

---

# MODELOS LINEALES

---

**Francesc Carmona Pontaque**

Electronic-University Mathematical Books

Consejo editor:

T. Aluja  
M.J. Bayarri  
F. Carmona  
C.M. Cuadras (coordinador)  
F.R. Fernández  
J. Fortiana  
G. Gómez  
W. González-Manteiga  
M.J. Greenacre  
J.M. Oller  
J. Puerto  
A. Satorra

e-UMAB

Electronic-University Mathematical Books

© EDICIONS DE LA UNIVERSITAT DE BARCELONA, 2004

Copia impresa del libro electrónico con ISBN: XX-XXXX-XXX-X

D.L.: B-XX.XXX-2004

Impresión: Gráficas Rey, S.L.

Impreso en España / Printed in Spain

*A la meva esposa Carme  
i els nostres fills Mireia i Guillem.*

*"Soñemos con un mundo unido  
sin ninguna otra soberanía  
que la del pueblo universal.  
No hacer daño nunca, nunca, a nadie."*

José María de Llanos (Padre Llanos)



---

# Presentación

La teoría y aplicaciones de los modelos lineales ocupan un papel fundamental en la Estadística. Tales modelos engloban la regresión simple, múltiple y polinómica, el análisis de la varianza, el diseño de experimentos, el estudio de curvas de crecimiento, los modelos log-lineales, y algunos contrastes sobre medias como caso particular. Basta consultar revistas especializadas como *Biometrics*, para comprobar que muchos problemas de estadística aplicada se pueden enfocar linealmente, siguiendo la omnipresente ecuación: Observación = Modelo + Error.

Algunos han creído que por el hecho de ser el modelo “lineal”, su tratamiento era más bien fácil. En realidad es todo lo contrario. Este tipo de modelo, que se adecua tan bien a la naturaleza, exige un estudio riguroso y posee múltiples facetas que por sí sólo constituye una especialidad en Estadística.

La obra de mi compañero y amigo Francesc Carmona, que hace más de veinticinco años fue un destacado alumno mío, nace precisamente de las clases que sobre el mismo tema impartí en la Facultad de Matemáticas de la Universidad de Barcelona, y que él continuó, ampliando y mejorando la materia. Diversos profesores editamos entonces unos apuntes, que luego ampliamos y publicamos dentro de la colección Publicaciones de Bioestadística y Biomatemática, editada por el Departamento de Estadística. Hacía falta convertir estos apuntes en un libro de verdad, una labor que ha sido llevada a cabo con entusiasmo por Francesc Carmona, consiguiendo una visión ampliada, moderna y mejorada del anterior material didáctico.

Me complace enormemente presentar el libro **Modelos Lineales**, editado en la colección e-UMAB de EUB, por estar muy bien escrito y documentado y ser muy completo. En efecto, además de contener los temas clásicos, incluye los modelos no paramétricos, el análisis de residuos, numerosos ejemplos ilustrativos, instrucciones en el lenguaje de programación R y adecuados hipervínculos. Esta obra es una contribución didáctica de alto nivel, que será de gran utilidad para investigadores, profesores y alumnos de Estadística.

Dr. Carles M. Cuadras  
[ccuadras@ub.edu](mailto:ccuadras@ub.edu)



---

# Prólogo

Las páginas que siguen constituyen una parte de las exposiciones teóricas y prácticas de asignaturas que se han impartido a lo largo de algunos años en varias licenciaturas y cursos de doctorado. En particular en la licenciatura de Matemáticas, la licenciatura de Biología y la diplomatura de Estadística de la Universidad de Barcelona. Se ha intentado un cierto equilibrio entre las explicaciones teóricas y los problemas prácticos. Sin embargo, nuestra intención siempre ha sido fundamentar sólidamente la utilización de los modelos lineales como base de las aplicaciones de la regresión, el análisis de la varianza y el diseño de experimentos. Por ello, en este libro la base matemática y estadística es considerable y creemos importante la correcta definición de los conceptos y la rigurosidad de las demostraciones. Una sólida base impedirá cometer ciertos errores, habituales cuando se aplican los procedimientos ciegamente.

Por otra parte, la aplicación práctica de los métodos de regresión y análisis de la varianza requiere la manipulación de muchos datos, a veces en gran cantidad, y el cálculo de algunas fórmulas matriciales o simples. Para ello es absolutamente imprescindible la utilización de algún programa de ordenador que nos facilite el trabajo. En una primera instancia es posible utilizar cualquier programa de hojas de cálculo que resulta sumamente didáctico. También se puede utilizar un paquete estadístico que seguramente estará preparado para ofrecer los resultados de cualquier modelo lineal estándar como ocurre con el paquete SPSS. En cambio, en este libro se ha optado por incluir algunos ejemplos con el programa R. Las razones son varias. En primer lugar, se trata de un programa que utiliza el lenguaje S, está orientado a objetos, tiene algunos módulos específicos para los modelos lineales y es programable. R utiliza un lenguaje de instrucciones y al principio puede resultar un poco duro en su aprendizaje, sin embargo superada la primera etapa de adaptación, su utilización abre todo un mundo de posibilidades, no sólo en los modelos lineales, sino en todo cálculo estadístico. Además, la razón más poderosa es que el proyecto R es GNU y, por tanto, de libre distribución. De modo que los estudiantes pueden instalar en su casa el programa R y practicar cuanto quieran sin coste económico alguno. Por otra parte, el paquete S-PLUS es una versión comercial con el mismo conjunto de instrucciones básicas.

El tratamiento de algunos temas tiene su origen en unos apuntes de C.M. Cuadras y Pedro Sánchez Algarra (1996) que amablemente han cedido para su actualización en este libro y a los que agradezco profundamente su colaboración. También es evidente que algunas demostraciones tienen su origen en el clásico libro de Seber [66].

Por último, este libro ha sido escrito mediante el procesador de textos científico  $\text{\LaTeX}$  y presentado en formato electrónico. Gracias a ello se puede actualizar con relativa facilidad. Se agradecerá la comunicación de cualquier errata, error o sugerencia.

Barcelona, 6 de mayo de 2004.

