ell'(\mathbf{X}'\mathbf{X})^{-1}\ell\end{aligned}$$

que es un escalar. Sea $\tilde{\beta}$ otro estimador insesgado de β que sea una combinación lineal de los datos. De lo que se trata es, entonces, demostrar que $\text{Var}(\ell'\hat{\beta}) \geq \sigma^2\ell'(\mathbf{X}'\mathbf{X})^{-1}\ell$ con al menos una ℓ tal que $\text{Var}(\ell'\hat{\beta}) > \sigma^2\ell'(\mathbf{X}'\mathbf{X})^{-1}\ell$.

Primero, se observa que se puede escribir cualquier otro estimador de β que sea una combinación lineal de los datos en la forma

$$\tilde{\beta} = \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]\mathbf{y} + \mathbf{b}_0$$

siendo $\mathbf{B}$ una matriz de $p \times n$, y $\mathbf{b}_0$ es un vector de $p \times 1$ de constantes, que ajusta en forma adecuada al estimador OLS para formar el estimado alternativo. A continuación se observa que si el modelo es correcto, entonces

http://carlos2524.jimdo.com/

$$
\begin{aligned}
E(\tilde{\boldsymbol{\beta}}) &= E\left(\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]\mathbf{y} + \mathbf{b}_0\right) \\
&= \left[(\mathbf{X}'\mathbf{X}^{-1}\mathbf{X}' + \mathbf{B}\right]E(\mathbf{y}) + \mathbf{b}_0 \\
&= \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]\mathbf{X}\boldsymbol{\beta} + \mathbf{b}_0 \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{B}\mathbf{X}\boldsymbol{\beta} + \mathbf{b}_0 \\
&= \boldsymbol{\beta} + \mathbf{B}\mathbf{X}\boldsymbol{\beta} + \mathbf{b}_0
\end{aligned}
$$

Por consiguiente, $\tilde{\boldsymbol{\beta}}$ es insesgado si, y sólo si tanto $\mathbf{b}_0 = \mathbf{0}$ como $\mathbf{BX} = \mathbf{0}$. La varianza de $\tilde{\boldsymbol{\beta}}$ es

$$
\begin{aligned}
\operatorname{Var}(\tilde{\boldsymbol{\beta}}) &= \operatorname{Var}\left(\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]\mathbf{y}\right) \\
&= \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]\operatorname{Var}(\mathbf{y})\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]' \\
&= \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]\sigma^2\mathbf{I}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]' \\
&= \sigma^2\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{B}'\right] \\
&= \sigma^2\left[(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{B}\mathbf{B}'\right]
\end{aligned}
$$

porque $\mathbf{BX} = \mathbf{0}$, lo cual a su vez implica que $(\mathbf{BX})' = \mathbf{X}'\mathbf{B}' = \mathbf{0}$. Como resultado,

$$
\begin{aligned}
\operatorname{Var}(\boldsymbol{\ell}'\tilde{\boldsymbol{\beta}}) &= \boldsymbol{\ell}'\operatorname{Var}(\tilde{\boldsymbol{\beta}})\boldsymbol{\ell} \\
&= \boldsymbol{\ell}'\left(\sigma^2\left[(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{B}\mathbf{B}'\right]\right)\boldsymbol{\ell} \\
&= \sigma^2\boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\ell} + \sigma^2\boldsymbol{\ell}'\mathbf{B}\mathbf{B}'\boldsymbol{\ell} \\
&= \operatorname{Var}(\boldsymbol{\ell}'\tilde{\boldsymbol{\beta}}) + \sigma^2\boldsymbol{\ell}'\mathbf{B}\mathbf{B}'\boldsymbol{\ell}
\end{aligned}
$$

Primero se observa que $\mathbf{BB}'$ es al menos una matriz positiva semidefinida, y por consiguiente $\sigma^2\boldsymbol{\ell}'\mathbf{BB}'\boldsymbol{\ell} \geq \mathbf{0}$. A continuación, se ve que se puede definir a $\boldsymbol{\ell}^* = \mathbf{B}'\boldsymbol{\ell}$. Por consiguiente,

$$
\boldsymbol{\ell}'\mathbf{BB}'\boldsymbol{\ell} = \boldsymbol{\ell}^{*\prime}\boldsymbol{\ell}^* = \sum_{i=1}^{p}\ell_i^{*2}
$$

que debe ser estrictamente mayor que 0 para cierta $\boldsymbol{\ell} \neq \mathbf{0}$, a menos que $\mathbf{B} = \mathbf{0}$. Por lo anterior, el estimado OLS de $\boldsymbol{\beta}$ es el mejor estimador lineal insesgado.

C.5 ASPECTOS COMPUTACIONALES DE LA REGRESIÓN MÚLTIPLE

En esta sección se describirá en forma breve un importante procedimiento computacional para resolver el problema de regresión por mínimos cuadrados. El criterio de mínimos cuadrados es

http://carlos2524.jimdo.com/

$$\min_{\beta} S(\beta) = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

se recordará de la sección 3.2.2 que el vector solución de mínimos cuadrados es normal al espacio p-dimensional de estimación. Como la norma euclidiana es invariante bajo una transformación ortogonal, una formulación equivalente del problema de mínimos cuadrados es

$$\min_{\beta} S(\beta) = (\mathbf{Qy} - \mathbf{QX}\beta)'(\mathbf{Qy} - \mathbf{QX}\beta) \tag{C.5.1}$$

siendo $\mathbf{Q}$ una matriz ortogonal de $n \times n$. Ahora bien, $\mathbf{Q}$ se puede escoger de tal manera que

$$\mathbf{QX} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}$$

siendo $\mathbf{R}$ una matriz triangular superior de $p \times p$ (es decir, una matriz con ceros abajo de la diagonal principal). Si se define a

$$\mathbf{Qy} = \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1'\mathbf{y} \\ \mathbf{Q}_2'\mathbf{y} \end{bmatrix}$$

siendo $\mathbf{Q}_1'$ una matriz de $p \times n$ formada por los primeros p renglones de $\mathbf{Q}$, entonces $\mathbf{Q}_2'$ es una matriz de $(n - p) \times n$, formada por los últimos $n - p$ renglones de $\mathbf{Q}$, y $\mathbf{q}_1$ es un vector columna, de $p \times 1$; entonces la solución de (C.5.1) satisface a

$$\mathbf{R}\,\hat{\beta} = \mathbf{q}_1 \tag{C.5.2}$$

o sea

$$\hat{\beta} = \mathbf{R}^{-1}\mathbf{q}_1 = \mathbf{R}^{-1}\mathbf{Q}_1'\mathbf{y} \tag{C.5.3}$$

Una ventaja de este método es que se puede obtener un inverso numéricamente estable de $\mathbf{R}$ con el método de sustitución en reversa. Para ilustrarlo, supóngase que $\mathbf{R}$ y $\mathbf{q}_1 = \mathbf{Q}_1'\mathbf{y}$ son, para $p = 3$,

$$\mathbf{R} = \begin{bmatrix} 3 & 1 & 2 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \qquad \mathbf{q}_1 = \mathbf{Q}_1'\mathbf{y} = \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix}$$

Las ecuaciones (C.5.2) son

$$\begin{bmatrix} 3 & 1 & 2 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix}$$

y el sistema de ecuaciones que realmente se debe resolver es

$$3\hat{\beta}_0 + 1\hat{\beta}_1 + 2\hat{\beta}_2 = 3$$

$$3\hat{\beta}_1 + 1\hat{\beta}_2 = -1$$

$$2\hat{\beta}_2 = 4$$

http://carlos2524.jimdo.com/

Según la última ecuación, $2\hat{\beta}_2 = 4$, o sea que $\hat{\beta}_2 = 2$. Al sustituir directamente en la segunda ecuación resulta $3\hat{\beta}_1 + 1(2) = -1$, o sea que $\hat{\beta}_1 = -1$. Por último, la primera ecuación es $3\hat{\beta}_0 + 1\hat{\beta}_1 + 2\beta_2 = 3$, y entonces $\hat{\beta}_0 = 0$.

Los algoritmos para calcular la descomposición **QR** se describen en Golub [1969], Lawson y Hanson [1974] y Seber [1977]. Hay disponibles muchas subrutinas para estos cálculos y para muchas clases de computadoras.

Se puede calcular en forma directa la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ a partir de la factorización **QR**. Ya que

$$\mathbf{QX} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}$$

entonces

$$\mathbf{X} = \mathbf{Q}'\begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} = \mathbf{Q}_1\mathbf{R}$$

En consecuencia, ya que $\mathbf{Q}_1'\mathbf{Q}_1 = \mathbf{I}$,

$$(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{R}'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{R})^{-1} = (\mathbf{R}'\mathbf{R})^{-1} = \mathbf{R}^{-1}(\mathbf{R}')^{-1} \tag{C.5.4}$$

Esta descomposición también conduce al cálculo eficiente de los elementos de la matriz de sombrero, que se vio es de utilidad en varios aspectos. Obsérvese que

$$\begin{aligned} \mathbf{H} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{Q}_1\mathbf{R}\mathbf{R}^{-1}(\mathbf{R}')^{-1}\mathbf{R}'\mathbf{Q}_1' \\ &= \mathbf{Q}_1\mathbf{Q}_1' \end{aligned} \tag{C.5.5}$$

Por lo anterior, los elementos de la diagonal principal en la matriz de sombrero se pueden formar con las sumas de cuadrados de los renglones de $\mathbf{Q}_1$. Entonces se pueden calcular con facilidad muchos estadísticos importantes de diagnóstico de regresión, como por ejemplo los residuales estudentizados y la medida de distancia de Cook. Belsley, Kuh y Welsch [1980] indican cómo se pueden calcular varios diagnósticos de regresión partiendo de estas ideas.

C.6 UN RESULTADO ACERCA DE LA INVERSA DE UNA MATRIZ

El resultado que se presentará en esta sección tiene utilidad en la obtención de la forma computacional del estadístico PRESS, y de varios diagnósticos de influencia. Considérese la matriz $\mathbf{X}'\mathbf{X}$, de $p \times p$, y sea $\mathbf{x}'$ el i-ésimo renglón de $\mathbf{X}$. Obsérvese que $\mathbf{X}'\mathbf{X} - \mathbf{x}\mathbf{x}'$ es la matriz $\mathbf{X}'\mathbf{X}$ con su i-ésimo renglón eliminado. El resultado es

$$(\mathbf{X}'\mathbf{X} - \mathbf{x}\mathbf{x}')^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}}{1 - \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}} \tag{C.6.1}$$

Se puede demostrar este resultado multiplicando el lado derecho por $\mathbf{X}'\mathbf{X} - \mathbf{x}\mathbf{x}'$, para llegar como sigue a una matriz identidad:

http://carlos2524.jimdo.com/

$$\left[(\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}}{1 - \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}\right](\mathbf{X}'\mathbf{X} - \mathbf{x}\mathbf{x}')$$

$$= \mathbf{I} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}'}{1 - \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}' - \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}'}{1 - \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}'$$

$$= \mathbf{I} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}'\left(1 - \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\right) - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\left[\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\right]\mathbf{x}'}{1 - \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$$

$$= \mathbf{I} - \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}'\left[\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\right] - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\mathbf{x}'\left[\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\right]}{1 - \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$$

$$= \mathbf{I}$$

El resultado de (C.6.1) también se puede escribir como sigue:

$$\left[\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right]^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}}{1 - h_{ii}} \tag{C.6.2}$$

porque $h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_1$, y $\mathbf{X}_{(i)}$ representa la matriz original $\mathbf{X}$ donde se conserva el i-ésimo renglón $\mathbf{x}_i$.

C.7 DESARROLLO DEL ESTADÍSTICO PRESS

Se usó el estadístico PRESS como medida de la validez del modelo de regresión, y de la eficiencia potencial de predicción. Recuérdese que $e_{(i)} = y_i - \hat{y}_{(i)}$ es el residual de PRESS, donde $\hat{y}_{(i)}$ es el valor predicho, obtenido con un ajuste de modelo conservando la i-ésima observación. Entonces

$$\text{PRESS} = \sum_{i=1}^{n} e_{(i)}^2 = \sum_{i=1}^{n} \left[y_i - \hat{y}_{(i)}\right]^2 \tag{C.7.1}$$

Al principio parecería que para calcular PRESS se requiere ajustar n regresiones distintas. Sin embargo, es posible calcular PRESS a partir de los resultados de un solo ajuste por mínimos cuadrados para todas las n observaciones. Para ver cómo se logra eso, sea $\hat{\boldsymbol{\beta}}_{(i)}$ el vector de los coeficientes de regresión, obtenido conservando la i-ésima observación. Entonces

$$\hat{\boldsymbol{\beta}}_{(i)} = \left[\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right]^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)} \tag{C.7.2}$$

en donde $\mathbf{X}_{(i)}$ y $\mathbf{y}_{(i)}$ son los vectores $\mathbf{X}$ y $\mathbf{y}$ donde se conserva la i-ésima observación.

Por consiguiente, el i-ésimo residual PRESS se puede escribir como sigue:

$$\begin{aligned} e_{(i)} &= y_i - \hat{y}_{(i)} \\ &= y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}}_{(i)} \\ &= y_i - \mathbf{x}'_i\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)} \end{aligned}$$

http://carlos2524.jimdo.com/

Hay una estrecha relación entre las matrices $(\mathbf{X}'\mathbf{X})^{-1}$ y $[\mathbf{X}_{(i)}\mathbf{X}_{(i)}]^{-1}$; en forma específica, según la ecuación (C.6.2) del apéndice C.6,

$$\left[\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right]^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}}{1 - h_{ii}} \tag{C.7.3}$$

en donde $h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$. De acuerdo con la ecuación anterior, se puede escribir

$$\begin{aligned}
e_{(i)} &= y_i - \mathbf{x}'_i\left[(\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}}{1 - h_{ii}}\right]\mathbf{X}'_{(i)}\mathbf{y}_{(i)} \\
&= y_i - \mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)} - \frac{\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)}}{1 - h_{ii}} \\
&= \frac{(1 - h_{ii})y_i - (1 - h_{ii})\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)} - h_{ii}\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)}}{1 - h_{ii}} \\
&= \frac{(1 - h_{ii})y_i - \mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)}}{1 - h_{ii}}
\end{aligned}$$

Como $\mathbf{X}'\mathbf{y} = \mathbf{X}'_{(i)}\mathbf{y}_{(i)} + \mathbf{x}_i y_i$, esta última ecuación se transforma como sigue:

$$\begin{aligned}
e_{(i)} &= \frac{(1 - h_{ii})\mathbf{y}_i - \mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y} - \mathbf{x}_i y_i)}{1 - h_{ii}} \\
&= \frac{(1 - h_{ii})y_i - \mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i y_i}{1 - h_{ii}} \\
&= \frac{(1 - h_{ii})y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}} + h_{ii}y_i}{1 - h_{ii}} \\
&= \frac{y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}}}{1 - h_{ii}}
\end{aligned} \tag{C.7.4}$$

Ahora bien, el numerador de la ecuación (C.7.4) es el residual ordinario e_i de un ajuste por mínimos cuadrados con todas las n observaciones, por lo que el i-ésimo residual PRESS es

$$e_{(i)} = \frac{e_i}{1 - h_{ii}} \tag{C.7.5}$$

Así, como PRESS no es más que la suma de cuadrados de los residuales PRESS, una fórmula sencilla de cómputo es

$$\text{PRESS} = \sum_{i=1}^{n}\left(\frac{e_i}{1 - h_{ii}}\right)^2 \tag{C.7.6}$$

http://carlos2524.jimdo.com/

En esta forma es fácil ver que PRESS es una suma ponderada de cuadrados de los residuales, donde los factores de ponderación o pesos se relacionan con la influencia de las observaciones. PRESS pondera los residuales correspondientes a observaciones de alta influencia con más severidad que los que proceden de puntos menos influyentes.