Dr. Francesc Carmona  
[fcarmona@ub.edu](mailto:fcarmona@ub.edu)





---

# Índice general

<b>1. Las condiciones</b>	<b>15</b>
1.1. Introducción . . . . .	15
1.2. Un ejemplo . . . . .	15
1.3. El modelo . . . . .	17
1.4. El método de los mínimos cuadrados . . . . .	18
1.5. Las condiciones de Gauss-Markov . . . . .	20
1.6. Otros tipos de modelos lineales . . . . .	21
1.7. Algunas preguntas . . . . .	21
1.8. Ejemplos con R . . . . .	22
1.9. Ejercicios . . . . .	24
<b>2. Estimación</b>	<b>27</b>
2.1. Introducción . . . . .	27
2.2. El modelo lineal . . . . .	27
2.3. Suposiciones básicas del modelo lineal . . . . .	29
2.4. Estimación de los parámetros . . . . .	30
2.5. Estimación de la varianza . . . . .	35
2.6. Distribuciones de los estimadores . . . . .	36
2.7. Matriz de diseño reducida . . . . .	38
2.8. Matrices de diseño de rango no máximo . . . . .	40
2.8.1. Reducción a un modelo de rango máximo . . . . .	40
2.8.2. Imposición de restricciones . . . . .	41
2.9. Ejercicios . . . . .	41
<b>3. Funciones paramétricas estimables</b>	<b>45</b>
3.1. Introducción . . . . .	45
3.2. Teorema de Gauss-Markov . . . . .	46
3.3. Varianza de la estimación y multicolinealidad . . . . .	50
3.4. Sistemas de funciones paramétricas estimables . . . . .	51
3.5. Intervalos de confianza . . . . .	53
3.6. Ejercicios . . . . .	54

<b>4. Complementos de estimación</b>	<b>59</b>
4.1. Ampliar un modelo con más variables regresoras	59
4.1.1. Una variable extra	59
4.1.2. Una interpretación	61
4.1.3. Más variables	63
4.2. Mínimos cuadrados generalizados	64
4.3. Otros métodos de estimación	66
4.3.1. Estimación sesgada	66
4.3.2. Estimación robusta	67
4.3.3. Más posibilidades	68
4.4. Ejercicios	68
<b>5. Contraste de hipótesis lineales</b>	<b>69</b>
5.1. Hipótesis lineales contrastables	69
5.2. El modelo lineal de la hipótesis	70
5.3. Teorema fundamental del Análisis de la Varianza	73
5.3.1. Un contraste más general	78
5.3.2. Test de la razón de verosimilitud	80
5.4. Cuando el test es significativo	81
5.5. Contraste de hipótesis sobre funciones paramétricas estimables	81
5.6. Elección entre dos modelos lineales	82
5.6.1. Sobre los modelos	82
5.6.2. Contraste de modelos	83
5.7. Ejemplos con R	85
5.8. Ejercicios	86
<b>6. Regresión lineal simple</b>	<b>91</b>
6.1. Estimación de los coeficientes de regresión	91
6.2. Medidas de ajuste	94
6.3. Inferencia sobre los parámetros de regresión	96
6.3.1. Hipótesis sobre la pendiente	96
6.3.2. Hipótesis sobre el punto de intercepción	97
6.3.3. Intervalos de confianza para los parámetros	97
6.3.4. Intervalo para la respuesta media	98
6.3.5. Predicción de nuevas observaciones	98
6.3.6. Región de confianza y intervalos de confianza simultáneos	99
6.4. Regresión pasando por el origen	99
6.5. Correlación	100
6.6. Carácter lineal de la regresión simple	101
6.7. Comparación de rectas	104
6.7.1. Dos rectas	104
6.7.2. Varias rectas	108
6.7.3. Contraste para la igualdad de varianzas	111
6.8. Un ejemplo para la reflexión	112
6.9. Ejemplos con R	115
6.10. Ejercicios	117

<b>7. Una recta resistente</b>	<b>121</b>
7.1. Recta resistente de los tres grupos	121
7.1.1. Formación de los tres grupos	121
7.1.2. Pendiente e intercepción	122
7.1.3. Ajuste de los residuos e iteraciones	123
7.1.4. Mejora del método de ajuste	127
7.2. Métodos que dividen los datos en grupos	127
7.3. Métodos que ofrecen resistencia	129
7.4. Ejercicios	131
<b>8. Regresión lineal múltiple</b>	<b>133</b>
8.1. El modelo	133
8.2. Medidas de ajuste	134
8.3. Inferencia sobre los coeficientes de regresión	136
8.4. Coeficientes de regresión estandarizados	141
8.5. Multicolinealidad	144
8.6. Regresión polinómica	145
8.6.1. Polinomios ortogonales	147
8.6.2. Elección del grado	148
8.7. Comparación de curvas experimentales	150
8.7.1. Comparación global	150
8.7.2. Test de paralelismo	151
8.8. Ejemplos con R	152
8.9. Ejercicios	156
<b>9. Diagnóstico del modelo</b>	<b>161</b>
9.1. Residuos	161
9.1.1. Estandarización interna	161
9.1.2. Estandarización externa	163
9.1.3. Gráficos	164
9.2. Diagnóstico de la influencia	166
9.2.1. Nivel de un punto	166
9.2.2. Influencia en los coeficientes de regresión	167
9.2.3. Influencia en las predicciones	168
9.3. Selección de variables	169
9.3.1. Coeficiente de determinación ajustado	169
9.3.2. Criterio $C_p$ de Mallows	169
9.3.3. Selección paso a paso	170
9.4. Ejemplos con R	170
9.5. Ejercicios	172
<b>10. Regresión robusta</b>	<b>175</b>
10.1. Minimizar una función objetivo	175
10.1.1. Funciones objetivo	176
10.2. Regresión robusta mínimo-cuadrada recortada	178

10.3. Ejemplos con S-PLUS . . . . .	179
10.4. Ejercicios . . . . .	183
<b>11. Análisis de la Varianza</b>	<b>185</b>
11.1. Introducción . . . . .	185
11.2. Diseño de un factor . . . . .	186
11.2.1. Comparación de medias . . . . .	186
11.2.2. Un modelo equivalente . . . . .	189
11.3. Diseño de dos factores sin interacción . . . . .	192
11.4. Diseño de dos factores con interacción . . . . .	198
11.5. Descomposición ortogonal de la variabilidad . . . . .	203
11.5.1. Descomposición de la variabilidad en algunos diseños . . . . .	205
11.5.2. Estimación de parámetros y cálculo del residuo . . . . .	207
11.6. Diagnóstico del modelo . . . . .	209
11.7. Diseños no balanceados y observaciones faltantes . . . . .	212
11.8. Ejemplos con R . . . . .	213
11.9. Ejercicios . . . . .	219
<b>12. Análisis de Componentes de la Varianza</b>	<b>223</b>
12.1. Introducción . . . . .	223
12.2. Contraste de hipótesis . . . . .	224
12.2.1. Los test $F$ . . . . .	226
12.2.2. Estimación de los componentes de la varianza . . . . .	227
12.3. Los modelos más sencillos . . . . .	228
12.3.1. Diseño de un factor con efectos fijos . . . . .	228
12.3.2. Diseño de un factor con efectos aleatorios . . . . .	231
12.3.3. Diseño de dos factores sin interacción con efectos fijos . . . . .	235
12.3.4. Diseño de dos factores sin interacción con efectos aleatorios . . . . .	238
12.3.5. Diseño de dos factores aleatorios con interacción . . . . .	240
12.3.6. Diseño de tres factores aleatorios y réplicas . . . . .	241
12.3.7. Diseño anidado de dos factores aleatorios . . . . .	242
12.3.8. Resumen . . . . .	244
12.4. Correlación intraclásica . . . . .	245
12.5. Ejemplos con R . . . . .	246
12.6. Ejercicios . . . . .	247
<b>A. Matrices</b>	<b>251</b>
A.1. Inversa generalizada . . . . .	251
A.2. Derivación matricial . . . . .	252
A.3. Matrices idempotentes . . . . .	252
A.4. Matrices mal condicionadas . . . . .	253
<b>B. Proyecciones ortogonales</b>	<b>255</b>
B.1. Descomposición ortogonal de vectores . . . . .	255
B.2. Proyecciones en subespacios . . . . .	257