C.8 DESARROLLO DE $S_{(i)}^2$

En el capítulo 4 se presentó una ecuación del cuadrado medio de residuales en un modelo de regresión conservando la i-ésima observación. La cantidad que resulta es $S_{(i)}^2$ y se usa para calcular la R de Student. Esta fórmula de cómputo de $S_{(i)}^2$ se puede deducir partiendo de la identidad del apéndice C.6:

$$\left[\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right]^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}}{1 - h_{ii}}$$

Si se posmultiplican ambos lados por $\mathbf{X}'\mathbf{y} - \mathbf{x}_i y_i$, se obtiene

$$\hat{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i y_i + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y} - \mathbf{x}_i y_i)}{1 - h_{ii}}$$

que se reduce a

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)} = \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i e_i}{1 - h_{ii}} \tag{C.8.1}$$

Ahora bien,

$$(n - p - 1)S_{(i)}^2 = \sum_{j \neq i}\left(y_j - \mathbf{x}'_j\hat{\boldsymbol{\beta}}_{(i)}\right)^2 \tag{C.8.2}$$

y después de usar la ecuación (C.8.1), esto se transforma en

$$\begin{aligned}\sum_{j \neq i}\left(y_j - \mathbf{x}'_j\hat{\boldsymbol{\beta}}_{(i)}\right)^2 &= \sum_{j=1}^{n}\left(y_i - \mathbf{x}'_j\hat{\boldsymbol{\beta}} + \frac{\mathbf{x}'_j(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i e_i}{1 - h_{ii}}\right)^2 \\ &\quad - \left(y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}} + \frac{h_{ii}e_i}{1 - h_{ii}}\right)^2 \\ &= \sum_{j=1}^{n}\left(e_j + \frac{h_{ii}e_i}{1 - h_{ii}}\right)^2 - \frac{e_i^2}{(1 - h_{ii})^2}\end{aligned} \tag{C.8.3}$$

Cuando se desarrolla el primer término del lado derecho de la ecuación anterior, se obtiene

$$\sum_{j=1}^{n}\left(e_j + \frac{h_{ij}e_i}{1 - h_{ii}}\right)^2 = \sum_{j=1}^{n}e_j^2 + \frac{2e_i}{1 - h_{ii}}\sum_{j=1}^{n}e_j h_{ij} + \frac{e_i^2}{(1 - h_{ii})^2}\sum_{j=1}^{n}h_{ij}^2$$

http://carlos2524.jimdo.com/

Sin embargo, como $\mathbf{Hy} = \mathbf{H}\hat{\mathbf{y}}$, entonces $\Sigma_{j=1}^{n} e_j h_{ij} = 0$. También, como $\mathbf{H}$ es idempotente, $\Sigma_{j=1}^{n} h_{ij}^2 = h_{ii}$. Por lo anterior, la ecuación (C.8.2) se puede escribir en la forma siguiente:

$$
\begin{aligned}
(n-p-1)S_{(i)}^2 &= \sum_{j=1}^{n} e_j^2 + \frac{h_{ii}e_i^2}{(1-h_{ii})^2} - \frac{e_i^2}{(1-h_{ii})^2} \\
&= \sum_{j=1}^{n} e_j^2 - \frac{e_i^2}{1-h_{ii}} \\
&= (n-p)MS_{\text{Res}} - \frac{e_i^2}{1-h_{ii}}
\end{aligned}
$$

Por último, se llega al resultado de la ecuación (4.12), esto es,

$$
S_{(i)}^2 = \frac{(n-p)MS_{\text{Res}} - e_i^2/(1-h_{ii})}{n-p-1}
$$

C.9 UNA PRUEBA PARA VALORES ATÍPICOS BASADA EN *R* DE STUDENT

Una forma frecuente de modelar un valor atípico es con el **modelo del valor atípico con media desplazada**. Supóngase que se ajusta el modelo $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, cuando el modelo verdadero es

$$
\mathbf{y} = \mathbf{X}\beta + \delta + \varepsilon
$$

donde δ es un vector de $n \times 1$, de ceros excepto para la u-ésima observación, cuyo valor es δ_u. Así,

$$
\delta = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \delta_u \\ 0 \\ \vdots \\ 0 \end{bmatrix}
$$

Supóngase que, tanto para el modelo que se ajusta y el modelo del valor atípico con media desplazada, $E(\varepsilon) \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Se trata de determinar un estadístico adecuado de prueba para las hipótesis

$$
\begin{aligned}
H_0&: \delta_u = 0 \\
H_0&: \delta_u \neq 0
\end{aligned}
$$

http://carlos2524.jimdo.com/

En este procedimiento se supone que el interés específico es la u-ésima observación, es decir, que se cuenta con información *a priori* que la u-ésima observación puede ser un valor atípico.

El primer paso es determinar un estimado adecuado de δ_u. Un candidato lógico es el u-ésimo residual. Sea $\mathbf{e} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$ el vector de residuales de $n \times 1$. El valor esperado de $\mathbf{e}$ es

$$\begin{aligned}
E(\mathbf{e}) &= E\left(\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y}\right) \\
&= \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]E(\mathbf{y}) \\
&= \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right][\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}] \\
&= \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{X}\boldsymbol{\beta} + \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\boldsymbol{\delta} \\
&= [\mathbf{X} - \mathbf{X}]\boldsymbol{\beta} + \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\boldsymbol{\delta} \\
&= \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\boldsymbol{\delta}
\end{aligned}$$

Así,

$$E(e_u) = (1 - h_{uu})\delta_u$$

siendo h_{uu} la u-ésima diagonal de sombrero, o el u-ésimo elemento diagonal de $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. En consecuencia, un estimador insesgado de δ_u es

$$\hat{\delta}_u = \frac{e_u}{1 - h_{uu}}$$

En el capítulo 4 se demostró que $\hat{\delta}_u$ es tan sólo el u-ésimo residual PRESS. El siguiente paso es determinar la varianza de ese estimador. Obsérvese que

$$\begin{aligned}
\operatorname{Var}(\mathbf{e}) &= \operatorname{Var}\left(\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y}\right) \\
&= \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\sigma^2\mathbf{I}\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]' \\
&= \sigma^2\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \\
&= \sigma^2\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]
\end{aligned}$$

Así, $\operatorname{Var}(e_u) = (1 - h_{uu})\sigma^2$. Entonces, la varianza de $\hat{\delta}_u$ es

$$\begin{aligned}
\operatorname{Var}\left(\frac{e_u}{1 - h_{uu}}\right) &= \frac{1}{(1 - h_{uu})^2}\operatorname{Var}(e_u) \\
&= \frac{(1 - h_{uu})\sigma^2}{(1 - h_{uu})^2}
\end{aligned}$$

http://carlos2524.jimdo.com/

$$= \frac{\sigma^2}{1 - h_{uu}}$$

A continuación, se observará que **e** es una combinación lineal de **y**. Así, **e** es una combinación lineal de variables aleatorias normalmente distribuidas. Por lo anterior, **e** sigue una distribución normal, al igual que $\hat{\delta}_u$. Entonces, bajo la hipótesis H_0: $\delta_u = 0$,

$$\frac{e_u/(1 - h_{uu})}{\sigma/\left(\sqrt{1 - h_{uu}}\right)} = \frac{e_u}{\sigma\sqrt{1 - h_{uu}}}$$

sigue una distribución normal estándar. Se ve que esta cantidad no es más que un ejemplo de un residual estudentizado, como se vio en el capítulo 4. En general, se desconoce σ^2. Se vio que MS_{Res} es un estimador insesgado de σ^2. Además, se vio que

$$\frac{MS_{\text{Res}}}{\sigma^2}$$

es una variable aleatoria χ^2, dividida entre sus grados de libertad. En consecuencia, un estadístico probable de prueba es

$$\frac{e_u}{\sqrt{MS_{\text{Res}}(1 - h_{uu})}}$$

que sigue una distribución t si $\mathbf{e} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$, y $SS_{\text{Res}} = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$ son independientes. Se puede demostrar que **e** y SS_{Res} son independientes si

$$\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\sigma^2\mathbf{I}\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] = \mathbf{0}$$

Desafortunadamente,

$$\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\sigma^2\mathbf{I}\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] = \sigma^2\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \neq \mathbf{0}.$$

El problema es que

$$SS_{\text{Res}} = \mathbf{e}'\mathbf{e} = \sum_{i=1}^{n} e_i^2$$

y eso quiere decir que SS_{Res} está correlacionado con cada residual individual, porque el cuadrado de cada residual individual es un componente de SS_{Res}. En el apéndice C.8 se desarrolló un estimado de σ^2 omitiendo la u-ésima observación. Este estimado es independiente de e_u, por la hipótesis básica de independencia para los errores aleatorios. En consecuencia, un estadístico adecuado de prueba para el modelo del valor atípico con media desplazada es

$$\frac{e_u}{s_{(u)}\sqrt{1 - h_{uu}}}$$

http://carlos2524.jimdo.com/

que es el residual estudentizado externamente, o R de Student. De acuerdo con H_0: $\delta_u = 0$, este estadístico sigue la distribución central T_{n-p-1}, y de acuerdo con H_0: $\delta_u \neq 0$, sigue la distribución $t'_{n-p-1,\gamma}$ siendo

$$\gamma = \frac{\delta_u}{\sigma/\left(\sqrt{1-h_{uu}}\right)} = \frac{\delta_i\sqrt{1-h_{uu}}}{\sigma}$$

Es importante notar que la potencia de esta prueba depende de h_{uu}. Recuérdese que si se ajusta una ordenada al origen al modelo, entonces $1/n \leq h_{uu} \leq 1$. La potencia máxima se tiene cuando $h_{uu} = 1/n$, que está en el centro de la nube de datos, en términos de las $\mathbf{X}$. Cuando $h_{uu} \to 1$, la potencia baja a 0. En otras palabras, esta prueba tiene menos capacidad de detectar valores atípicos en los puntos de datos de alta influencia.

C.10 EL TEOREMA DE GAUSS-MARKOV, Var(ε) = V

El teorema de Gauss-Markov establece que el estimador de mínimos cuadrados generalizados (GLS, de *generalized least-squares*) de β, que es $\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ es BLUE (de *best linear unbiased estimator*), **mejor estimador lineal insesgado**. De nuevo, por mejor se quiere indicar que $\hat{\beta}$ minimiza la varianza para cualquier combinación lineal de los coeficientes estimados, $\ell'\hat{\beta}$. Se nota que si el modelo es correcto,

$$\begin{aligned} E\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right] &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}E(\mathbf{y}) \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \end{aligned}$$

Así, $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ es un estimador insesgado de β. La varianza de este estimador es

$$\begin{aligned} \mathrm{Var}\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right] &= \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]\mathrm{Var}(\mathbf{y})\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]' \\ &= \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]\mathbf{V}\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]' \\ &= \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]\mathbf{V}\left[\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\right] \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) \end{aligned}$$

Así,

$$\begin{aligned} \mathrm{Var}(\ell'\hat{\beta}) &= \ell'\,\mathrm{Var}(\hat{\beta})\,\ell \\ &= \ell'\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\right]\ell \end{aligned}$$

Sea $\tilde{\beta}$ otro estimador insesgado de β que sea una combinación lineal de los datos. De lo que se trata es, en consecuencia, de demostrar que $\mathrm{Var}(\ell'\tilde{\beta}) \geq \ell'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\ell$ con al menos una ℓ tal que $\mathrm{Var}(\ell'\tilde{\beta}) > \ell'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\ell$.

http://carlos2524.jimdo.com/

Primero se observa que se puede escribir cualquier otro estimador de β que sea combinación lineal de los datos en la forma

$$\tilde{\beta} = \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{B}\right]\mathbf{y} + \mathbf{b}_0$$

donde $\mathbf{B}$ es una matriz de $p \times n$, y $\mathbf{b}_0$ es un vector de constantes de $p \times 1$, que ajusta en forma adecuada al estimador GLS, para formar el estimado alternativo. A continuación se ve que si el modelo es correcto, entonces

$$\begin{aligned}
E(\tilde{\beta}) &= E\left(\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{B}\right]\mathbf{y} + \mathbf{b}_0\right) \\
&= \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{B}\right]E(\mathbf{y}) + \mathbf{b}_0 \\
&= \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{B}\right]\mathbf{X}\beta + \mathbf{b}_0 \\
&= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta + \mathbf{B}\mathbf{X}\beta + \mathbf{b}_0 \\
&= \beta + \mathbf{B}\mathbf{X}\beta + \mathbf{b}_0
\end{aligned}$$

Por consiguiente, $\tilde{\beta}$ es insesgado si, y sólo si tanto $\mathbf{b}_0 = \mathbf{0}$ como $\mathbf{BX} = \mathbf{0}$. La varianza de $\tilde{\beta}$ es

$$\begin{aligned}
\mathrm{Var}(\tilde{\beta}) &= \mathrm{Var}\left(\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{B}\right]\mathbf{y}\right) \\
&= \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{B}\right]\mathrm{Var}(\mathbf{y})\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{V}'\mathbf{V}^{-1} + \mathbf{B}\right]' \\
&= \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{B}\right]\mathbf{V}\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{B}\right]' \\
&= \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \mathbf{B}\right]\mathbf{V}\left[\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} + \mathbf{B}'\right] \\
&= \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} + \mathbf{B}\mathbf{V}\mathbf{B}'\right]
\end{aligned}$$

como $\mathbf{BX} = \mathbf{0}$, lo cual a su vez implica que $(\mathbf{BX})' = \mathbf{X}'\mathbf{B}' = \mathbf{0}$. El resultado es

$$\begin{aligned}
\mathrm{Var}(\ell'\hat{\beta}) &= \ell'\,\mathrm{Var}(\tilde{\beta})\,\ell \\
&= \ell'\left(\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} + \mathbf{B}\mathbf{V}\mathbf{B}'\right]\right)\ell \\
&= \ell'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\ell + \ell'\mathbf{B}\mathbf{V}\mathbf{B}'\ell \\
&= \mathrm{Var}(\ell'\hat{\beta}) + \ell'\mathbf{B}\mathbf{V}\mathbf{B}'\ell
\end{aligned}$$

Se ve que $\mathbf{V}$ es una matriz positiva definida. Por consiguiente, existe una matriz no singular, Γ, tal que $\mathbf{V} = \Gamma'\Gamma$. El resultado es que $\mathbf{BVB}' = \mathbf{B}\Gamma'\Gamma\mathbf{B}'$ es, cuando menos, una matriz positiva semidefinida. Por lo anterior, $\ell'\mathbf{BVB}'\ell \geq 0$. Después, obsérvese que se puede definir a $\ell^* = \Gamma\mathbf{B}'\ell$, y por tanto

$$\ell'\mathbf{BVB}'\ell = \ell^{*\prime}\ell^* = \sum_{i=1}^{p} \ell_i^{*2}$$

que debe ser estrictamente mayor que 0 para una $\ell \neq \mathbf{0}$, a menos que $\mathbf{B} = \mathbf{0}$. Por lo anterior, el estimado GLS de β es el mejor estimador lineal insesgado.

http://carlos2524.jimdo.com/

C.11 EL SESGO EN MS_{RES} CUANDO EL MODELO ESTÁ SUBESPECIFICADO

En el apéndice C.3 se demostró que si el modelo está especificado en forma correcta, entonces $E(MS_{\text{Res}}) = \sigma^2$, y por consiguiente MS_{Res} es un estimador insesgado de σ^2. Ahora supóngase que se ajusta el modelo $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$, siendo $\mathbf{X}_1$ la matriz de modelo de $n \times p$, asociada con $\boldsymbol{\beta}_1$, el vector de parámetros que se ajusta.