<b>C. Estadística multivariante</b>	<b>259</b>
C.1. Esperanza, varianza y covarianza . . . . .	259
C.2. Normal multivariante . . . . .	260
<b>Bibliografía</b>	<b>261</b>
<b>Índice alfabético</b>	<b>265</b>



# Las condiciones

## I.1. Introducción

Los métodos de la Matemática que estudian los fenómenos deterministas relacionan, por lo general, una variable dependiente con diversas variables independientes. El problema se reduce entonces a resolver un sistema lineal, una ecuación diferencial, un sistema no lineal, etc.. Sin embargo, la aplicación de los métodos cuantitativos a las Ciencias Experimentales ha revelado la poca fiabilidad de las relaciones deterministas. En tales Ciencias, el azar, la aleatoriedad, la variabilidad individual, las variables no controladas, etc. justifican el planteo, en términos muy generales, de la ecuación fundamental

$$\text{"observación"} = \text{"modelo"} + \text{"error aleatorio"}$$

El experimentador puede, fijando las condiciones de su experimento, especificar la estructura del modelo, pero siempre debe tener en cuenta el error aleatorio o desviación entre lo que observa y lo que espera observar según el modelo.

Los modelos de regresión utilizan la ecuación anterior fijando el modelo como una función lineal de unos parámetros. El objetivo consiste, casi siempre, en la predicción de valores mediante el modelo ajustado.

El *Análisis de la Varianza* es un método estadístico introducido por R.A. Fisher de gran utilidad en las Ciencias Experimentales, que permite controlar diferentes variables cualitativas y cuantitativas (llamadas factores), a través de un modelo lineal, suponiendo normalidad para el error aleatorio. Fisher(1938) definió este método como "la separación de la varianza atribuible a un grupo de la varianza atribuible a otros grupos". Como veremos, los tests en Análisis de la Varianza se construyen mediante estimaciones independientes de la varianza del error.

Ambos conjuntos de modelos se pueden abordar con una teoría común: los modelos lineales.

Iniciaremos este capítulo con un ejemplo de modelización de un problema y su aplicación práctica. A continuación explicaremos en qué consiste esencialmente el método de los mínimos cuadrados y estableceremos las condiciones para que este método sea válido para su utilización en Estadística.

## I.2. Un ejemplo

En el libro de Sen and Srivastava en [67, pág. 2] se explica este ejemplo que nosotros hemos adaptado a las medidas europeas.

Sabemos que cuantos más coches circulan por una carretera, menor es la velocidad del tráfico. El estudio de este problema tiene como objetivo la mejora del transporte y la reducción del tiempo de viaje.

La tabla adjunta proporciona los datos de la densidad (en vehículos por km) y su correspondiente velocidad (en km por hora).

Dato	Densidad	Velocidad	Dato	Densidad	Velocidad
1	12,7	62,4	13	18,3	51,2
2	17,0	50,7	14	19,1	50,8
3	66,0	17,1	15	16,5	54,7
4	50,0	25,9	16	22,2	46,5
5	87,8	12,4	17	18,6	46,3
6	81,4	13,4	18	66,0	16,9
7	75,6	13,7	19	60,3	19,8
8	66,2	17,9	20	56,0	21,2
9	81,1	13,8	21	66,3	18,3
10	62,8	17,9	22	61,7	18,0
11	77,0	15,8	23	66,6	16,6
12	89,6	12,6	24	67,8	18,3

Tabla I.1: Datos del problema de tráfico

Como la congestión afecta a la velocidad, estamos interesados en determinar el efecto de la densidad en la velocidad. Por razones que explicaremos más adelante (ver ejercicio 9.2), tomaremos como variable dependiente la raíz cuadrada de la velocidad.

El gráfico 1.1 presenta la nube de puntos o diagrama de dispersión (*scatter plot*) con la variable independiente (densidad) en el eje horizontal y la variable dependiente (raíz cuadrada de la velocidad) en el eje vertical.

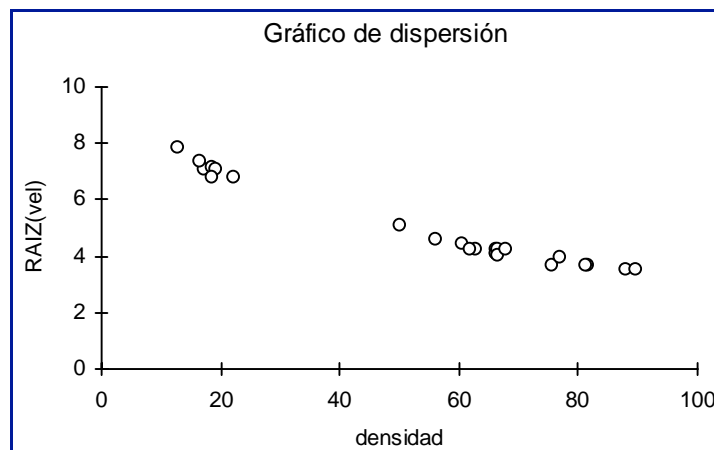


Figura I.1: Nube de puntos del problema de tráfico

Como primera aproximación podríamos tomar, como modelo de ajuste, la recta que une dos puntos representativos, por ejemplo, los puntos  $(12,7, \sqrt{62,4})$  y  $(87,8, \sqrt{12,4})$ . Dicha recta es  $y = 8,6397 - 0,0583x$ .

Inmediatamente nos proponemos hallar la mejor de las rectas, según algún criterio. Como veremos, el método de los mínimos cuadrados proporciona una recta, llamada recta de regresión, que goza de muy buenas propiedades. Este método consiste en hallar  $a$  y  $b$  tales que se minimice la suma de los errores al cuadrado.

$$\sum_{i=1}^n (y_i - (a + bx_i))^2$$

En este caso la recta de regresión es  $y = 8,0898 - 0,0566x$ .

Para estudiar la bondad del ajuste se utilizan los residuos

$$e_i = y_i - \hat{y}_i$$

donde  $\hat{y}_i = 8,0898 - 0,0566x_i$ . Los gráficos de la figura 1.2 nos muestran estos residuos.



Para mejorar el modelo podemos añadir el término cuadrático y considerar el modelo parabólico

$$y_i = a + bx_i + cx_i^2$$

También aquí, el método de los mínimos cuadrados proporciona un ajuste que es óptimo en varios aspectos. Se trata de hallar los valores de  $a$ ,  $b$  y  $c$  que minimizan la suma de los errores al cuadrado

$$\sum_{i=1}^n (y_i - (a + bx_i + cx_i^2))^2$$

El cálculo de estos valores con los datos del tráfico se deja como ejercicio (ver ejercicio 1.3).

La figura 1.3 muestra los gráficos de los residuos para el modelo parabólico.

Finalmente, podemos utilizar el modelo concreto que hemos obtenido para sustituir la velocidad en la ecuación

$$\text{flujo} = \text{velocidad} \times \text{densidad}$$

de modo que el flujo queda en función de la densidad. Por último, el máximo valor de esta función es la capacidad de la carretera.

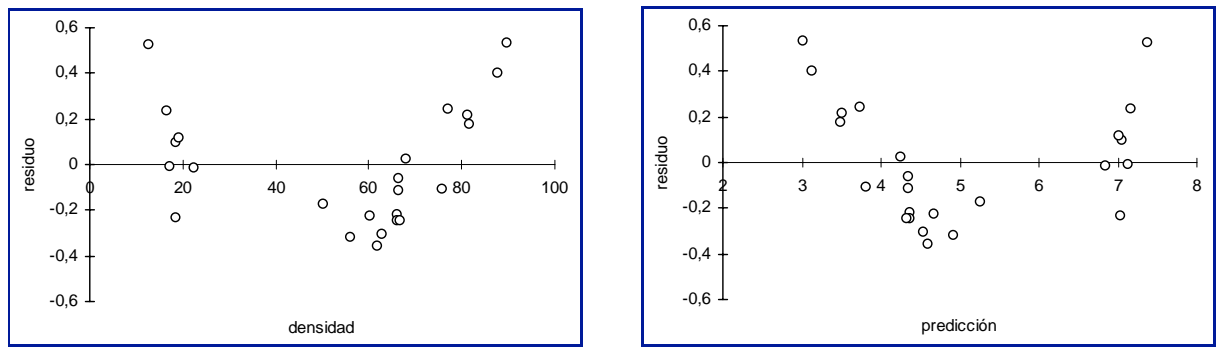


Figura 1.2: Gráficos de los residuos del modelo recta de regresión.

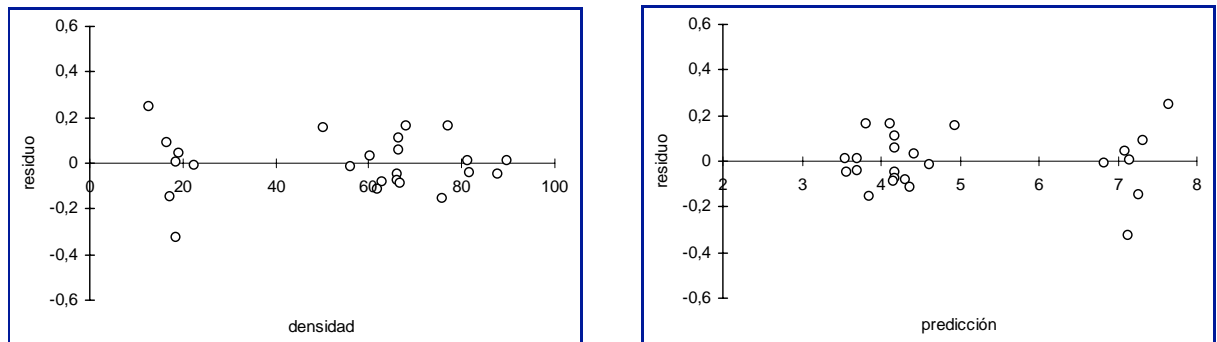


Figura 1.3: Gráficos de los residuos del modelo parabólico.

### 1.3. El modelo

Cuando en el ejemplo anterior ajustamos los datos a una recta, implícitamente estamos asumiendo la hipótesis de que los datos siguen un patrón lineal subyacente del tipo

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

Pero el ajuste no es perfecto y contiene errores. La ecuación que define el modelo es

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, \dots, n$$

donde  $\epsilon_i$  son los errores aleatorios. Éste es el modelo de *regresión simple* o con una sola variable independiente.

En el mismo ejemplo anterior, ajustamos mejor con el modelo

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i \quad i = 1, \dots, n$$

que continúa siendo un *modelo lineal*.

Un modelo es lineal si lo es para los parámetros. Por ejemplo, el modelo  $\ln y_i = \beta_0 + \beta_1 \ln(x_i) + \epsilon_i$  es lineal, mientras que  $y_i = \beta_0 \exp(-\beta_1 x_i) + \epsilon_i$  no.

En general, suponemos que una cierta variable aleatoria  $Y$  es igual a un valor fijo  $\eta$  más una desviación aleatoria  $\epsilon$

$$Y = \eta + \epsilon$$

$\eta$  representa la verdadera medida de la variable, es decir, la parte *determinista* de un experimento, que depende de ciertos factores cualitativos y variables cuantitativas que son controlables por el experimentador.

El término  $\epsilon$  representa el *error*. Es la parte del modelo no controlable por el experimentador debido a múltiples causas aleatorias, inevitables en los datos que proceden de la Biología, Psicología, Economía, Medicina, . . . El error  $\epsilon$  convierte la relación matemática  $Y = \eta$  en la relación estadística  $Y = \eta + \epsilon$ , obligando a tratar el modelo desde la perspectiva del análisis estadístico.

En particular, los modelos de la forma

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i \quad i = 1, \dots, n$$

con  $k > 1$  variables independientes, predictoras o regresoras, se llaman modelos de *regresión múltiple*. La variable cuyos datos observados son  $y_i$  es la llamada variable dependiente o respuesta.

Los parámetros  $\beta_j$  son desconocidos y nuestro objetivo principal es su estimación. En cuanto a los errores  $\epsilon_i$ , su cálculo explícito nos permitirá, como veremos extensamente, la evaluación del modelo.