Supóngase además que el modelo real es

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} = [\mathbf{X}_1\ \mathbf{X}_2]\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} + \boldsymbol{\varepsilon}$$

en donde $\mathbf{X}_2$ es la matriz de modelo asociada con $\boldsymbol{\beta}_2$, el vector de términos importantes que no se ajustaron, sea cual fuere la razón. Para ambos modelos, supóngase que $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, y que $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$. En la sección 9.1.2 se demostró que, para este caso, $\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$ es un estimador sesgado de $\boldsymbol{\beta}_1$. Considérese el valor esperado de SS_{Res}, que es

$$\begin{aligned}
E(SS_{\text{Res}}) &= E\left(\mathbf{y}'\left[\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\right]\mathbf{y}\right)\\
&= \text{traza}\left(\left[\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\right]\sigma^2\mathbf{I}\right)\\
&\quad + [\boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2']\begin{bmatrix}\mathbf{X}_1'\\ \mathbf{X}_2'\end{bmatrix}\left[\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\right][\mathbf{X}_1\ \mathbf{X}_2]\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix}\\
&= \sigma^2\,\text{traza}\left[\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\right]\\
&\quad + [\boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2']\begin{bmatrix}\mathbf{X}_1' - \mathbf{X}_1'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\\ \mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\end{bmatrix}[\mathbf{X}_1\ \mathbf{X}_2]\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix}\\
&= (n-p)\sigma^2 + [\boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2']\begin{bmatrix}\mathbf{0}'\\ \mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\end{bmatrix}[\mathbf{X}_1\ \mathbf{X}_2]\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix}\\
&= (n-p)\sigma^2 + [\boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2']\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix}\\
&= (n-p)\sigma^2 + \boldsymbol{\beta}_2'\left[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\right]\boldsymbol{\beta}_2
\end{aligned}$$

El valor esperado de MS_{Res}, en este caso, es

$$\begin{aligned}
E(MS_{\text{Res}}) &= E\left(\frac{SS_{\text{Res}}}{n-p}\right)\\
&= \sigma^2 + \frac{\boldsymbol{\beta}_2'\left[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\right]\boldsymbol{\beta}_2}{n-p}
\end{aligned}$$

http://carlos2524.jimdo.com/

El resultado es que MS_{Res} no es un estimador insesgado de σ^2 cuando el modelo está subespecificado. El sesgo es

$$\frac{\boldsymbol{\beta}_2'\left[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\right]\boldsymbol{\beta}_2}{n-p}$$

C.12 CÁLCULO DE DIAGNÓSTICO DE INFLUENCIA

En esta sección se desarrollarán las formas computacionales, muy útiles, de los diagnósticos de influencia *DFFITS*, *DFBETAS* y la *D* de Cook, que se mencionaron por primera vez en el capítulo 6.

C.12.1 $DFFITS_i$

Recuérdese que, según la ecuación (6.9),

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{S_{(i)}^2 h_{ii}}}, \qquad i = 1, 2, \ldots, n \tag{C.12.1}$$

También, de acuerdo con el apéndice C.8,

$$\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{(i)} = \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\, e_i}{1 - h_{ii}} \tag{C.12.2}$$

Multiplicando ambos lados de la ecuación (C.12.2) por $\mathbf{x}_i'$ se obtiene

$$\hat{y}_i - \hat{y}_{(i)} = \frac{h_{ii}\, e_i}{1 - h_{ii}} \tag{C.12.3}$$

Ambos lados de esta ecuación se dividen entre $\sqrt{S_{(i)}^2 h_{ii}}$, para obtener $DFFITS_i$:

$$\begin{aligned} DFFITS_i &= \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{S_{(i)}^2 h_{ii}}} = \frac{h_{ii}\, e_i}{1 - h_{ii}}\left[\frac{1}{S_{(i)}^2 h_{ii}}\right]^{1/2} \\ &= \frac{e_i}{\sqrt{S_{(i)}^2\left(1 - h_{ii}\right)}}\left(\frac{h_{ii}}{1 - h_{ii}}\right)^{1/2} \\ &= t_i\left(\frac{h_{ii}}{1 - h_{ii}}\right)^{1/2} \end{aligned} \tag{C.12.4}$$

en donde t_i es *R* de Student.

http://carlos2524.jimdo.com/

C.12.2 D_i de Cook

Se usará la ecuación (C.12.2) para desarrollar una forma computacional para la D_i de Cook. Recuérdese que el estadístico D_i de Cook es

$$D_i = \frac{\left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{(i)}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{(i)}\right)}{pMS_{\text{Res}}}, \qquad i = 1, 2, \ldots, n \tag{C.12.5}$$

Se usa la ecuación (C.12.2) en la ecuación (C.12.5) para obtener

$$\begin{aligned} D_i &= \frac{\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\, e_i^2}{(1 - h_{ii})^2\, pMS_{\text{Res}}} \\ &= \left(\frac{e_i}{1 - h_{ii}}\right)^2\left(\frac{h_{ii}}{pMS_{\text{Res}}}\right) \\ &= \frac{r_i^2}{p}\left(\frac{h_{ii}}{1 - h_{ii}}\right) \end{aligned}$$

donde r_i es el residual estudentizado.

C.12.3 $DFBETAS_{j,\,i}$

El estadístico $DFBETAS_{j,\,i}$ se definió como sigue, en la ecuación (6.7):

$$DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 C_{jj}}}$$

Así, $DFBETAS_{j,\,i}$ no es más que el j-ésimo elemento de $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$ en la ecuación (C.12.2), dividido entre un factor de estandarización. Ahora bien,

$$\hat{\beta}_j - \hat{\beta}_{j(i)} = \frac{r_{j,i}\, e_i}{1 - h_{ii}} \tag{C.12.6}$$

y recuérdese que $\mathbf{R} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$, por lo que

$$(\mathbf{R}\mathbf{R}')' = \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right]' = (\mathbf{X}'\mathbf{X})^{-1} = \mathbf{C} = \mathbf{R}'\mathbf{R}$$

Por lo anterior, $C_{jj} = \mathbf{r}_j'\mathbf{r}_j$, y entonces se puede escribir el factor de estandarización como sigue:

$$\sqrt{S_{(i)}^2 C_{jj}} = \sqrt{S_{(i)}^2 \mathbf{r}_j'\mathbf{r}_j}$$

Por último, la forma computacional de $DFBETAS_{j,\,i}$ es

$$DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 C_{jj}}}$$

http://carlos2524.jimdo.com/

$$= \left[\frac{r_{j,i}\, e_i}{1 - h_{ii}}\right] \frac{1}{\sqrt{S_{(i)}^2 \mathbf{r}_j' \mathbf{r}_j}}$$

$$= \frac{\mathbf{r}_{j,i}}{\sqrt{\mathbf{r}_j' \mathbf{r}_j}} \frac{t_i}{\sqrt{1 - h_{ii}}}$$

en donde t_i es R de Student.

C.13 MODELOS LINEALES GENERALIZADOS

C.13.1 Estimación de parámetros en la regresión logística

El logaritmo de máxima verosimilitud para un modelo de regresión logística se presentó en la ecuación (13.8) como sigue:

$$\ln L(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \mathbf{x}_i' \boldsymbol{\beta} - \sum_{i=1}^{n} \ln[1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})]$$

En muchas aplicaciones de modelos de regresión logística se tienen intentos u observaciones repetidas en cada valor de las variables x. Sea y_i la cantidad de unos observada para la i-ésima observación, y sea n_i la cantidad de intentos en cada observación. Entonces, el logaritmo de máxima verosimilitud es

$$\ln L(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \pi_i + \sum_{i=1}^{n} n_i \ln(1 - \pi_i) - \sum_{i=1}^{n} y_i \ln(1 - \pi_i)$$

Los estimados de máxima verosimilitud (MLE, de *maximum likelihood estimator* o *estimate*) se pueden calcular con un algoritmo de mínimos cuadrados iterativamente reponderados (IRLS, de *iteratively reweighted least squares*). Para visualizarlo, recuérdese que los estimados MLE son las soluciones de

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = 0$$

que se puede expresar como sigue:

$$\frac{\partial L}{\partial \pi_i} \frac{\partial \pi_i}{\partial \boldsymbol{\beta}} = 0$$

Nótese que

$$\frac{\partial L}{\partial \pi_i} = \sum_{i=1}^{n} \frac{n_i}{\pi_i} - \sum_{i=1}^{n} \frac{n_i}{1 - \pi_i} + \sum_{i=1}^{n} \frac{y_i}{1 - \pi_i}$$

y que

$$\frac{\partial \pi_i}{\partial \boldsymbol{\beta}} = \left\{ \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})} - \left[\frac{\exp(\mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})}\right]^2 \right\} \mathbf{x}_i$$

http://carlos2524.jimdo.com/

Al reunir todo lo anterior se obtiene

$$\begin{aligned}\frac{\partial L}{\partial \boldsymbol{\beta}} &= \left[\sum_{i=1}^{n}\frac{n_i}{\pi_i} - \sum_{i=1}^{n}\frac{n_i}{1-\pi_i} + \sum_{i=1}^{n}\frac{y_i}{1-\pi_i}\right]\pi_i(1-\pi_i)\mathbf{x}_i \\ &= \sum_{i=1}^{n}\left[\frac{y_i}{\pi_i} - \frac{n_i}{1-\pi_i} + \frac{y_i}{1-\pi_i}\right]\pi_i(1-\pi_i)\mathbf{x}_i \\ &= \sum_{i=1}^{n}(y_i - n_i\pi_i)\mathbf{x}_i\end{aligned}$$

Por consiguiente, el estimador de máxima verosimilitud resuelve la ecuación

$$\mathbf{X}'(\mathbf{y}-\boldsymbol{\mu}) = \mathbf{0}$$

en donde $\mathbf{y}' = [y_1, y_2, \ldots, y_n]$ y $\boldsymbol{\mu}' = [n_1\pi_1, n_2\pi_2, \ldots, n_n\pi_n]$. Este conjunto de ecuaciones se llama con frecuencia **ecuaciones de puntuación de máxima verosimilitud**. Son en realidad de la misma forma de las ecuaciones normales que se vieron antes para los mínimos cuadrados lineales, porque en el modelo de regresión lineal, $E(y) = \mathbf{X}\beta = \boldsymbol{\mu}$, y las ecuaciones normales son

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$

que se pueden escribir como sigue:

$$\begin{aligned}\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= \mathbf{0} \\ \mathbf{X}'(\mathbf{y} - \boldsymbol{\mu}) &= \mathbf{0}\end{aligned}$$

En realidad, el método de **Newton-Raphson** es el que se usa para resolver las ecuaciones de puntuación para el modelo de regresión logística. En este procedimiento se observa que en la proximidad de la solución se puede usar un desarrollo en serie de Taylor de primer orden para formar la aproximación

$$p_i - \pi_i \approx \left(\frac{\partial \pi_i}{\partial \boldsymbol{\beta}}\right)'(\boldsymbol{\beta}^* - \boldsymbol{\beta}) \tag{C.13.1}$$

donde

$$p_i = \frac{y_i}{n_i}$$

y β^* es el valor de β que resuelve las ecuaciones de puntuación. Ahora bien, $\eta_i = \mathbf{x}_i'\beta$, y

$$\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \mathbf{x}_i$$

así que

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$$

http://carlos2524.jimdo.com/

Según la regla de la cadena,

$$\frac{\partial \pi_i}{\partial \boldsymbol{\beta}} = \frac{\partial \pi_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \frac{\partial \pi_i}{\partial \eta_i}\mathbf{x}_i$$

Por lo anterior, la ecuación (C.13.1) se puede escribir en la forma

$$\begin{aligned} p_i - \pi_i &\approx \left(\frac{\partial \pi_i}{\partial \eta_i}\right)\mathbf{x}_i'(\boldsymbol{\beta}^* - \boldsymbol{\beta}) \\ p_i - \pi_i &\approx \left(\frac{\partial \pi_i}{\partial \eta_i}\right)(\mathbf{x}_i'\boldsymbol{\beta}^* - \mathbf{x}_i'\boldsymbol{\beta}) \\ p_i - \pi_i &\approx \left(\frac{\partial \pi_i}{\partial \eta_i}\right)(\eta_i^* - \eta_i) \end{aligned} \tag{C.13.2}$$

siendo η_i^* el valor de η_i evaluado en β^*. Se nota que

$$(y_i - n_i\pi_i) = (n_i p_i - n_i\pi_i) = n_i(p_i - \pi_i)$$

y en vista de que

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$$

se puede escribir

$$\frac{\partial \pi_i}{\partial \eta_i} = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} - \left[\frac{\exp(\eta_i)}{1 + \exp(\eta_i)}\right]^2$$

En consecuencia,

$$y_i - n_i\pi_i \approx [n_i\pi_i(1 - \pi_i)](\eta_i^* - \eta_i)$$

Ahora bien, la varianza del predictor lineal $\eta_i^* = \mathbf{x}_i'\beta^*$ es, con una primera aproximación,

$$\operatorname{Var}(\eta_i^*) \approx \frac{1}{n_i\pi_i(1 - \pi_i)}$$

Por lo anterior,

$$y_i - n_i\pi_i \approx \left[\frac{1}{\operatorname{Var}(\eta_i^*)}\right](\eta_i^* - \eta_i)$$

y se pueden expresar las ecuaciones de puntuación en la forma

$$\sum_{i=1}^{n}\left[\frac{1}{\operatorname{Var}(\eta_i)}\right](\eta_i^* - \eta_i) = 0$$

http://carlos2524.jimdo.com/

o bien, en notación matricial,

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \eta) = \mathbf{0}$$

siendo $\mathbf{V}$ una matriz diagonal de los factores de ponderación obtenidos con las varianzas de las η_i. Como $\eta = \mathbf{X}\beta$, las ecuaciones de puntuación se pueden escribir como sigue:

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \mathbf{X}\beta) = \mathbf{0}$$

y el estimado de máxima verosimilitud de β es

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\eta^*$$

Sin embargo, se presenta el problema que no se conoce η^*. La solución de este problema emplea la ecuación (C.13.2):

$$p_i - \pi_i \approx \left(\frac{\partial \pi_i}{\partial \eta_i}\right)(\eta_i^* - \eta_i)$$

de donde se pueden despejar las η_i^*,

$$\eta_i^* \approx \eta_i + (p_i - \pi_i)\frac{\partial \eta_i}{\partial \pi_i}$$

Sean $z_i = \eta_i + (p_i - \pi_i)(\partial\eta_i)/(\partial\pi_i)$ y $\mathbf{z}' = [z_1, z_2, \ldots, z_n]$. Entonces, el estimado de Newton-Raphson de β es

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{z}$$

Nótese que la parte aleatoria de z_i es

$$(p_i - \pi_i)\frac{\partial \eta_i}{\partial \pi_i}$$

Así,

$$\begin{aligned}
\operatorname{Var}\left[(p_i - \pi_i)\frac{\partial \eta_i}{\partial \pi_i}\right] &= \left[\frac{\pi_i(1-\pi_i)}{n_i}\right]\left(\frac{\partial \eta_i}{\partial \pi_i}\right)^2 \\
&= \left[\frac{\pi_i(1-\pi_i)}{n_i}\right]\left(\frac{1}{\pi_i(1-\pi_i)}\right)^2 \\
&= \frac{1}{n_i\pi_i(1-\pi_i)}
\end{aligned}$$

Entonces, $\mathbf{V}$ es la matriz diagonal de los factores de ponderación obtenidos con las varianzas de la parte aleatoria de $\mathbf{z}$. Por consiguiente, el algoritmo IRLS basado en el método de Newton-Raphson se puede describir como sigue:

http://carlos2524.jimdo.com/

1. Usar mínimos cuadrados ordinarios para obtener un estimado inicial de β, por ejemplo $\hat{\beta}_0$;
2. Usar $\hat{\beta}_0$ para estimar $\mathbf{V}$ y π;
3. Definir $\eta_0 = \mathbf{X}\,\hat{\beta}_0$;
4. Basar $\mathbf{z}_1$ en η_0;
5. Obtener un nuevo estimado de $\hat{\beta}_1$ e iterar, hasta que se satisfaga un criterio adecuado de convergencia.