*Observación:*

En el modelo de regresión simple puede suceder que los datos  $x_i$   $i = 1, \dots, n$  correspondan a los valores observados de una v.a.  $X$  o de una variable controlada no aleatoria. En cualquier caso, vamos a considerar los valores  $x_i$  como constantes y no como observaciones de una variable aleatoria.

En la regresión simple

$$Y = \phi(x) + \epsilon$$

donde  $Y$  es aleatoria y  $\epsilon$  es aleatoria con  $E(\epsilon) = 0$ . De manera que, para cada valor  $X = x$ ,  $Y$  es una v.a. con esperanza  $\phi(x)$ . Si asumimos

$$\phi(x) = E[Y|X = x] = \beta_0 + \beta_1 x$$

podemos proceder considerando las inferencias como condicionadas a los valores observados de  $X$ .

En cualquier caso, también en regresión múltiple, vamos a considerar los valores de las variables regresoras  $X_1, \dots, X_k$  como simplemente números.

## 1.4. El método de los mínimos cuadrados

La paternidad de este método se reparte entre Legendre que lo publicó en 1805 y Gauss que lo utilizó en 1795 y lo publicó en 1809.

Obviamente, cuanto menores son los residuos, mejor es el ajuste. De todos los posibles valores de los  $\beta_j$ , el método de los mínimos cuadrados selecciona aquellos que minimizan

$$S = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}))^2$$

En el caso de la regresión lineal simple

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

de modo que derivando e igualando a cero, se obtienen los estimadores MC (mínimo-cuadráticos) ó LS (*least squares*)

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}\end{aligned}$$

También se puede considerar el *modelo centrado*, que consiste en centrar los datos de la variable regresora

$$y_i = \gamma_0 + \beta_1(x_i - \bar{x}) + \epsilon_i \quad i = 1, \dots, n$$

La estimación MC de  $\gamma_0, \beta_1$  es equivalente a la estimación de  $\beta_0, \beta_1$ , ya que  $\gamma_0 = \beta_0 + \beta_1 \bar{x}$ . De modo que  $\hat{\gamma}_0 = \bar{y}$  y la estimación de  $\beta_1$  es la misma que en el modelo anterior.

Con las estimaciones de los parámetros, podemos proceder al cálculo de predicciones  $\hat{y}_i$  y residuos  $e_i$

$$\begin{aligned}\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x}) \\ e_i &= y_i - \hat{y}_i = y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x})\end{aligned}$$

Como consecuencia resulta que

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

lo que no ocurre en un modelo sin  $\beta_0$ .

Finalmente, si queremos una medida del ajuste de la regresión podemos pensar en la suma de cuadrados  $\sum_{i=1}^n e_i^2$ , pero es una medida que depende de las unidades de  $y_i$  al cuadrado. Si  $\beta_0 \neq 0$ , la medida que se utiliza es el coeficiente de determinación

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Sabemos que  $0 \leq R^2 \leq 1$  y cuando  $R^2 \approx 1$  el ajuste es bueno.

En el caso  $\beta_0 = 0$ , el coeficiente de determinación es

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n y_i^2}$$

de modo que los modelos que carecen de término independiente no se pueden comparar con los que sí lo tienen.

## I.5. Las condiciones de Gauss-Markov

Hasta aquí, el método de los mínimos cuadrados es analítico ¿dónde está la estadística?

A lo largo de los siguientes capítulos vamos a ver que un modelo estadístico y la imposición de algunas condiciones, hacen que podamos utilizar el modelo con toda la potencia de los métodos estadísticos y calibrar la bondad del ajuste desde esa óptica.

Una primera pregunta es ¿qué tan bueno es el método de los mínimos cuadrados para estimar los parámetros? La respuesta es que este método proporciona un buen ajuste y buenas predicciones si se verifican las condiciones de Gauss-Markov.

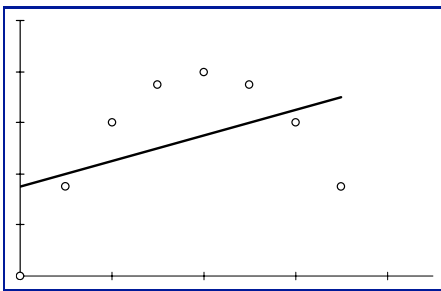
En el modelo lineal que hemos definido anteriormente, se supone que los errores  $\epsilon_i$  son desviaciones que se comportan como variables aleatorias. Vamos a exigir que estos errores aleatorios verifiquen las siguientes condiciones:

1.  $E(\epsilon_i) = 0 \quad i = 1, \dots, n$
2.  $\text{var}(\epsilon_i) = \sigma^2 \quad i = 1, \dots, n$
3.  $E(\epsilon_i \cdot \epsilon_j) = 0 \quad \forall i \neq j$

Veamos con detalle estas condiciones:

### Primera condición

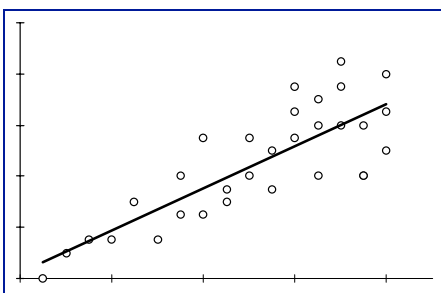
$$E(\epsilon_i) = 0 \quad i = 1, \dots, n$$



Se trata de una condición natural sobre un error. De este modo nos aseguramos que  $E(y_i) = \beta_0 + \beta_1 x_i$ , el modelo lineal es correcto y la situación que representa el gráfico no se puede dar.

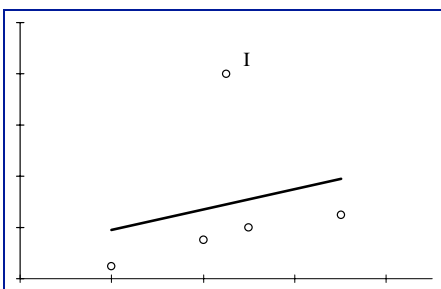
### Segunda condición

$$\text{var}(\epsilon_i) = E(\epsilon_i^2) = \sigma^2 \text{ constante} \quad i = 1, \dots, n$$

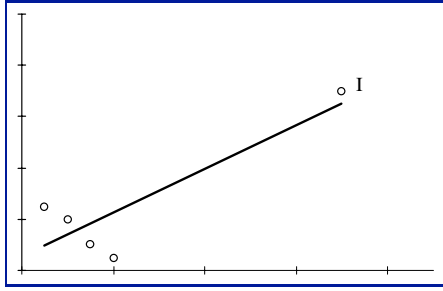


Es la propiedad de *homocedasticidad*. En el gráfico se representa una situación anómala llamada de *heterocedasticidad*, en la que la  $\text{var}(\epsilon_i)$  crece con  $x_i$ . El parámetro desconocido  $\sigma^2$  es la llamada varianza del modelo.