C.13.2 La familia exponencial

Es fácil demostrar que las distribuciones normal, binomial y de Poisson son miembros de la familia exponencial. Recuérdese que la familia exponencial de distribuciones se define con la ecuación (13.37), que repetimos a continuación para mayor comodidad:

$$f(y_i, \theta_i, \phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + h(y_i, \phi)\}$$

1. ***La distribución normal***

$$\begin{aligned} f(y_i, \theta_i, \phi) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right) \\ &= \exp\left[-\ln(2\pi\sigma^2) - \frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right] \\ &= \exp\left[\frac{1}{\sigma^2}\left(-\frac{y^2}{2} + y\mu - \frac{\mu^2}{2}\right) - \frac{1}{2}\ln(2\pi\sigma^2)\right] \\ &= \exp\left[\frac{1}{\sigma^2}\left(y\mu - \frac{\mu^2}{2}\right) - \frac{y^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right] \end{aligned}$$

Así, para la distribución normal,

$$\theta_i = \mu$$

$$b(\theta_i) = \frac{\mu^2}{2}$$

$$a(\phi) = \sigma^2$$

$$h(y_i, \phi) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)$$

$$E(y) = \frac{db(\theta_i)}{d\theta_i} = \mu \quad \text{y} \quad \text{Var}(y) = \frac{d^2b(\theta_i)}{d\theta_i^2}\,a(\phi) = \sigma^2$$

http://carlos2524.jimdo.com/

2. *La distribución binomial*

$$
\begin{aligned}
f(y_i, \theta_i, \phi) &= \binom{n}{m} \pi^y (1-\pi)^{n-y} \\
&= \exp\left\{ \ln\binom{n}{y} + y \ln \pi + (n-y)\ln(1-\pi) \right\} \\
&= \exp\left\{ \ln\binom{n}{y} + y \ln \pi + n \ln(1-\pi) - y \ln(1-\pi) \right\} \\
&= \exp\left\{ y \ln\left[\frac{\pi}{1-\pi}\right] + n \ln(1-\pi) + \ln\binom{n}{y} \right\}
\end{aligned}
$$

Por consiguiente, para la distribución binomial,

$$
\theta_i = \ln\left[\frac{\pi}{1-\pi}\right] \quad \text{y} \quad \pi = \frac{\exp(\theta_i)}{1+\exp(\theta_i)}
$$

$$
\begin{aligned}
b(\theta_i) &= -n \ln(1-\pi) \\
a(\phi) &= 1 \\
h(y_i, \phi) &= \ln\binom{n}{y} \\
E(y) &= \frac{db(\theta_i)}{d\theta_i} = \frac{db(\theta_i)}{d\pi}\frac{d\pi}{d\theta_i}
\end{aligned}
$$

Se observará que

$$
\begin{aligned}
\frac{d\pi}{d\theta_i} &= \frac{\exp(\theta_i)}{1+\exp(\theta_i)} - \left[\frac{\exp(\theta_i)}{1+\exp(\theta_i)}\right]^2 \\
&= \pi(1-\pi)
\end{aligned}
$$

Por consiguiente,

$$
\begin{aligned}
E(y) &= \left(\frac{n}{1-\pi}\right)\pi(1-\pi) \\
&= n\pi
\end{aligned}
$$

Se ve que éste es la media de la distribución binomial. También,

$$
\begin{aligned}
\text{Var}(y) &= \frac{dE(y)}{d\theta_i} \\
&= \frac{dE(y)}{d\pi}\frac{d\pi}{d\theta_i} \\
&= n\pi(1-\pi)
\end{aligned}
$$

Esta última ecuación no es más que la varianza de la distribución binomial.

http://carlos2524.jimdo.com/

3. *La distribución de Poisson*

$$f(y_i, \theta_i, \phi) = \frac{\lambda^y e^{-\lambda}}{y!}$$
$$= \exp[y \ln \lambda - \lambda - \ln(y!)]$$

Por tanto, para la distribución de Poisson,

$$\theta_i = \ln(\lambda) \quad \text{y} \quad \lambda = \exp(\theta_i)$$
$$b(\theta_i) = \lambda$$
$$a(\phi) = 1$$
$$h(y_i, \phi) = -\ln(y!)$$

Ahora bien,

$$E(y) = \frac{db(\theta_i)}{d\theta_i} = \frac{db(\theta_i)}{d\lambda}\frac{d\lambda}{d\theta_i}$$

Sin embargo, como

$$\frac{d\lambda}{d\theta_i} = \exp(\theta_i) = \lambda$$

la media de la distribución de Poisson es

$$E(y) = 1 \cdot \lambda = \lambda$$

La varianza de la distribución de Poisson es

$$\text{Var}(y) = \frac{dE(y)}{d\theta_i} = \lambda$$

C.13.3 Estimación de parámetros en el modelo lineal generalizado

Considérese el método de máxima verosimilitud aplicado al modelo lineal generalizado, y supóngase que se usa la cadena canónica. La función logaritmo de verosimilitud es

$$\ell(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} [y_i\theta_i - b(\theta_i)]/a(\phi) + h(y_i, \phi)$$

Para la cadena canónica, $\eta_i = g[E(y_i)] = g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$; por consiguiente,

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \frac{\partial \ell}{\partial \theta_i}\frac{\partial \theta_i}{\partial \boldsymbol{\beta}}$$
$$= \frac{1}{a(\phi)}\sum_{i=1}^{n}\left[y_i - \frac{db(\theta_i)}{d\theta_i}\right]\mathbf{x}_i$$
$$= \frac{1}{a(\phi)}\sum_{i=1}^{n}(y_i - \mu_i)\mathbf{x}_i$$

http://carlos2524.jimdo.com/

En consecuencia, se pueden calcular los estimados de máxima verosimilitud de los parámetros resolviendo el sistema de ecuaciones

$$\frac{1}{a(\phi)} \sum_{i=1}^{n} (y_i - \mu_i)\mathbf{x}_i = 0$$

En la mayor parte de los casos, $a(\phi)$ es constante, y entonces esas ecuaciones se transforman en

$$\sum_{i=1}^{n} (y_i - \mu_i)\mathbf{x}_i = 0$$

Éste es, en realidad, un **sistema** de $p = k + 1$ ecuaciones, una para cada parámetro de modelo. En forma matricial, las ecuaciones son

$$\mathbf{X}'(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}$$

en donde $\boldsymbol{\mu}' = [\mu_1, \mu_2, \ldots, \mu_p]$. Éstas son las llamadas ecuaciones de puntaje de máxima verosimilitud, y son exactamente las mismas que se vieron antes, en el caso de la regresión logística, donde $\mu' = [n_1\pi_1, n_2\pi_2, \ldots, n_n\pi_n]$.

Para resolver las ecuaciones de puntaje se puede usar IRLS, igual que en el caso de la regresión logística. Se comienza determinando una aproximación de primer orden con serie de Taylor, en la proximidad de la solución

$$y_i - \mu_i \approx \frac{d\mu_i}{d\eta_i}(\eta_i^* - \eta_i)$$

Ahora bien, para una cadena canónica, $\eta_i = \theta_i$, y

$$y_i - \mu_i \approx \frac{d\mu_i}{d\theta_i}(\eta_i^* - \eta_i) \tag{C.13.3}$$

Por consiguiente,

$$\eta_i^* - \eta_i \approx (y_i - \mu_i)\frac{d\theta_i}{d\mu_i}$$

Esta ecuación es una base para aproximar la varianza de $\hat{\eta}_i$.

En la estimación de máxima verosimilitud se reemplaza η_i por su estimado, $\hat{\eta}_i$. En ese caso,

$$\operatorname{Var}(\eta_i^* - \eta_i) \approx \operatorname{Var}\left[(y_i - \mu_i)\frac{d\theta_i}{d\mu_i}\right]$$

Como η_i^* y μ_i son constantes,

$$\operatorname{Var}(\hat{\eta}_i) \approx \left[\frac{d\theta_i}{d\mu_i}\right]^2 \operatorname{Var}(y_i)$$

http://carlos2524.jimdo.com/

Pero

$$\frac{d\theta_i}{d\mu_i} = \frac{1}{\text{var}(\mu_i)}$$

y $\text{Var}(y_i) = \text{var}(\mu_i)a(\phi)$. En consecuencia,

$$\text{Var}(\hat{\eta}_i) \approx \left[\frac{1}{\text{var}(\mu_i)}\right]^2 \text{var}(\mu_i)a(\phi)$$

$$\approx \frac{1}{\text{var}(\mu_i)}a(\phi)$$

Por comodidad, se definirá $\text{var}(\eta_i) = [\text{var}(\mu_i)]^{-1}$, por lo que

$$\text{Var}(\hat{\eta}_i) \approx \text{var}(\eta_i)a(\phi)$$

Esto se sustituye en la ecuación (C.13.3), para dar como resultado

$$y_i - \mu_i \approx \frac{1}{\text{var}(\eta_i)}(\eta_i^* - \eta) \tag{C.13.4}$$

Si se define a $\mathbf{V}$ como una matriz diagonal de $n \times n$, cuyos elementos diagonales son las $\text{var}(\eta_i)$, entonces la forma matricial de la ecuación (C.13.4) es

$$\mathbf{y} - \mu \approx \mathbf{V}^{-1}(\eta^* - \eta)$$

Entonces se pueden reexpresar las ecuaciones de puntaje en la siguiente forma:

$$\mathbf{X}'(\mathbf{y} - \mu) = \mathbf{0}$$

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \eta) = \mathbf{0}$$

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \mathbf{X}\beta) = \mathbf{0}$$

Y el estimado de máxima verosimilitud de β es

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\eta^*$$

Ahora bien, de la misma manera que en el caso de la regresión logística, no se conoce η^*, por lo que se hará un esquema iterativo basado en

$$z_i = \hat{\eta}_i + (y_i - \hat{\mu}_i)\frac{d\eta_i}{d\mu_i}$$

Al usar los mínimos cuadrados iterativamente reponderados, con el método de Newton-Raphson, la solución se determina partiendo de

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{z}$$

http://carlos2524.jimdo.com/

En forma asintótica, el componente aleatorio de $\mathbf{z}$ se origina en las observaciones y_i. Los elementos diagonales de la matriz $\mathbf{V}$ son las varianzas de las z_i, además de $a(\phi)$.

Por ejemplo, considérese el caso de la regresión logística:

$$\eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

$$\begin{aligned}
\frac{d\eta_i}{d\mu_i} = \frac{d\eta_i}{d\pi_i} &= \frac{d\ln\left(\dfrac{\pi_i}{1-\pi_i}\right)}{d\pi_i} \\
&= \frac{1-\pi_i}{\pi_i}\left[\frac{\pi_i}{1-\pi_i} + \frac{\pi_i}{(1-\pi_i)^2}\right] \\
&= \frac{(1-\pi_i)}{\pi_i(1-\pi_i)}\left[1 + \frac{\pi_i}{1-\pi_i}\right] \\
&= \frac{1}{\pi_i}\left[\frac{1-\pi_i+\pi_i}{1-\pi_i}\right] \\
&= \frac{1}{\pi_i(1-\pi_i)}
\end{aligned}$$

Así, para la regresión logística, los elementos diagonales de la matriz $\mathbf{V}$ son

$$\begin{aligned}
\left(\frac{d\eta_i}{d\mu_i}\right)^2 \operatorname{Var}(y_i) &= \left[\frac{1}{\pi_i(1-\pi_i)}\right]^2 \frac{\pi_i(1-\pi_i)}{n_i} \\
&= \frac{1}{n_i\pi_i(1-\pi_i)}
\end{aligned}$$

que es exactamente lo que se obtuvo antes.

Por lo anterior, el IRLS basado en el método de Newton-Raphson se puede describir como sigue:

1. Con mínimos cuadrados ordinarios obtener un estimado inicial de β, que será $\hat{\beta}_0$;
2. Usar $\hat{\beta}_0$ para estimar $\mathbf{V}$ y $\boldsymbol{\mu}$;
3. Sea $\eta_0 = \mathbf{X}\hat{\beta}_0$;
4. Definir $\mathbf{z}_1$ basado en η_0;
5. Obtener un nuevo estimado $\hat{\beta}_1$ e iterar hasta que se satisfaga algún criterio adecuado de convergencia.

Si no se usa la cadena canónica, entonces $\eta_i \neq \theta_i$, y la derivada adecuada del logaritmo de verosimilitud es

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \frac{d\ell}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}$$

http://carlos2524.jimdo.com/

Nótese que:

1. $\dfrac{d\ell}{d\theta_i} = \dfrac{1}{a(\phi)}\left[y_i = \dfrac{db(\theta_i)}{d\theta_i}\right] = \dfrac{1}{a(\phi)}(y_i - \mu_i)$
2. $\dfrac{d\theta_i}{d\mu_i} = \dfrac{1}{\text{var}(\mu_i)}$ y
3. $\dfrac{\partial\eta_i}{\partial\boldsymbol{\beta}} = \mathbf{x}_i$

Al reunir todo lo anterior se obtiene

$$\frac{\partial\ell}{\partial\boldsymbol{\beta}} = \frac{y_i - \mu_i}{a(\phi)}\frac{1}{\text{var}(\mu_i)}\frac{d\mu_i}{\partial\eta_i}\mathbf{x}_i$$

También aquí se puede usar un desarrollo en serie de Taylor, para llegar a

$$y_i - \mu_i \approx \frac{d\mu_i}{d\eta_i}(\eta_i^* - \eta_i)$$

Siguiendo un argumento parecido al que se usó antes,

$$\text{Var}(\hat{\eta}_i) \approx \left[\frac{d\theta_i}{d\mu_i}\right]^2 \text{Var}(y_i)$$

y al final se puede demostrar que

$$\frac{\partial\ell}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\frac{\eta_i^* - \eta_i}{a(\phi)\text{var}(\eta_i)}\mathbf{x}_i$$

Igualando a cero esta última ecuación, y escribiéndola en forma matricial, se obtiene

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \eta) = \mathbf{0}$$

o bien, ya que $\eta = \mathbf{X}\beta$,

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \mathbf{X}\beta) = \mathbf{0}$$

La solución de Newton-Raphson se basa en

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{z},$$

siendo

$$z_i = \hat{\eta}_i + (y_i - \hat{\mu}_i)\frac{d\eta_i}{d\mu_i}$$

Igual que en el caso de la cadena canónica, la matriz $\mathbf{V}$ es una matriz diagonal formada por las varianzas de los predictores lineales estimados, además de $a(\phi)$.

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

REFERENCIAS

Adichie, J. N., "Estimates of regression parameters based on rank tests", *Ann. Math. Stat.*, **38**, 1967, pp. 894-904.

Aitkin, M. A., "Simultaneous inference and the choice of variable subsets", *Technometrics*, **16**, 1974, pp. 221-227.

Allen, D. M., "Mean square error of prediction as a criterion for selecting variables", *Technometrics*, **13**, 1971, pp. 469-475.

Allen, D. M., "The relationship between variable selection and data augmentation and a method for prediction", *Technometrics*, **16**, 1974, pp. 125-127.

Andrews, D. F., "Significance tests based on residuals", *Biometrika*, **58**, 1971, pp. 139-148.

Andrews, D. F., "A robust method for multiple linear regression", *Technometrics*, **16**, 1974, pp. 523-531.

Andrews, D. F., "The robustness of residual displays", en R. L. Launer y G. N. Wilkinson (eds.), *Robustness in Statistics*, Academic Press, Nueva York, 1979, pp. 19-32.

Andrews, D. F., P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers y J. W. Tukey, *Robust Estimates of Location*, Princeton University Press, Princeton, N.J., 1972.

Anscombe, F. J., "Examination of residuals", en *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, University of California, Berkeley, 1961, pp. 1-36.

Anscombe, F. J., "Topics in the investigation of linear relations fitted by the method of least squares", *J. R. Stat. Soc. Ser. B*, **29**, 1967, pp. 1-52.

Anscombe, F. J., "Graphs in statistical analysis", *Am. Stat.*, **27**(1), 1973, pp. 17-21.

Anscombe, F. J. y J. W. Tukey, "The examination and analysis of residuals", *Technometrics*, **5**, 1963, pp. 141-160.

Askin, R. G. y D. C. Montgomery, "Augmented robust estimators", *Technometrics*, **22**, 1980, pp. 333-341.

Askin, R. G. y D. C. Montgomery, "An analysis of constrained robust regression estimators", *Nav. Res. Logistics Q*, **31**, 1984, pp. 283-296.

Atkinson, A. C., "Diagnostic regression for shifted power transformations", *Technometrics*, **25**, 1983, pp. 23-33.

Atkinson, A. C., *Plots, Transformations, and Regression*, Clarendon Press, Oxford, 1985.

Atkinson, A. C., "Fast very robust methods for the detection of multiple outliers", *J. Am. Stat. Assoc.*, **89**, 1994, pp. 1329-1339.

Barnett, V. y T. Lewis, *Outliers in Statistical Data*, segunda edición, Wiley, Nueva York, 1994.

Bates, D. M. y D. G. Watts, *Nonlinear Regression Analysis and Its Aplications*, Wiley, Nueva York, 1988.

http://carlos2524.jimdo.com/

Beaton, A. E., *The Use of Special Matrix Operators in Statistical Calculus*, Research Bulletin RB-64-51, Educational Testing Service, Princeton, N.J., 1964.

Beaton, A. E. y J. W. Tukey, "The fitting of power series, meaning polynomials, illustrated on band spectroscopic data", *Technometrics*, **16**, 1974, pp. 147-185.

Belsley, D. A., E. Kuh y R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, Nueva York, 1980.

Bendel, R. B. y A. A. Afifi, "Comparison of stopping rules in forward stepwise regressions", presentado en la Joint Statistical Meeting, St. Louis, Mo., 1974.

Berk, K. N., "Comparing subset regression procedures", *Technometrics*, **20**, 1978, pp. 1-6.

Berkson, J., "Are there two regressions?", *J. Am. Stat. Assoc.*, **45**, 1950, pp. 164-180.

Berkson, J., "Estimation of a linear function for a calibration line; consideration of a recent proposal", *Technometrics*, **11**, 1969, pp. 649-660.

Bishop, C. M., *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.

Bloomfield, P. y W. L. Steiger, *Least Absolute Deviations: Theory, Applications, and Algorithms*, Birkhuser Verlag, Boston, 1983.

Book, D., J. Booker, H. O. Hartley y R. I. Sielken, Jr., "Unbiased L1 estimators and their covariances", ONR THEMIS Technical Report núm. 64, Institute of Statistics, Texas A & M University, 1980.

Box, G. E. P., "Use and abuse of regression", *Technometrics*, **8**, 1966, pp. 625-629.

Box, G. E. P. y D. W. Behnken, "Some new three level designs for the study of quantitative variables", *Technometrics*, **2**, 1960, pp. 455-475.

Box, G. E. P. y D. R. Cox, "An analysis of transformations", *J. R. Stat. Soc. Ser. B*, **26**, 1964, pp. 211-243.

Box, G. E. P. y N. R. Draper, "A basis for the selection of a response surfase design", *J. Am. Stat. Assoc.*, **54**, 1959, pp. 622-654.

Box, G. E. P. y N. R. Draper, "The choice of a second-order rotatable design", *Biometrika*, **50**, 1963, pp. 335-352.

Box, G. E. P. y N. R. Draper, *Empirical Model Building and Response Surfaces*, Wiley, Nueva York, 1987.

Box, G. E. P. y J. S. Hunter, "Multifactor experimental designs for exploring response surfaces", *Ann. Math. Stat.*, **28**, 1957, pp. 195-242.

Box, G. E. P., W. G. Hunter y J. S. Hunter, *Statistics for Experimenters*, Wiley, Nueva York, 1978.

Box, G. E. P., G. M. Jenkins y G. C. Reinsel, *Time Series Analysis, Forecasting, and Control*, tercera edición, Prentice-Hall, Englewood Cliffs, N.J., 1994.

Box, G. E. P. y P. W. Tidwell, "Transformation of the independent variables", *Technometrics*, **4**, 1962, pp. 531-550.

Box, G. E. P. y J. M. Wetz, "Criterion for judging the adequacy of estimation by an approximating response polynomial", Technical Report núm. 9, Department of Statistics, University of Wisconsin, Madison, 1973.

Bradley, R. A. y S. S. Srivastava, "Correlation and polynomial regression", *Am. Stat.*, **33**, 1979, pp. 11-14.

http://carlos2524.jimdo.com/

Breiman, L., J. H. Friedman, R. A. Olshen y C. J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, Calif., 1984.

Brown, P. J., "Centering and scaling in ridge regression", *Technometrics*, **19**, 1977, pp. 35-36.

Brown, R. L., J. Durbin y J. M. Evans, "Techniques for testing the constancy of regression relationships over time (with discussion)", *J. R. Stat. Soc. Ser. B*, **37**, 1975, pp. 149-192.

Buse, A. y L. Lim, "Cubic splines as a special case of restricted least squares", *J. Am. Stat. Assoc.*, **72**, 1977, pp. 64-68.

Cady, F. B. y D. M. Allen, "Combining experiments to predict future yield data", *Agron. J.*, **64**, 1972, pp. 211-214.

Carroll, R. J. y D. Ruppert, "Transformation in regression: A robust analysis", *Technometrics*, **27**, 1985, pp. 1-12.

Carroll, R. J. y D. Ruppert, *Transformation and Weighting in Regression*, Chapman & Hall, Londres, 1988.

Chatterjee, S. y B. Price, *Regression Analysis by Example*, Wiley, Nueva York, 1977.

Coakley, C. W. y T. P. Hettmansperger, "A bounded influence, high breakdown, efficient regression estimator", *J. Am. Stat. Assoc.*, **88**, 1993, pp. 872-880.

Cochrane, D. y G. H. Orcutt, "Application of least squares regression to relationships containing autocorrelated error terms", *J. Am. Stat. Assoc.*, **44**, 1949, pp. 32-61.

Conniffe, D. y J. Stone, "A critical view of ridge regression", *The Statistician*, **22**, 1973, pp. 181-187.

Conniffe, D. y J. Stone, "A reply to Smith and Goldstein", *The Statistician*, **24**, 1975, pp. 67-68.

Cook, R. D., "Detection of influential observation in linear regression", *Technometrics*, **19**, 1977, pp. 15-18.

Cook, R. D., "Influential observations in linear regression", *J. Am. Stat. Assoc.*", **74**, 1979, pp. 169-174.

Cook, R. D., "Exploring partial residual plots", *Technometrics*, **35**, 1993, pp. 351-362.

Cook, R. D. y P. Prescott, "On the accuracy of Bonferroni significance levels for detecting outliers in linear models", *Technometrics*, **22**, 1981, pp. 59-63.

Cook, R. D. y S. Weisberg, "Diagnostics for heteroscedasticity in regression", *Biometrika*, **70**, 1983, pp. 1-10.

Cox, D. R. y E. J. Snell, "The choice of variables in observational studies", *Appl. Stat.*, **23**, 1974, pp. 51-59.

Curry, H. B. e I. J. Schoenberg, "On Polya frequency functions IV: The fundamental spline functions and their limits", *J. Anal. Math.*, **17**, 1966, pp. 71-107.

Daniel, C., *Applications of Statistics to Industrial Experimentation*, Wiley, Nueva York, 1976.

Daniel, C. y F. S. Wood, *Fitting Equations to Data*, segunda edición, Wiley, Nueva York, 1980.

Davies, R. B. y B. Hutton, "The effects of errors in the independent variables in linear regression", *Biometrika*, **62**, 1975, pp. 383-391.

http://carlos2524.jimdo.com/

Davidson, A. C. y D. V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, Londres, 1997.

De Jong, P. J., T. De Wet y A. H. Welsh, "Mallows-type bounded-influence-regression trimmed means", *J. Am. Stat. Assoc.*, **83**, 1988, pp. 805-810.

DeLury, D. B., *Values and Integrals of the Orthogonal Polynomials up to N = 26*, University of Toronto Press, Toronto, 1960.

Dempster, A. P., M. Schatzoff y N. Wermuth, "A simulation study of alternatives to ordinary least squares", *J. Am. Stat. Assoc.*, **72**, 1977, pp. 77-90.

Denby, L. y W. A. Larson, "Robust regression estimators compared via Monte Carlo", *Commun. Stat.*, **A6**, 1977, pp. 335-362.

Dodge, Y., *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, North-Holland, Amsterdam, 1987.

Dolby, G. R., "The ultrastructural relation: A synthesis of the functional and structural relations", *Biometrika*, **63**, 1976, pp. 39-50.

Dolby, J. L., "A quick method for choosing a transformation", *Technometrics*, **5**, 1963, pp. 317-325.

Draper, N. R., J. Guttman y H. Kanemasa, "The distribution of certain regression statistics", *Biometrika*, **58**, 1971, pp. 295-298.

Draper, N. R. y H. Smith, *Applied Regression Analysis*, tercera edición, Wiley, Nueva York, 1998.

Draper, N. R. y R. C. Van Nostrand, "Shrinkage estimators: Review and comments", Technical Report núm. 500, Department of Statistics, University of Wisconsin, Madison, 1977a.

Draper, N. R. y R. C. Van Nostrand, "Ridge regression: Is it worthwhile?" Technical Report núm. 501, Department of Statistics, University of Wisconsin, Madison, 1977b.

Draper, N. R. y R. C. Van Nostrand, "Ridge regression and James-Stein estimators: Review and comments", *Technometrics*, **21**, 1979, pp. 451-466.

Durbin, J., "Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables", *Econometrica*, **38**, 1970, pp. 410-421.

Durbin, J. y G. S. Watson, "Testing for serial correlation in least squares regression I", *Biometrika*, **37**, 1950, pp. 409-438.

Durbin, J. y G. S. Watson, "Testing for serial correlation in least squares regression II", *Biometrika*, **38**, 1951, pp. 159-178.

Durbin, J. y G. S. Watson, "Testing for serial correlation in least squares regression III", *Biometrika*, **58**, 1971, pp. 1-19.

Dutter, R., "Numerical solution of robust regression problems: Computational aspects, a comparison", *J. Stat. Comput. Simul.*, **5**, 1977, pp. 207-238.

Dutter, R. y P. J. Hober, "Numerical methods for the robust nonlinear regression problem", *J. Stat. Comput. Simul.*, **13**, 1981, pp. 79-114.

Dykstra, O., Jr., "The augmentation of experimental data to maximize $|\mathbf{X}'\mathbf{X}|$", *Technometrics*, **13**, 1971, pp. 682-688.

Edwards, J. B., "The relation between the F-test and R^2", *Am. Stat.*, **23**, 1969, p. 28.

Efron, B., "Bootstrap methods: Another look at the jackknife", *Ann. Stat.*, **7**, 1979, pp. 1-26.

http://carlos2524.jimdo.com/

Efron, B., *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia, 1982.

Efron, B., "Better bootstrap confidence intervals (with discussion)", *J. Am. Stat. Assoc.*, **82**, 1987, pp. 172-200.

Efron, B. y R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy", *Stat. Sci.*, **1**, 1986, pp. 54-77.

Efron, B. y R. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, Londres, 1993.

Efroymson, M. A., "Multiple regression analysis", en A. Ralston y H. S. Wilf (eds.), *Mathematical Methods for Digital Computers*, Wiley, Nueva York, 1960.

Ellerton, R. R. W., "Is the regression equation adequate-A generalization", *Technometrics*, **20**, 1978, pp. 313-316.

Eubank, R. L., *Spline Smoothing and Nonparametric Regression*, Dekker, Nueva York, 1988.

Eubank, R. L. y P. Speckman, "Curve fitting by polynomial-trigonometric regression", *Biometrika*, **77**, 1990, pp. 1-9.

Everitt, B. S., *Cluster Analysis*, tercera edición, Halsted Press, Nueva York, 1993.

Farrar, D. E. y R. R. Glauber, "Multicollinearity in regression analysis: The problem revisited", *Rev. Econ. Stat.*, **49**, 1967, pp. 92-107.

Feder, P. I., "Graphical techniques in statistical data analysis-Tools for extracting information from data", *Technometrics*, **16**, 1974, pp. 287-299.

Forsythe, A. B., "Robust estimation of straight-line regression coefficients by minimizing *p*th power deviations", *Technometrics*, **14**, 1972, pp. 159-166.

Forsythe, G. E., "Generation and use of orthogonal polynomials for data-fitting with a digital computer", *J. Soc. Ind. Appl. Math.*, **5**, 1957, pp. 74-87.

Fuller, W. A., *Introduction to Statistical Time Series*, Wiley, Nueva York, 1976.

Furnival, G. M., "All possible regressions with less computation", *Technometrics*, **13**, 1971, pp. 403-408.

Furnival, G. M. y R. W. M. Wilson, Jr., "Regression by leaps and bounds", *Technometrics*, **16**, 1974, pp. 499-511.

Gallant, A. R. y W. A. Fuller, "Fitting segmented polynomial regression models whose join points have to be estimated", *J. Am. Stat. Assoc.*, **63**, 1973, pp. 144-147.

Garside, M. J., "The best subset in multiple regression analysis", *Appl. Stat.*, **14**, 1965, pp. 196-200.

Gartside, P. S., "A study of methods for comparing several variances", *J. Am. Stat. Assoc.*, **67**, 1972, pp. 342-346.

Gaylor, D. W. y J. A. Merrill, "Augmenting existing data in multiple regression", *Technometrics*, **10**, 1968, pp. 73-81.

Geisser, S., "The predictive sample reuse method with applications", *J. Am. Stat. Assoc.*, **70**, 1975, pp. 320-328.

Gentle, J. M., W. J. Kennedy y V. A. Sposito, "On least absolute deviations estimators", *Commun. Stat.*, **A6**, 1977, pp. 839-845.

Gibbons, D. G., *A Simulation Study of Some Ridge Estimators*, General Motors Research Laboratories, Mathematics Department, GMR-2659 (rev. ed.), Warren, Mich., 1979.

http://carlos2524.jimdo.com/

Gnanadesikan, R., *Methods for Statistical Analysis of Multivariate Data*, Wiley, Nueva York, 1977.

Goldberger, A. S., *Econometric Theory*, Wiley, Nueva York, 1964.

Goldstein, M. y A. F. M. Smith, "Ridge-type estimators for regression analysis", *J. R. Stat. Soc. Ser. B*, **36**, 1974, pp. 284-291.

Golub, G. H., "Matrix decompositions and statistical calculations", en R. C. Milton y J. A. Welder (eds.), *Statistical Computation*, Academic, Nueva York, 1969.

Gorman, J. W. y R. J. Toman, "Selection of variables for fitting equations to data", *Technometrics*, **8**, 1966, pp. 27-51.

Graybill, F. A., *An Introduction to Linear Statistical Models*, vol. 1, McGraw-Hill, Nueva York, 1961.

Graybill, F. A., *Theory and Application of the Linear Model*, Duxbury, North Scituate, Mass., 1976.

Guilkey, D. K. y J. L. Murphy, "Directed ridge regression techniques in cases of multicollinearity", *J. Am. Stat. Assoc.*, **70**, 1975, pp. 769-775.

Gunst, R. F., "Similarities among least squares, principal component, and latent root regression estimators", presentado en Washington, D.C., Joint Statistical Meetings, 1979.

Gunst, R. F. y R. L. Mason, "Biased estimation in regression: An evaluation using mean squared error", *J. Am. Stat. Assoc.*, **72**, 1977, pp. 616-628.

Gunst, R. F. y R. L. Mason, "Some considerations in the evaluation of alternative prediction equations", *Technometrics*, **21**, 1979, pp. 55-63.

Gunst, R. F., J. T. Webster y R. L. Mason, "A comparison of least squares and latent root regression estimators", *Technometrics*, **18**, 1976, pp. 75-83.

Gunter, B., "Tree-based classification and regression. Part I: Background and fundamentals", *Qual. Prog.*, **28**, agosto, 1997a, pp. 159-163.

Gunter, B., "Tree-based classification and regression. Part II: Assessing classification performance", *Qual. Prog.*, **28**, diciembre, 1997b, pp. 83-84.

Gunter, B., "Tree-based classification and regression. Part III: Tree-based procedures", *Qual. Prog.*, **31**, febrero, 1998, p. 121.

Hadi, A. S. y J. S. Simonoff, "Procedures for the identification of multiple outliers in linear models", *J. Am. Stat. Assoc.*, **88**, 1993, pp. 1264-1272.

Hahn, G. J., "Simultaneous prediction intervals for a regression model", *Technometrics*, **14**, 1972, pp. 203-214.

Hahn, G. J., "The coefficient of determination exposed!" *Chem. Technol.*, **3**, 1973, pp. 609-614.

Hahn, G. J., "Fitting regression models with no intercept term", *J. Qual. Technol.*, **9**(2), 1979, pp. 56-61.

Hahn, G. J. y R. W. Hendrickson, "A table of percentage points of the largest absolute value of k student t variates and its applications", *Biometrika*, **58**, 1971, pp. 323-332.

Haitovski, Y., "A note on the maximization of R^2", *Am. Stat.*, **23**(1), 1969, pp. 20-21.

Hald, A., *Statistical Theory with Engineering Applications*, Wiley, Nueva York, 1952.