Otras situaciones extrañas, que también se pretende prevenir, son:



El punto I del gráfico representa un punto influyente y atípico (*outlier*). En general es un punto a estudiar, un error o incluso una violación de la primera condición.



El punto I del gráfico es claramente influyente, aunque no es atípico (*outlier*), ya que proporciona un residuo pequeño.

### Tercera condición

$$E(\epsilon_i \epsilon_j) = 0 \quad \forall i \neq j$$

Las observaciones deben ser incorrelacionadas. Con dos puntos tenemos una recta de regresión. Con 20 copias de esos dos puntos, tenemos 40 puntos y la misma recta, poco fiable.

Tales condiciones pueden expresarse en forma matricial como

$$E(\epsilon) = \mathbf{0} \quad \text{var}(\epsilon) = \sigma^2 \mathbf{I}_n$$

donde  $E(\epsilon)$  es el vector de esperanzas matemáticas y  $\text{var}(\epsilon)$  es la matriz de covarianzas de  $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ .

Como demostraremos en los siguientes capítulos, la adopción de estas condiciones evitará teóricamente las situaciones anómalas que aquí hemos esquematizado.

## 1.6. Otros tipos de modelos lineales

Por suerte, con el mismo tratamiento podremos resolver otros modelos lineales, que aunque tienen diferentes objetivos, gozan de las mismas bases teóricas.

Por ejemplo, el Análisis de la Varianza con un factor (one-way Analysis of Variance), representado por el modelo lineal

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad \text{con } \epsilon_{ij} \sim N(0, \sigma^2) \text{ indep.,}$$

se resuelve de forma similar al modelo de regresión.

El Análisis de la Covarianza, que utiliza como variables independientes tanto variables cuantitativas como factores, y el Análisis Multivariante de la Varianza, con varias variables dependientes, son dos de los análisis que generalizan el estudio y aplicaciones de los modelos lineales que vamos a investigar.

## 1.7. Algunas preguntas

Un típico problema de estadística consiste en estudiar la relación que existe, si existe, entre dos variables aleatorias  $X$  e  $Y$ . Por ejemplo, altura y peso, edad del hombre y la mujer en una pareja, longitud y anchura de unas hojas, temperatura y presión de un determinado volumen de gas.

Si tenemos  $n$  pares de observaciones  $(x_i, y_i)$   $i = 1, 2, \dots, n$ , podemos dibujar estos puntos en un gráfico o *scatter diagram* y tratar de ajustar una curva a los puntos de forma que los puntos se hallen lo más *cerca* posible de la curva. No podemos esperar un ajuste perfecto porque ambas variables están expuestas a fluctuaciones al azar debido a factores incontrolables. Incluso aunque en algunos casos pudiera existir una relación exacta entre variables físicas como temperatura y presión, también aparecerían fluctuaciones debidas a errores de medida.

Algunas cuestiones que podemos plantearnos en nuestras investigaciones son:

- Si existe un modelo físico teórico y lineal, podemos utilizar la regresión para estimar los parámetros.

- Si el modelo teórico no es lineal, se puede, en muchos casos, transformar en lineal. Por ejemplo:

$$PV^\gamma = c \longrightarrow \log P = \log c - \gamma \log V$$

- Si no es una recta, se puede estudiar un modelo de regresión polinómico. ¿De qué grado?
- En el modelo múltiple intervienen varias variables “predictoras” ¿son todas necesarias? ¿son linealmente independientes las llamadas “variables independientes”?
- ¿Se verifican realmente las condiciones de Gauss-Markov?
- ¿Qué ocurre si las variables predictoras son discretas?
- ¿Qué ocurre si la variable dependiente es discreta o una proporción?
- ¿Y si faltan algunos datos?
- ¿Qué hacemos con los puntos atípicos y los puntos influyentes?

Algunas de estas preguntas las iremos trabajando y resolviendo en los siguientes capítulos, otras pueden quedar para una posterior profundización.

## I.8. Ejemplos con R

En esta sección vamos a ver como se calculan las regresiones que se han sugerido a partir del ejemplo inicial con los datos de la tabla 1.1.

En primer lugar procedemos a introducir los datos en los vectores correspondientes.

```
> dens<-c(12.7,17.0,66.0,50.0,87.8,81.4,75.6,66.2,81.1,62.8,77.0,89.6,
+ 18.3,19.1,16.5,22.2,18.6,66.0,60.3,56.0,66.3,61.7,66.6,67.8)
> vel<-c(62.4,50.7,17.1,25.9,12.4,13.4,13.7,17.9,13.8,17.9,15.8,12.6,
+ 51.2,50.8,54.7,46.5,46.3,16.9,19.8,21.2,18.3,18.0,16.6,18.3)
> rvel<-sqrt(vel)
```

Las siguientes instrucciones generan el gráfico de puntos para estos datos.

```
> par(pty="m")
> plot(dens,rvel,type="p",xlab="densidad",ylab="RAIZ(vel) ")
```

El cálculo de la regresión simple se realiza con la función `lsfit(x,y)` que asignamos al objeto `recta.ls`

```
> recta.ls<-lsfit(dens,rvel)
```

Aunque esta última instrucción no muestra ninguna información en pantalla, ahora ya podemos utilizar su resultado. Por ejemplo, podemos añadir la recta de regresión al gráfico anterior.

```
> abline(recta.ls)
```

Los coeficientes de la recta son:

```
> recta.ls$coef
      Intercept           X
8.08981299 -0.05662558
```

También se puede obtener una información más completa con la instrucción `ls.print`, aunque su resultado no se explicará hasta el capítulo correspondiente.

```
> ls.print(recta.ls, digits=4, print.it=T)
Residual Standard Error=0.2689
R-Square=0.9685
F-statistic (df=1, 22)=676.3944
p-value=0
```

	Estimate	Std.Err	t-value	Pr(> t )
Intercept	8.0898	0.1306	61.9295	0
X	-0.0566	0.0022	-26.0076	0

La estimación de la desviación estándar de los errores y otros elementos de diagnóstico del modelo se obtienen con la función `ls.diag` como

```
> ls.diag(recta.ls)$std.dev
[1] 0.2689388
```