Halperin, M., "Fitting of straight lines and prediction when both variables are subject to error", *J. Am. Stat. Assoc.*, **56**, 1961, pp. 657-669.

http://carlos2524.jimdo.com/

Halperin, M., "On inverse estimation in linear regression", *Technometrics*, **12**, 1970, pp. 727-736.

Hawkins, D. M., "On the investigation of alternative regressions by principal components analysis", *Appl. Stat.*, **22**, 1973, pp. 275-286.

Hawkins, D. M., "The feasible solution algorithm for least trimmed squares regression", *Comput. Stat. Data Anal.*, **17**, 1994, pp. 185-196.

Hawkins, D. M., D. Bradu y G. V. Kass, "Location of several outliers in multiple regression using elemental sets", *Technometrics*, **26**, 1984, pp. 197-208.

Hayes, J. G. (ed.), *Numerical Approximations to Functions and Data*, Athlone Press, Londres, 1970.

Hayes, J. G., "Numerical methods for curve and surface fitting", J. Inst. Math. Appl., **10**, 1974, pp. 144-152.

Haykin, S., *Neural Networks: A Comprehensive Foundation*, Macmillan Co., Nueva York, 1994.

Hemmerle, W. J., "An explicit solution for generalized ridge regression", *Technometrics*, **17**, 1975, pp. 309-314.

Hemmerle, W. J. y T. F. Brantle, "Explicit and constrained generalized ridge regression", *Technometrics*, **20**, 1978, pp. 109-120.

Hettmansperger, T. P. y J. W. Mckean, *Robust Nonparametric Statistical Methods*, vol. 5 of Kendall's Library of Statistics, Arnold, Londres, 1998.

Hill, R. C., G. G. Judge y T. B. Fomby, "On testing the adequacy of a regression model", *Technometrics*, **20**, 1978, pp. 491-494.

Hill, R. W., "On estimating the covariance matrix of robust regression M-estimates", Commun. Stat., **A8**, 1979, pp. 1183-1196.

Himmelblau, D. M., *Process Analysis by Statistical Methods*, Wiley, Nueva York, 1970.

Hoadley, B., "A Bayesian look at inverse linear regression", *J. Am. Stat. Assoc.*, **65**, 1970, pp. 356-369.

Hoaglin, D. C. y R. E. Welsch, "The hat matrix in regression and ANOVA", *Am. Stat.*, **32**(1), 1978, pp. 17-22.

Hocking, R. R., "Criteria for selection of a subset regression: Which one should be used", *Technometrics*, **14**, 1972, pp. 967-970.

Hocking, R. R., "Misspecification in regression", *Am. Stat.*, **28**, 1974, pp. 39-40.

Hocking, R. R., "The analysis and selection of variables in linear regression", *Biometrics*, **32**, 1976, pp. 1-49.

Hocking, R. R. y L. R. LaMotte, "Using the SELECT program for choosing subset regressions", en W. O. Thompson y F. B. Cady (eds.), *Proceedings of the University of Kentucky Conference on Regression with a Large Number of Predictor Variables*, Department of Statistics, University of Kentucky, Lexington, 1973.

Hocking, R. R., F. M. Speed y M. J. Lynn, "A class of biased estimators in linear regression", *Technometrics*, **18**, 1976, pp. 425-437.

Hodges, S. D. y P. G. Moore, "Data uncertainties and least squares regression", *Appl. Stat.*, **21**, 1972, pp. 185-195.

Hoerl, A. E., "Optimum solution of many variable equations", *Chem. Eng. Prog.*, **55**, 1959, p. 69.

http://carlos2524.jimdo.com/

Hoerl, A. E. y R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics*, **12**, 1970a, pp. 55-67.

Hoerl, A. E. y R. W. Kennard, "Ridge regression: Applications to nonorthogonal problems", *Technometrics*, **12**, 1970b, pp. 69-82.

Hoerl, A. E. y R. W. Kennard, "Ridge regression: Iterative estimation of the biasing parameter", *Commun. Stat.*, **A5**, 1976, pp. 77-88.

Hoerl, A. E., R. W. Kennard y K. F. Baldwin, "Ridge regression: Some simulations", *Commun. Stat.*, **4**, 1975, pp. 105-123.

Hogg, R. V., "Adaptive robust procedures: A partial review and some suggestions for future applications and theory", *J. Am. Stat. Assoc.*, **69**, 1974, pp. 909-925.

Hogg, R. V., "Statistical robustness: One view of its use in applications today", *Am. Stat.*, **33**(3), 1979a, pp. 108-115.

Hogg, R. V., "An introduction to robust estimation", en R. L. Launer y G. N. Wilkinson (eds.), *Robustness in Statistics*, Academic, Nueva York, 1979b, pp. 1-18.

Hogg, R. V. y R. H. Randles, "Adaptive distribution-free regression methods and their applications", *Technometrics*, **17**, 1975, p. 399-407.

Holland, P. W. y R. E. Welsch, "Robust regression using iteratively reweighted least squares", *Commun. Stat.*, **A6**, 1977, pp. 813-828.

Huber, P. J., "Robust estimation of a location parameter", *Ann. Math. Stat.*, **35**, 1964, pp. 73-101.

Huber, P. J., "Robust statistics: A review", *Ann. Math. Stat.*, **43**, 1972, pp. 1041-1067.

Huber, P. J., "Robust regression: Asymptotics, conjectures, and Monte Carlo", *Ann. Stat.*, **1**, 1973, pp. 799-821.

Huber, P. J., *Robust Statistics*, Wiley, Nueva York, 1981.

Jaeckel, L. A., "Estimating regression coefficients by minimizing the dispersion of the residuals", *Ann. Math. Stat.*, **43**, 1972, pp. 1449-1458.

Joglekar, G., J. H. Schuenemeyer y V. LaRiccia, "Lack-of-fit testing when replicates are not available", *Am. Stat.*, **43**, 1989, pp. 135-143.

Johnson, R. A. y D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1992.

Johnston, J., *Econometric Methods*, McGraw-Hill, Nueva York, 1972.

Jurečková, J., "Asymptotic relations of *M*-estimates and *R*-estimates in linear regression models", *Ann. Stat.*, **5**, 1977, pp. 464-472.

Kalotay, A. J., "Structural solution to the linear calibration problem", *Technometrics*, **13**, 1971, pp. 761-769.

Kendall, M. G. y G. U. Yule, *An Introduction to the Theory of Statistics*, Charles Griffin, Londres, 1950.

Kennard, R. W. y L. Stone, "Computer aided design of experiments", *Technometrics*, **11**, 1969, pp. 137-148.

Kennedy, W. J. y T. A. Bancroft, "Model-building for prediction in regression using repeated significance tests", *Ann. Math. Stat.*, **42**, 1971, pp. 1273-1284.

Khuri, A. H. y J. A. Cornell, *Response Surfaces: Designs and Analyses*, segunda edición, Dekker, Nueva York, 1996.

http://carlos2524.jimdo.com/

Krasker, W. S. y R. E. Welsch, "Efficient bounded-influence regression estimation", *J. Am. Stat. Assoc.*, **77**, 1982, pp. 595-604.

Krutchkoff, R. G., "Classical and inverse regression methods of calibration", *Technometrics*, **9**, 1967, pp. 425-439.

Krutchkoff, R. G., "Classical and inverse regression methods of calibration in extrapolation", *Technometrics*, **11**, 1969, pp. 605-608.

Kunugi, T., T. Tamura y T. Naito, "New acetylene process uses hydrogen dilution", *Chem. Eng. Prog.*, **57**, 1961, pp. 43-49.

Land, C. E., "Confidence interval estimation for means after data transformation to normality", *J. Am. Stat. Assoc.*, **69**, 1974, pp. 795-802 (Correction, *ibid.*, **71**, p. 255).

Larsen, W. A. y S. J. McCleary, "The use of partial residual plots in regression analysis", *Technometrics*, **14**, 1972, pp. 781-790.

Lawless, J. F., "Ridge and related estimation procedures: Theory and practice", *Commun. Stat.*, **A7**, 1978, pp. 139-164.

Lawless, J. F. y P. Wang, "A simulation of ridge and other regression estimators", *Commun. Stat.*, **A5**, 1976, pp. 307-323.

Lawrense, K. D. y J. L. Arthur, "Robust nonlinear regression", en K. D. Lawrence y J. L. Arthur (eds.), *Robust Regression: Analysis and Applications*, Dekker, Nueva York, 1990, pp. 59-86.

Lawson, C. R. y R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N. J., 1974.

Leamer, E. E., "Multicollinearity: A Bayesian interpretation", *Rev. Econ. Stat.*, **55**, 1973, pp. 371-380.

Leamer, E. E., *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, Wiley, Nueva York, 1978.

Levine, H., "Robust tests for equality of variances", en I. Olkin (ed.), *Contributions to Probability and Statistics*, Stanford University Press, Palo Alto, Calif., 1960, pp. 278-292.

Lieberman, G. J., R. G. Miller, Jr. y M. A. Hamilton, "Unlimited simultaneous discrimination intervals in regression", *Biometrika*, **54**, 1967, pp. 133-145.

Lindley, D. V., "Regression lines and the linear functional relationship", *J. R. Stat. Soc. Suppl.*, **9**, 1974, pp. 218-244.

Lindley, D. V. y A. F. M. Smith, "Bayes estimates for the linear model (with discussion)", *J. R. Stat. Soc. Ser. B*, **34**, 1972, pp. 1-41.

Looney, S. W. y T. R. Gulledge, Jr., "Use of the correlation coefficient with normal probability plots", *Am. Stat.*, **35**, 1985, pp. 75-79.

Lowerre, J. M., "On the mean square error of parameter estimates for some biased estimators", *Technometrics*, **16**, 1974, pp. 461-464.

McCarthy, P. J., "The use of balanced half-sample replication in cross-validation studies", *J. Am. Stat. Assoc.*, **71**, 1976, pp. 596-604.

McCullagh, P. y J. A. Nelder, *Generalized Linear Models*, segunda edición, Chapman & Hall, Londres, 1989.

McDonald, G. C. y D. I. Galarneau, "A Monte Carlo evaluation of some ridge-type estimators", *J. Am. Stat. Assoc.*, **70**, 1975, pp. 407-416.

http://carlos2524.jimdo.com/

Mallows, C. L., "Choosing variables in a linear regression: A graphical aid", presentado en Central Regional Meeting of the Institute of Mathematical Statistics, Manhattan, Kans, 1964.

Mallows, C. L., "Choosing a subset regression", presentado en Joint Statistical Meetings, Los Ángeles, 1966.

Mallows, C. L., "Some comments on C_p", *Technometrics*, **15**, 1973, pp. 661-675.

Mallows, C. L., "Augmented partial residuals", *Technometrics*, **28**, 1986, pp. 313-319.

Mallows, C. L., "More comments on C_p", *Technometrics*, **37**, 1995, pp. 362-372. (Véase también, **39**, 1997, pp. 115-116.)

Mandansky, A., "The fitting of straight lines when both variables are subject to error", *J. Am. Stat. Assoc.*, **54**, 1959, pp. 173-205.

Mansfield, E. R., J. T. Webster y R. F. Gunst, "An analytic variable selection procedure for principal component regression", *Appl. Stat.*, **26**, 1977, pp. 34-40.

Mansfield, E. R. y M. D. Conerly, "Diagnostic value of residual and partial residual plots", *Am. Stat.*, **41**, 1987, pp. 107-116.

Mantel, N., "Why stepdown procedures in variable selection", *Technometrics*, **12**, 1970, pp. 621-625.

Marazzi, A., *Algorithms, Routines and S Functions for Robust Statistics*, Wadsworth and Brooks/Cole, Pacific Grove, Calif., 1993.

Maronna, R. A., "Robust *M*-estimators of multivariate location and scatter", *Ann. Stat.*, **4**, 1976, pp. 51-67.

Marquardt, D. W., "An algorithm for least squares estimation of nonlinear parameters", *J. Soc. Ind. Appl. Math.*, **2**, 1963, pp. 431-441.

Marquardt, D. W., "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation", *Technometrics*, **12**, 1970, pp. 591-612.

Marquardt, D. W. y R. D. Snee, "Ridge regression in practice", *Am. Stat.*, **29**(1), 1975, pp. 3-20.

Mason, R. L., R. F. Gunst y J. T. Webster, "Regression analysis and problems of multicollinearity", *Commun. Stat.*, **4**(3), 1975, pp. 277-292.

Mayer, L. S. y T. A. Willke, "On biased estimation in linear models", *Technometrics*, **16**, 1973, pp. 494-508.

Miller, D. M., "Reducing transformation bias in curve fitting", *Am. Stat.*, **38**, 1984, pp. 124-126.

Miller, R. G., Jr., *Simultaneous Statistical Inference*, McGraw-Hill, Nueva York, 1966.

Montgomery, D. C., *Design and Analysis of Experiments*, quinta edición, Wiley, Nueva York, 2001.

Montgomery, D.C., L. A. Johnson y J. S. Gardiner, *Forecasting and Time Series Analysis*, segunda edición, McGraw-Hill, Nueva York, 1990.

Montgomery, D. C., E. W. Martin y E. A. Peck, "Interior analysis of the observations in multiple linear regression", *J. Qual. Technol.*, **12**(3), 1980, pp. 165-173.

Morgan, J. A. y J. F. Tatar, "Calculation of the residual sum of squares for all posible regressions", *Technometrics*, **14**, 1972, pp. 317-325.

Mosteller, F. y J. W. Tukey, "Data analysis including statistics", en G. Lindzey y E. Aronson (eds.), *Handbook of Social Psychology*, vol. 2, Addison-Wesley, Reading, Mass., 1968.

http://carlos2524.jimdo.com/

Mosteller, F. y J. W. Tukey, *Data Analysis and Regression: A Second Course in Statistics*, Addison-Wesley, Reading, Mass., 1977.

Moussa-Hamouda, E. y F. C. Leone, "The 0-blue estimators for complete and censored samples in linear regression", *Technometrics*, **16**, 1974, pp. 441-446.

Moussa-Hamouda, E. y F. C. Leone, "The robustness of efficiency of adjusted trimmed estimators in linear regression", *Technometrics*, **19**, 1977a, pp. 19-34.

Moussa-Hamouda, E. y F. C. Leone, "Efficiency of ordinary least squares from trimmed and Winsorized samples in linear regression", *Technometrics*, **19**, 1977b, pp. 265-273.

Mullet, G. M., "Why regression coefficients have the wrong sign", *J. Qual. Technol.*, **8**, 1976, pp. 121-126.

Myers, R. H., *Classical and Modern Regression with Applications*, segunda edición, PWS-Kent Publishers, Boston, 1990.

Myers, R. H. y D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley, Nueva York, 1995.

Myers, R. H. y D. C. Montgomery, "A tutorial on generalized linear models", *Journal of Quality Technology*, **29**, 1997, pp. 274-291.

Narula, S. y J. S. Ramberg, Carta al editor, *Am. Stat.*, **26**, 1972, p. 42.

Naszódi, L. J., "Elimination of the bias in the course of calibration", *Technometrics*, **20**, 1978, pp. 201-205.

Nelder, J. A. y R. W. M. Wedderburn, "Generalized linear models", *J. R. Stat. Soc. Ser. A*, **153**, 1972, pp. 370-384.

Neter, J., M. H. Kuther, C. J. Nachtsheim y W. Wasserman, *Applied Linear Statistical Models*, cuarta edición, Richard D. Irwin, Homewood, Ill, 1996.

Neyman, J. y E. L. Scott, "Correction for bias introduced by a transformation of variables", *Ann. Math. Stat.*, **31**, 1960, pp. 643-655.

Obenchain, R. L., "Ridge analysis following a preliminary test of the shrunken hypothesis", *Technometrics*, **17**, 1975, pp. 431-441.

Obenchain, R. L., "Classical *F*-tests and confidence intervals for ridge regression", *Technometrics*, **19**, 1977, pp. 429-439.

Ott, R. L. y R. H. Myers, "Optimal experimental designs for estimating the independent variable in regression", *Technometrics*, **10**, 1968, pp. 811-823.

Pearson, E. S. y H. O. Hartley, *Biometrika Tables for Statisticians*, vol. 1, tercera edición, Cambridge University Press, Londres, 1966.

Peixoto, J. L., "Hierarchical variable selection in polynomial regression models", *Am. Stat.*, **41**, 1987, pp. 311-313.