Con el vector de residuos y las predicciones se pueden dibujar unos gráficos similares a los de la figura 1.2. La instrucción `par(mfrow=c(1,2))` permite dos gráficos en la misma figura.

```
> e<-recta.ls$residuals
> par(mfrow=c(1,2))
> par(pty="s")
> plot(dens,e,type="p",xlab="densidad",ylab="residuos",ylim=c(-0.6,0.6))
> abline(h=0)
> pred<-rvel-e
> plot(pred,e,type="p",xlab="predicción",ylab="residuos",ylim=c(-0.6,0.6))
> abline(h=0)
```

Finalmente, podemos repetir los cálculos para el modelo parabólico. Simplemente debemos introducir los valores de la variable densidad y sus cuadrados en una matriz de datos. El resto es idéntico al modelo de regresión simple.

```
> matriz.frame<-data.frame(dens,dens^2)
> parabola.ls<-lsfit(matriz.frame,rvel)
> parabola.ls$coef
      Intercept      dens      dens.2
8.8814208199 -0.1035152795 0.0004892585
> round(parabola.ls$coef,5)
Intercept      dens      dens.2
8.88142 -0.10352 0.00049
> e<-parabola.ls$residuals
> par(mfrow=c(1,2))
> par(pty="s")
> plot(dens,e,type="p",xlab="densidad",ylab="residuos",ylim=c(-0.6,0.6))
> abline(h=0)
> pred<-rvel-e
> plot(pred,e,type="p",xlab="predicción",ylab="residuos",ylim=c(-0.6,0.6))
> abline(h=0)
```

Los gráficos serán muy similares a los de la figura 1.3.

En los siguientes capítulos veremos otras instrucciones de R, en especial la función `lm`, que permiten ajustar un modelo de regresión a unos datos.

## I.9. Ejercicios

### Ejercicio 1.1

Hallar las estimaciones de los parámetros en un modelo de regresión lineal simple, minimizando la suma de los cuadrados de los errores:

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

Hallar una expresión para las predicciones  $\hat{y}_i$  y los residuos  $e_i = y_i - \hat{y}_i$ .

### Ejercicio 1.2

Hallar las estimaciones de los parámetros en un modelo de regresión parabólico, minimizando la suma de los cuadrados de los errores:

$$S = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2$$

Hallar una expresión para las predicciones  $\hat{y}_i$  y los residuos  $e_i = y_i - \hat{y}_i$ .

### Ejercicio 1.3

Consideremos el problema de tráfico planteado en el apartado 1.2 de este capítulo, con la variable independiente *densidad* y la variable dependiente *raíz cuadrada de la velocidad*. Con los datos proporcionados en la tabla 1.1 realizar el siguiente proceso:

- Dibujar la nube de puntos y la recta que pasa por los puntos  $(12.7, \sqrt{62.4})$  y  $(87.8, \sqrt{12.4})$ . Dibujar el gráfico de los residuos con la densidad y el gráfico con las predicciones. Calcular la suma de cuadrados de los residuos.
- Hallar la recta de regresión simple. Dibujar el gráfico de los residuos con la densidad y el gráfico con las predicciones. Calcular la suma de cuadrados de los residuos.
- Mejorar el modelo anterior considerando una regresión parabólica. Dibujar el gráfico de los residuos con la densidad y el gráfico con las predicciones. Calcular la suma de cuadrados de los residuos.
- Calcular la capacidad de la carretera o punto de máximo flujo. Recordar que  $\text{flujo} = \text{vel} \times \text{densidad}$ .

### Ejercicio 1.4

La siguiente tabla contiene los mejores tiempos conseguidos en algunas pruebas de velocidad en atletismo en los Juegos Olímpicos de Atlanta:

distancia	hombres tiempo	mujeres tiempo
100	9,84	10,94
200	19,32	22,12
400	43,19	48,25
800	102,58	117,73
1500	215,78	240,83
5000	787,96	899,88
10000	1627,34	1861,63
42192	7956,00	8765,00



Si tomamos como variable regresora o independiente la distancia (metros) y como variable respuesta o dependiente el tiempo (segundos):

- (a) Calcular la recta de regresión simple con los datos de los hombres y dibujarla. Dibujar el gráfico de los residuos con la distancia y el gráfico con las predicciones. Calcular la suma de cuadrados de los residuos y el  $R^2$ .
- (b) Repetir el apartado anterior utilizando los logaritmos de las variables tiempo y distancia.
- (c) Repetir los dos apartados anteriores utilizando los datos de las mujeres.



# Estimación

## 2.1. Introducción

En primer lugar concretaremos la definición general de un modelo lineal y hallaremos la estimación por mínimos cuadrados de los parámetros del modelo.

Veremos que la estimación será única si la matriz de diseño es de rango máximo. En caso contrario, resulta importante definir el concepto de función paramétrica estimable y probar, para estas funciones, la unicidad del estimador mínimo-cuadrático, como estudiaremos en el siguiente capítulo.

Estudiaremos las propiedades de estos estimadores, entre las que destacaremos el *Teorema de Gauss-Markov* que demuestra que los estimadores mínimo-cuadráticos son los mejores, en el sentido de que son insesgados y de mínima varianza.

Además, con la introducción de la hipótesis de normalidad de los errores, podremos estudiar las distribuciones de los estimadores y de otros estadísticos, así como la relación con los estimadores de máxima verosimilitud.

Más adelante, trabajaremos la generalización del método de los mínimos cuadrados cuando la matriz de varianzas-covarianzas de los errores no es  $\sigma^2 \mathbf{I}$ . Por otra parte, también profundizaremos el caso de matrices de diseño de rango no máximo.

## 2.2. El modelo lineal

Sea  $Y$  una variable aleatoria que fluctúa alrededor de un valor desconocido  $\eta$ , esto es

$$Y = \eta + \epsilon$$

donde  $\epsilon$  es el error, de forma que  $\eta$  puede representar el valor verdadero e  $Y$  el valor observado.

Supongamos que  $\eta$  toma valores distintos de acuerdo con diferentes situaciones experimentales según el modelo lineal

$$\eta = \beta_1 x_1 + \cdots + \beta_m x_m$$

donde  $\beta_i$  son parámetros desconocidos y  $x_i$  son valores conocidos, cada uno de los cuales ilustra situaciones experimentales diferentes.

En general se tienen  $n$  observaciones de la variable  $Y$ . Diremos que  $y_1, y_2, \dots, y_n$  observaciones independientes de  $Y$  siguen un *modelo lineal* si

$$y_i = x_{i1}\beta_1 + \cdots + x_{im}\beta_m + \epsilon_i \quad i = 1, \dots, n$$

Estas observaciones de  $Y$  se pueden considerar variables aleatorias independientes y distribuidas como  $Y$  (son copias) o también realizaciones concretas (valores numéricos) para los cálculos.