Peixoto, J. L., "A property of well-formulated polynomial regression models", *Am. Stat.*, **44**, 1990, pp. 26-30. (Véase también **45**, 1991, p. 82.)

Pena, D. y V. J. Yohai, "The detection of influential subsets in linear regression by using an influence matrix", *J. R. Stat. Soc. Ser. B*, **57**, 1995, pp. 145-156.

Perng, S. K. y Y. L. Tong, "A sequential solution to the inverse linear regression problem", *Ann. Stat.*, **2**, 1974, pp. 535-539.

Pesaran, M. H. y L. J. Slater, *Dynamic Regression: Theory and Algorithms*, Halsted Press, Nueva York, 1980.

http://carlos2524.jimdo.com/

Pfaffenberger, R. C. y T. E. Dielman, "A comparison of robust ridge estimators", *Business and Economics Section Proceedings of the American Statistical Association*, 1985, pp. 631-635.

Poirier, D. J., "Piecewise regression using cubic splines", *J. Am. Stat. Assoc.*, **68**, 1973, pp. 515-524.

Poirier, D. J., "On the use of bilinear splines in economics", *J. Econ.*, **3**, 1975, pp. 23-24.

Pope, P. T. y J. T. Webster, "The use of an *F*-statistic in stepwise regression procedures", *Technometrics*, **14**, 1972, pp. 327-340.

Pukelsheim, F., *Optimum Design of Experiments*, Chapman & Hall, Londres, 1995.

Ramsay, J. O., "A comparative study of several robust estimates of slope, intercept, and scale in linear regression", *J. Am. Stat. Assoc.*, **72**, 1977, pp. 608-615.

Rao, P., "Some notes on misspecification in regression", *Am. Stat.*, **25**, 1971, pp. 37-39.

Ripley, B. D., "Statistical ideas for selecting network architectures", en B. Kappen and S. Grielen (eds.), *Neural Networks: Artificial Intelligence and Industrial Applications*, Springer-Verlag, Berlín, 1994, pp. 183-190.

Rocke, D. M. y D. L. Woodruff, "Identification of outliers in multivariate data", *J. Am. Stat. Assoc.*, **91**, 1996, pp. 1047-1061.

Rosenberg, S. H. y P. S. Levy, "A characterization on misspecification in the general linear regression model", *Biometrics*, **28**, 1972, pp. 1129-1132.

Rousseeuw, P. J., "Least median of squares regression", *J. Am. Stat. Assoc.*, **79**, 1984, pp. 871-880.

Rousseeuw, P. J., "Robust estimation and identifying outliers", en H. M. Wadsworth (ed.), *Handbook of Statistical Methods for Engineers and Scientists*, McGraw-Hill, Nueva York, Capítulo 17, 1998.

Rousseeuw, P J. y A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, Nueva York, 1987.

Rousseeuw, P. J. y B. L. van Zomeren, "Unmasking multivariate outliers and leverage points", *J. Am. Stat. Assoc.*, **85**, 1990, pp. 633-651.

Rousseeuw, P. J. y V. Yohai, "Robust regression by means of S-estimators", en J. Franke, W. Härdle y R. D. Martin (eds.), *Robust and Nonlinear Time Series Analysis: Lecture Notes in Statistics*, vol. 26, Springer, Berlín, 1984, pp. 256-272.

Ryan, T. P., *Modern Regression Methods*, Wiley, Nueva York, 1997.

SAS Institute, *SAS Views: SAS Principles of Regression Analysis*, SAS Institute, Cary, N.C., 1987.

Schatzoff, M., R. Tsao y S. Fienberg, "Efficient calculation of all possible regressions", *Technometrics*, **10**, 1968, pp. 769-779.

Scheffé, H., "A method for judging all contrasts in the analysis of variance", *Ann. Math. Stat.*, **40**, 1953, pp. 87-104.

Scheffé, H., *The Analysis of Variance*, Wiley, Nueva York, 1959.

Scheffé, H., "A statistical theory of calibration", *Ann. Stat.*, **1**, 1973, pp. 1-37.

Schilling, E. G., "The relationship of analysis of variance to regression. Part I. Balanced designs", *J. Qual. Technol.*, **6**, 1974a, pp. 74-83.

Schilling, E. G., "The relationship of analysis of variance to regression. Part II. Unbalanced designs", *J. Qual. Technol.*, **6**, 1974b, pp. 146-153.

http://carlos2524.jimdo.com/

Sclove, S. L., "Improved estimators for coefficients in linear regression", *J. Am. Stat. Assoc.*, **63**, 1968, pp. 596-606.

Searle, S. R., *Linear Models*, Wiley, Nueva York, 1971.

Searle, S. R. y J. G. Udell, "The use of regression on dummy variables in market research", *Manage. Sci. B*, **16**, 1970, pp. 397-409.

Seber, G. A.F., *Linear Regression Analysis*, Wiley, Nueva York, 1977.

Sebert, D. M., D. C. Montgomery y D. A. Rollier, "A clustering algorithm for identifying multiple outliers in linear regression", *Comput. Stat. Data Anal.*, **27**, 1998, pp. 461-484.

Sielken, R. L., Jr. y H. O. Hartley, "Two linear programming algorithms for unbiased estimation of linear models", *J. Am. Stat. Assoc.*, **68**, 1973, pp. 639-641.

Silvey, S. D., "Multicollinearity and imprecise estimation", *J. R. Stat. Soc. Ser. B*, **31**, 1969, pp. 539-552.

Simpson, D. G., D. Ruppert y R. J. Carroll, "On one-step *GM*-estimates and stability of inference in linear regression", *J. Am. Stat. Assoc.*, **87**, 1992, pp. 439-450.

Simpson, J. R. y D. C. Montgomery, "A biased-robust regression technique for the combined-outlier multicollinearity problem", *J. Stat. Comput. Simul.*, **56**, 1996, pp. 1-22.

Simpson, J. R. y D. C. Montgomery, "A robust regression technique using compound estimation", *Nav. Res. Logistics*, **45**, 1998a, pp. 125-139.

Simpson, J. R. y D. C. Montgomery, "The development and evaluation of alternative generalized M-estimation techniques", *Commun. Stat. Simul. Comput.*, **27**, 1998b, pp. 999-1018.

Simpson, J. R. y D. C. Montgomery, "A performance-based assessment of robust regression methods", *Commun. Stat. Simul. Comput.*, **27**, 1998c, pp. 1031-1099.

Smith, A. F. M. y M. Goldstein, "Ridge regression: Some comments on a paper of Conniffe and Stone", *The Statistician*, **24**, 1975, pp. 61-66.

Smith, B. T., J. M. Boyle, B. S. Garbow, Y. Ikebe, V. C. Klema y C. B. Moler, *Matrix Eigensystem Routines*, Springer-Verlag, Berlín, 1974.

Smith, G. y F. Campbell, "A critique of some ridge regression methods (with discussion)", *J. Am. Stat. Assoc.*, **75**, 1980, pp. 74-103.

Smith, J. H., "Families of transformations for use in regression analysis", *Am. Stat.*, **26**(3), 1972, pp. 59-61.

Smith, P. L., "Splines as a useful and convenient statistical tool", *Am. Stat.*, **33**(2), 1979, pp. 57-62.

Snee, R. D., "Some aspects of nonorthogonal data analysis, Part I. Developing prediction equations", *J. Qual. Technol.*, **5**, 1973, pp. 67-79.

Snee, R. D., "Validation of regression models: Methods and examples", *Technometrics*, **19**, 1977, pp. 415-428.

Sprent, P., *Models in Regression and Related Topics*, Methuen, Londres, 1969.

Sprent, P. y G. R. Dolby, "The geometric mean functional relationship", *Biometrics*, **36**, 1980, pp. 547-550.

Staudte, R. G. y S. J. Sheather, *Robust Estimation and Testing*, Wiley, Nueva York, 1990.

Stefanski, W., "A note on high-breakdown estimators", *Stat. Probab. Lett.*, **11**, 1991, pp. 353-358.

http://carlos2524.jimdo.com/

Stefansky, W., "Rejecting outliers by maximum normed residual", *Ann. Math. Stat.*, **42**, 1971, pp. 35-45.

Stefansky, W., "Rejecting outliers in factorial designs", *Technometrics*, **14**, 1972, pp. 469-479.

Stein, C., "Multiple regression", en I. Olkin (ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, Stanford, Calif., 1960.

Stewart, G. W., *Introduction to Matrix Computations*, Academic, Nueva York, 1973.

Stone, M., "Cross-validating choice and assessment of statistical predictions (with discussion)", *J. R. Stat. Soc. Ser. B*, **36**, 1974, pp. 111-147.

Stromberg, A. J., "Computation of high breakdown nonlinear regression parameters", *J. Am. Stat. Assoc.*, **88**, 1993, pp. 237-244.

Stromberg, A. J. y D. Ruppert, "Breakdown in nonlinear regression", *J. Am. Stat. Assoc.*, **83**, 1989, pp. 991-997.

Suich, R. y G. C. Derringer, "Is the regression equation adequate–one criterion", *Technometrics*, **19**, 1977, pp. 213-216.

Theil, H. y A. S. Goldberg, "On pure and mixed statistical estimation in economics", *Int. Econ. Rev.*, **2**, 1961, pp. 65-78.

Theobald, C. M., "Generalizations of mean square error applied to ridge regression", *J. R. Stat. Soc. Ser. B*, **36**, 1974, pp. 103-106.

Thompson, M. L., "Selection of variables in multiple regression: Part I. A review and evaluation", *Int. Stat. Rev.*, **46**, 1978a, pp. 1-19.

Thompson, M. L., "Selection of variables in multiple regression: Part II. Chosen procedures", computations and examples, *Int. Stat. Rev.*, **46**, 1978b, pp. 129-146.

Tucker, W. T., "The linear calibration problem revisited", presentado en ASQC Fall Technical Conference, Cincinnati, Ohio, 1980.

Tufte, E. R., *Data Analysis for Politics and Policy*, Prentice-Hall, Englewood Cliffs, N.J., 1974.

Tukey, J. W., "On the comparative anatomy of transformations", *Ann. Math. Stat.*, **28**, 1957, pp. 602-632.

Wagner, H. M., "Linear programming techniques for regression analysis", *J. Am. Stat. Assoc.*, **54**, 1959, pp. 206-212.

Wahba, G., G. H. Golub y C. G. Health, "Generalized cross-validation as a method for choosing a good ridge parameter", *Technometrics*, **21**, 1979, pp. 215-223.

Walker, E., *Influence, Collinearity, and Robust Estimation in Regression*, disertación doctoral, Department of Statistics, Virginia Tech, Blacksburg, 1984.

Walker, E. y J. B. Birch, "Influence measures in ridge regression", *Technometrics*, **30**, 1988, pp. 221-227.

Walls, R. E. y D. L. Weeks, "A note on the variance of a predicted response in regression", *Am. Stat.*, **23**, 1969, pp. 24-26.

Webster, J. T., R. F. Gunst y R. L. Mason, "Latent root regression analysis", *Technometrics*, **16**, 1974, pp. 513-522.

Weisberg, S., *Applied Linear Regression*, segunda edición, Wiley, Nueva York, 1985.

http://carlos2524.jimdo.com/

Welsch, R. E., "Confidence regions for robust regression", *Statistical Computing Section Proceedings of the American Statistical Association*, Washington, D.C., 1975.

Welsch, R. E. y S. C. Peters, "Finding influential subsets of data in regression models", en A. R. Gallant y T. M. Gerig (eds.), *Proceedings of the Eleventh Interface Symposium on Computer Science and Statistics*, Institute of Statistics, North Carolina State University, 1978, pp. 240-244.

White, J. W. y R. F. Gunst, "Latent root regression: Large sample analysis", *Technometrics*, **21**, 1979, pp. 481-488.

Wichern, D. W. y G. A. Churchill, "A comparison of ridge estimators", *Technometrics*, **20**, 1978, pp. 301-311.

Wilcox, R. R., *Introduction to Robust Estimation and Hypothesis Testing*, Academic, Nueva York, 1997.

Wilde, D. J. y C. S. Beightler, *Foundations of Optimization*, Prentice-Hall, Englewood Cliffs, N. J., 1967.

Wilkinson, J. W., *The Algebraic Eigenvalue Problem*, Oxford University Press, Londres, 1965.

Willan, A. R. y D. G. Watts, "Meaningful multicollinearity measures", *Technometrics*, **20**, 1978, pp. 407-412.

Williams, D. A., Carta al editor, *Appl. Stat.*, **22**, 1973, pp. 407-408.

Williams, E. J., "A note on regression methods in calibration", *Technometrics*, **11**, 1969, pp. 189-192.

Wold, S., "Spline functions in data analysis", *Technometrics*, **16**, 1974, pp. 1-11.

Wonnacott, R. J. y T. H. Wonnacott, *Econometrics*, Wiley, Nueva York, 1970.

Wood, F. S., "The use of individual effects and residuals in fitting equations to data", *Technometrics*, **15**, 1973, pp. 677-695.

Working, H. y H. Hotelling, "Application of the theory of error to the interpretation of trends", *J. Am. Stat. Assoc. Suppl. (Proc.)*, **24**, 1929, pp. 73-85.

Wu, C. F. J., "Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion)", *Ann. Stat.*, **14**, 1986, pp. 1261-1350.

Yale, C. y A. B. Forsythe, "Winsorized regression", *Technometrics*, **18**, 1976, pp. 291-300.

Yohai, V. J., "High breakdown-point and high efficiency robust estimates for regression", *Ann. Stat.*, **15**, 1987, pp. 642-656.

Younger, M. S., *A Handbook for Linear Regression*, Duxbury Press, North Scituate, Mass., 1979.

Zellner, A., *An Introduction to Bayesian Inference in Econometrics*, Wiley, Nueva York, 1971.

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

ÍNDICE

Adecuación del modelo en la regresión, 5, 15, 26, 143, 155, 266, 477
Algoritmo DUPLEX, 483
Análisis
de conglomerados, 195
de covarianza, 249
de eigensistema, 303, 304
de residuales, 24, 117-118, 261. *Véase también* Comprobación de adecuación de modelo; Gráficas de resiuales de varianza, 27, 72, 79, 251, 258
de varianza, clasificación de una vía, 251
Autocorrelación, 131, 441, 442

Bondad de ajuste y regresión, 141
Bootstrap, casos, 460
Bootstrap, pares, 460
Boootstrap, procedimientos en regresión, 394, 459-466

Cadena
canónica, 421, 561
de identidad, 415
inversa, 421
logarítmica, 415, 422, 424
logística, 422
CART (árboles de clasificación y regresión), 466
Causa y efecto en regresión, 4, 8, 39
Centrado de datos, 23, 198, 303, 306, 307, 338
Cociente de ventaja, 407
Códigos asignados como regresores, 249
Coeficiente
de correlación muestral, 50
de determinación múltiple, *véase* R^2
de determinación, *véase* R^2
de regresión estandarizados, 102, 104
de variación (CV), 54
Coeficientes de regresión, 13, 62
parcial, 62
Colinealidad, 7. *Véase también* Multicolinealidad
Columnas ortogonales en la matriz **X**, 87
Comparación de modelos de regresión, 137, 248, 275
Comprobación de la adecuación de modelo, 5, 477. *Véase también* Premisas del modelo de regresión; Gráficas de residuales
Compromiso de Marquardt, 407
Correlación, 49
Correlación parcial, 278
Correlación serial, *véase* Autocorrelación
Corridas de confirmación y validación de modelo, 479, 491
Corridas repetidas comparadas con réplicas, 142
Covarianza, 20, 49, 74
de y y de β_1, 23
$COVRATIO_i$, 193
Cuadrado medio de residuales, 22, 75, 267
Cuadrados mínimos restringidos, 92, 115
Curvas SPLINE, 203-205, 209

Dependencia casi lineal en regresores, 106. *Véase también* Multicolinealidad.
Derivadas matriciales, 534
Descenso más pronunciado, 389, 470
Descomposición de varianza, proporciones de, 306
Descomposición en valores singulares, 305
Descomposición **QR**, 543
Descripción de datos, 10
Desviación, 408, 409
Deterioramiento, 198, 270
Determinante de **X'X**, 307, 472

http://carlos2524.jimdo.com/

DFBETAS$_{j,i}$, 191, 554
DFFITS$_i$, 191, 553
Diagnóstico de eliminación, 191
Diagrama de dispersión, 1, 2
Diseño central compuesto, 219
Diseño *G* óptimo, 472
Diseños *D* óptimo, 472
Distribución
  exponencial doble, 346
  F no central, 28, 79, 531
  F, 531
  ji cuadrada, 530
  ji cuadrada no central, 531, 536
  normal bivariada, 49
  t, 530
  t no central, 531
Distribuciones conjuntas, 49, 50
División de datos, 482

Ecuación de regresión, examen de. *Véase* Adecuación del modelo, comprobación de; Gráficas de residuales
Ecuación óptima de regresión, 261
Ecuaciones normales, 14, 42, 66, 67, 376
Ecuaciones normales de mínimos cuadrados, *véase* Ecuaciones normales
Eficiencia asintótica, 359
Eficiencia de un estimador, 358
Eigenvalores, 303
Eigenvectores, 303
Eliminación hacia atrás, 199, 278, 280, 310
Error
  absoluto, minimización de la suma de, 347
  cuadrático medio de un estimador, 312
  estándar, 24, 25, 92
  estándar de la regresión, 22
  puro, 142, 224
Errores en regresores, 452
Escalamiento de variables de longitud unitaria, 103, 106
Escalamiento normal unitario de variables, 103
Escalamiento y centrado, 303, 306
Estabilización de la varianza, 156
Estadística de Durbin-Watson, 443, 446, 447
Estadísticas suficientes, 48
Estadístico C_p, 268, 310, 321
Estimación
  de (suma de) mínimos cuadrados recortada, 362
  de Bayes, estimador ridge obtenido con, 319
  de error con vecinos cercanos, 145. *Véase también* Error puro
  de mínima mediana de cuadrados, estándar, 24, 25, 92 360
  de mínimos cuadrados, 15, 65, 375
  de parámetro en modelos de regresión, 5, 8-9, 14, 21-22, 41, 46, 65, 74, 311, 319, 327, 332-333, 343, 345, 349, 351, 362-365, 374, 378, 382, 385, 389, 403, 415, 423, 446, 447, 454, 555, 561
  L, 365
  mezclada, 320
  R, 365
  sesgada en regresión, 311, 327, 328, 330, 331, 332, 333, 336
Estimado de error dependiente del modelo, 22, 75
Estimado de error independiente del modelo, 141, 145
Estimador, 15
  de contracción, 319
  insesgado de varianza mínima, 48, 74
  lineal insesgado óptimo, 20, 74, 312, 540, 550
Estimadores
  compuestos de regresión, 365
  consistentes, 48
  de punto de quiebre alto, 358, 359
  de rango fraccionados, 336
  GM, 364
  insesgados, 19, 22, 48, 74, 75, 176, 312
  M, 347, 357
  MM, 363
  S, 363
Estudio en retrospectiva en la regresión, 6, 7, 14
Estudio observacional en una regresión, 6, 8, 14, 65
Experimentos diseñados, 6, 8, 65, 218, 251, 459, 470, 491

http://carlos2524.jimdo.com/

$F = t^2$ en regresión lineal simple, 30
F y R^2, 114
Factor de inflación de varianza (VIF), 107, 302, 478
Falta de ajuste, 141, 145
  suma de cuadrados, 142
Familia de distribuciones exponenciales, 399, 420, 559
F_{IN}, 278, 281, 283
Forma
  cuadrática, 532
  de correlación del modelo de regresión, 291-292
F_{OUT}, 280, 281, 283
Función
  cadena, 415, 421
  de cuasi-probabilidad, 430
  de puntuación, 365, 556
  de respuesta logística, 401
  del valor esperado, 374
  influencia, 349, 350

Generación de subconjuntos de modelos de regresión, 271, 277, 280
Geometría de lineales cuadrados mínimos, 72
Geometría de mínimos cuadrados no lineales, 376, 388
Gompertz, modelo de crecimiento, 394
Gráfica de curvas de nivel, 61, 225, 226
Gráfica de variable agregada, 131
Gráficas de probabilidad normal de residuales, 124
  de regresión parcial, 131, 132
  de residuales, 123-124, 126, 128-129, 133-134, 429, 442
  de residuales parciales, 130, 131

Hipótesis lineal, 89-92

Incrementos fraccionados, 389
Índices de condición, 303, 305
Inferencia simultánea en regresión, 35, 94-95
Influencia, 185, 186, 261, 345
Interacción, 63, 242, 246, 248
Intervalo
  de confianza de Bonferroni, 97
  de confianza de la respuesta media, 32, 42, 43
  de predicción, 35, 37, 42
Intervalos de confianza de los coeficientes de regresión, 30, 42, 92, 94

Jerarquía, 200

Linealización, 382
Líneas segmentadas, véase Regresión por segmentos; Curvas SPLINE

Matriz
  alias, 262
  de correlación, 103, 292, 299
  de covarianza de coeficientes de regresión, 74
  de diagrama de dispersión, 68, 76
  de información, 439
  de sombrero, 68, 99, 187
  hessiana, 413
  idempotente, 119, 532
  rango de, 531, 533
  traza de una, 533
  X, 67
Máxima verosimilitud, 46, 74, 77, 346, 347, 378, 403, 415, 422, 556
  función de, 47, 77, 346, 373, 403, 408, 415
  pruebas del cociente de, 410, 416
Medida de la distancia de Cook, 189, 194, 554
Método
  de Box-Cox, 166
  de Box-Tidwell, 170
  de Cochrane-Orcutt, 447
  de Gauss-Newton, 382
  de suma extra de cuadrados, 83, 84, 203, 255, 539
  Newton-Raphson, 556
  S de Scheffé para intervalos de confianza conjunta, 97
  t de módulo máximo para intervalos de confianza conjuntos, 98
Mínimos cuadrados
  generalizados, 173, 550
  limitados, véase Mínimos cuadrados restringidos
  ponderados, 173, 174-175. *Véase también* Mínimos cuadrados

http://carlos2524.jimdo.com/

ponderados iterativamente
ponderados iterativamente, 348, 404, 422
Minitab, 19, 27, 29, 277
Modelo
completo, 84, 87, 89, 248, 255, 262, 409
de crecimiento logístico, 394
de red neural, 468-470
de regresión global, propiedades del, 117
de regresión lineal simple, 1, 13
de regresión poblacional y modelo correspondiente de regresión muestral, 14, 65
de regresión sin intercepto. *Véase* Regresión por el origen
de regresión útil, 143
de regresión, términos trigonométricos en un, 211
de regresión, validación de un, 285, 477
lineal, 1-2, 61, 63, 197
mecanístico, 3, 200, 394
parsimonioso, 198
reducido, 84, 87, 89, 248, 256, 409
sobredefinido, 293
Modelos
cuadráticos, *véase* Modelos de segundo orden
de regresión no lineal, 367, 373, 374, 414, 470
de regresión, 1, 13, 41, 61, 197, 373, 374, 399
de segundo orden, 63, 197
empíricos, 3, 62
intrínsecamente lineales, 159, 379
lineales generalizados, 399, 403, 420, 426, 555, 561
Muestra Bootstrap, 459
Multicolinealidad, 7, 106, 270, 291, 479
en la regresión, fuentes de, 292

No linealidad, 127, 159, 162
Nudo, 204
Número de condición, 303

Observación influyente conjunta, 194
Observaciones influyentes, 39, 121, 185, 189, 194, 261
influyentes, grupos de, 194
Orden de un polinomio, 198

Papel de la computadora, 10-11
Parámetro de no centralidad, 28, 79, 531, 536, 538
Parámetros del modelo, 5, 13, 14, 61
Polinomios ortogonales, 226, 228-229
precauciones en la interpretación, 37-38
Predicción y regresión, 9-10, 35, 42, 143, 285, 297-299, 310-311
Predictor lineal, 402, 421
Premisas del análisis de regresión, *véase* Premisas del modelo de regresión
Premisas del modelo de regresión, 1, 13, 23, 47, 64, 78, 117, 155
PRESS Estadístico, 136, 137, 270, 322, 598
PRESS Residuales, 120
Probit, modelo de regresión, 403
Problema de calibración, 454
Problema en la construcción del modelo, 1, 82, 199, 261
Proceso autorregresivo de primer orden, 442, 446
Propiedades asintóticas de estimados de parámetro, 48, 359, 362, 386, 392
Propiedades de estimadores, 19-20, 20-21, 22-23, 48, 74, 78, 174, 176, 262-265, 360, 361, 362, 392, 441, 540, 550, 552
Prueba
con regresores ortogonales, 87
de hipótesis
de valores atípicos basada en *R* de Student, 121, 547
del coeficiente de correlación, 49
en modelos lineales generalizados, 408, 412
en parámetros individuales en regresión múltiple, 82
en pendiente e intercepto al origen en regresión lineal simple, 24-25
en regresión, 23-24
en subconjuntos de parámetros, 83, 86
en términos polinomiales, 85, 198, 202, 203, 220, 221
lineal general, 89-92
significación de la regresión, 25, 27, 79

http://carlos2524.jimdo.com/

F, 28, 79, 538
F parcial, 85
t, 24, 82
Pruebas estadísticas para los residuales, 136
Punto de quiebre, 358

R de Student, 121, 137, 546-547
R^2, 37, 44, 82, 265-266
R^2, ajustada, véase R^2_{Aj}.
$R_{Aj}{}^2$, 82, 266-267
Recolección de datos en regresión, 6, 292
Reducción del peso de observaciones, 195. *Véase también* Regresión robusta.
Región de confianza conjunta, 95
Región de confianza de los coeficientes de regresión, 95
Regresión
bipeso, 369
de componentes principales, 330-331, 332-333
de influencia acotada, 364
de línea recta, 1, 2. *Véase también* Modelo de regresión lineal simple
de Poisson, 414
de raíz latente, 336
extrapolación en, 5, 34, 38, 97, 199, 299
extrapolación oculta en, 99
inversa y calibración, 458
kernel, 212
lineal múltiple, modelo de, 4, 61
localmente ponderada (loess), 213, 215
logística, 399, 403, 405, 555
no paramétrica, 211-213, 215-216
polinomial, 62, 197, 217-218
por el origen, 41
por segmentos, 203, 204, 209-210, 257, 278, 280-281
prueba de falta de ajuste en, 141, 220
ridge, 311-312, 313, 317-319, 324, 366
robusta, 124, 195, 343, 344, 347, 357, 359, 360, 362, 363, 365-367
sin modelo, *véase* Regresión no paramétrica y causalidad, *véase* Causa y efecto en regresión y programación lineal, 347
Regresores
aleatorios, 48-65
candidatos, 261
linealmente independientes, 67
ortogonales, 108, 291, 327
y respuesta con distribución conjunta, 48
Remuestreo, *véase* procedimientos Bootstrap en regresión
Residual, 15, 68, 117-118
Residuales
de desviación, 428
escalados, 118-123, 137
estandarizados, 118
estudentizados, 118-119, 121, 137
Ridge, regresión generalizada, 327, 328, 330
Rotabilidad, 225, 472
$R_{predic}{}^2$, 137

SAS PROC GENMOD, 411, 412, 417, 419, 424, 471
SAS PROC NLIN, 391
SAS PROC REG, 53, 72, 73, 81, 278-279, 280-281, 282
Selección de variable, 82, 261, 271, 278, 279, 281-284, 309, 324
Selección de variables en regresión, véase Problema en la construción del modelo; Selección de variable
Selección de variables hacia adelante, 199, 278, 309
Selección empírica de una transformación, 156, 166
Sesgo en estimación de regresión, 263, 264, 268, 551
Significación de la regresión, prueba de hipótesis, 25, 27, 79
Signos incorrectos en coeficientes de regresión, 108, 308
Sobreajuste de un modelo, 82, 470
Sobredispersión, 429
S-PLUS, 215, 216, 367, 368
STATGRAPHICS, 391
Suma
de cuadrados de error, *véase* Error puro; Suma de cuadrados residuales
de cuadrados de regresión, 28, 79, 80, 535
de cuadrados de residuales, 21, 28, 79, 80, 535, 538

http://carlos2524.jimdo.com/

de cuadrados del modelo, 28. *Véase* también Suma de cuadrados de la regresión
Superficie de respuesta, 218, 223-225
metodología de, 218, 472

Teorema de Gauss-Markoff, 20, 74, 312, 540, 550
Teorema del límite central, 530
Término de error en un modelo de regresión, 1, 13, 441
Todas las regresiones posibles, 271, 277, 310
Transformación
arco seno, 156
arco tanh, 52
de raíz cuadrada, 156
fuerza de una, 156
inversa, 156. *Véase también* Transformación recíproca
logarítmica, 156, 161
logit, 403
log-log complementaria, 403
recíproca, 156
Transformaciones, 127, 155, 156, 158-159, 161-162, 198-199, 217, 378-379
Transformaciones de estabilización de la varianza, 127, 156
Traza *ridge,* 314, 327

Validación cruzada, 482, 485
Valor atípico
de regresión, 344
en el espacio *Y*, 345
en el espacio *X*, 345
en espacio *X* y *Y*, 345
residual, 345
Valores
ajustados, 15, 16, 18, 68
atípicos, 7, 39, 118, 125, 137-141, 345
de corte para diagnóstico, 192
singulares, 305
Variable
binaria, 399
de respuesta, 1
dependiente, 1
independiente, 1
indicadora, 237, 238, 241, 245, 249-251
Variables
categóricas en regresión, 237. *Véase también* Variables indicadoras.
cualitativas, *véase* Variables categóricas en regresión
ficticias, 237. *Véase también* Variables indicadoras
predictoras, 1. *Véase también* Variables regresoras.
regresoras, 1, 13, 48
Varianza, 3, 13, 64, 74
de la respuesta pronosticada, 32-33,93, 224
del error de predicción, 35
generalizada, 193
Varianzas desiguales, 173, 174. *Véase también* Estabilización de la varianza; Transformaciones
Vector de incrementos, 382
Ventaja, 403
Violaciones en los supuestos de la regresión, 117, 129, 156-157

Wald, inferencia de, 413, 416
Weibull, modelo de crecimiento, 395

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

Esta obra se terminó de imprimir en abril del 2006
en los talleres de Overprint, S. A. de C. V.
Agustín Yáñez 1253, Col. Sector Popular
C.P. 09060, México, D.F.

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

http://carlos2524.jimdo.com/

Introducción al Análisis de Regresión Lineal

Esta técnica estadística para investigar y modelar la relación entre variables tiene aplicaciones en casi todos los campos, pues se emplea tanto en las ingenierías, las ciencias físicas y químicas, la economía y la administración como en las ciencias sociales y biológicas.

En esta obra se presenta un claro equilibro entre la teoría y las aplicaciones de la regresión lineal. A lo largo del texto, se describen los usos más convencionales de la técnica, así como las menos comunes. Al principio se presenta una introducción general al modelado de la regresión, incluyendo aplicaciones típicas. En los capítulos sucesivos, el lector encontrará una cobertura detallada de:

- Selección de variable y técnicas de construcción de modelos.
- Métodos de adecuación del modelo y de transformación de variables
- El problema de multicolinealidad, incluyendo sus fuentes, efectos perjudiciales, diagnósticos y medidas correctivas.
- Modelos de regresión no lineales, de Poisson.
- Variables indicadoras que hacen la conexión entre modelos de regresión y análisis de varianza.
- Técnicas de regresión robusta, como los estimadores de M, la mediana mínima de cuadrados y la estimación S.
- Modelos lineales generalizados.

Se presentan también modelos de regresión con errores autocorrelacionados, estimados de regresión de secuencia de llamada, árboles de clasificación y regresión, así como validación del modelo de regresión. Además incluye temas que por lo general no se encuentran en un texto de regresión lineal como *bootstrapping* y redes neuronales. Se detallan procedimientos analíticos por computadora, disponibles en paquetes de software como SAS, Minitab y S-Plus.

En esta obra el lector encontrará el texto más completo y actualizado del análisis de regresión, la técnica más ampliamente usada en la estadística de nuestros días.

ISBN 970-24-0327-8

9 789702 403272

GRUPO PATRIA CULTURAL

COMPAÑÍA EDITORIAL CONTINENTAL