La expresión del modelo lineal en forma matricial es

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

o en forma resumida

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.1)$$

Los elementos que constituyen el modelo lineal son:

1. El vector de observaciones  $\mathbf{Y} = (y_1, y_2, \dots, y_n)'$ .
2. El vector de parámetros  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m)'$ .
3. La matriz del modelo

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

cuyos elementos son conocidos.

En problemas de regresión,  $\mathbf{X}$  es la matriz de regresión. En los llamados diseños factoriales del Análisis de la Varianza,  $\mathbf{X}$  recibe el nombre de matriz de diseño.

4. El vector de errores o desviaciones aleatorias  $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'$ , donde  $\epsilon_i$  es la desviación aleatoria de  $y_i$ .

### Ejemplo 2.2.1

El modelo lineal más simple consiste en relacionar una variable aleatoria  $Y$  con una variable controlable  $x$  (no aleatoria), de modo que las observaciones de  $Y$  verifiquen

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, \dots, n$$

Se dice que  $Y$  es la variable de predicción o dependiente y  $x$  es la variable predictora, por ejemplo  $Y$  es la respuesta de un fármaco a una dosis  $x$ . Hallar  $\beta_0$  y  $\beta_1$  es el clásico problema de regresión lineal simple.

### Ejemplo 2.2.2

El modelo anterior se puede generalizar a situaciones en las cuales la relación sea polinómica.

Consideremos el modelo

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_p x_i^p + \epsilon_i \quad i = 1, \dots, n$$

Observemos que es lineal en los parámetros  $\beta_i$ . La matriz de diseño es

$$\begin{pmatrix} 1 & x_1 & \dots & x_1^p \\ 1 & x_2 & \dots & x_2^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^p \end{pmatrix}$$

### Ejemplo 2.2.3

En general, cualquier variable  $Y$  puede relacionarse con dos o más variables control. Así, son modelos lineales:

- a)  $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$
- b)  $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \beta_4 x_{i1}^2 + \beta_5 x_{i2}^2 + \epsilon_i$
- c)  $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 \cos(x_{i2}) + \beta_3 \sin(x_{i2}) + \epsilon_i$

Sin embargo, no es modelo lineal

$$y_i = \beta_0 + \beta_1 \log(\beta_2 x_{i1}) + \beta_3 x_{i2}^{\beta_4} + \epsilon_i$$

### Ejemplo 2.2.4

Supongamos que la producción  $Y$  de una planta depende de un factor  $F$  (fertilizante) y un factor  $B$  (bloque o conjunto de parcelas homogéneas). El llamado modelo del diseño del factor en bloques aleatorizados es

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

donde

$\mu$  es una constante (media general)

$\alpha_i$  el efecto del fertilizante

$\beta_j$  el efecto del bloque

Si tenemos 2 fertilizantes y 3 bloques, tendremos en total  $k = 2 \times 3 = 6$  situaciones experimentales y la siguiente matriz de diseño:

$\mu$	$\alpha_1$	$\alpha_2$	$\beta_1$	$\beta_2$	$\beta_3$
1	1	0	1	0	0
1	0	1	1	0	0
1	1	0	0	1	0
1	0	1	0	1	0
1	1	0	0	0	1
1	0	1	0	0	1

La utilización del fertilizante 1 en el bloque 3 queda descrita a través de la fila 5 de  $\mathbf{X}$ .

### Ejemplo 2.2.5

Para predecir la capacidad craneal  $C$ , en Antropología se utiliza la fórmula

$$C = \alpha L^{\beta_1} A^{\beta_2} H^{\beta_3}$$

donde  $L$  = longitud del cráneo,  $A$  = anchura parietal máxima y  $H$  = altura basio bregma.

La fórmula anterior se convierte en un modelo lineal tomando logaritmos

$$\log C = \log \alpha + \beta_1 \log L + \beta_2 \log A + \beta_3 \log H$$

El parámetro  $\alpha$  expresa el tamaño, mientras que los parámetros  $\beta$  expresan la forma del cráneo.

## 2.3. Suposiciones básicas del modelo lineal

En el modelo lineal definido en el apartado anterior, se supone que los errores  $\epsilon_i$  son desviaciones que se comportan como variables aleatorias que verifican las condiciones de Gauss-Markov:

1.  $E(\epsilon_i) = 0 \quad i = 1, \dots, n$
2.  $\text{var}(\epsilon_i) = \sigma^2 \quad i = 1, \dots, n$
3.  $E(\epsilon_i \cdot \epsilon_j) = 0 \quad \forall i \neq j$

Como sabemos, la condición (2) es la llamada condición de *homocedasticidad* del modelo y el parámetro desconocido  $\sigma^2$  es la llamada varianza del modelo. La condición (3) significa que las  $n$  desviaciones son mutuamente incorrelacionadas.

Estas condiciones pueden expresarse en forma matricial como

$$E(\epsilon) = \mathbf{0} \quad \text{var}(\epsilon) = \sigma^2 \mathbf{I}_n$$

donde  $E(\epsilon)$  es el vector de esperanzas matemáticas y  $\text{var}(\epsilon)$  es la matriz de covarianzas de  $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ .

Si además suponemos que cada  $\epsilon_i$  es  $N(0, \sigma)$  y que  $\epsilon_1, \dots, \epsilon_n$  son estocásticamente independientes, entonces diremos que el modelo definido es un *modelo lineal normal*. Así tendremos que

$$\mathbf{Y} \sim N_n(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$$

es decir,  $\mathbf{Y}$  sigue la distribución normal multivariante de vector de medias  $\mathbf{X}\beta$  y matriz de covarianzas  $\sigma^2 \mathbf{I}_n$ .

Se llama rango del diseño al rango de la matriz  $\mathbf{X}$

$$r = \text{rango } \mathbf{X}$$

y es un elemento muy importante en la discusión de los modelos. Evidentemente  $r \leq m$ . El valor de  $r$  es el número efectivo de parámetros del diseño, en el sentido de que si  $r < m$  es posible reparametrizar el modelo para que  $r$  sea igual al número de parámetros. En muchos casos el diseño verifica directamente que  $r = m$  y entonces se dice que es de *rango máximo*.

El modelo lineal que verifique las condiciones aquí expuestas, salvo la normalidad, diremos que está bajo las condiciones de Gauss-Markov ordinarias.

## 2.4. Estimación de los parámetros

La estimación de los parámetros  $\beta = (\beta_1, \dots, \beta_m)'$  se hace con el criterio de los mínimos cuadrados. Se trata de hallar el conjunto de valores de los parámetros  $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_m)'$  que minimicen la siguiente suma de cuadrados

$$\begin{aligned} \epsilon' \epsilon &= (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \\ &= \sum_{i=1}^n (y_i - x_{i1}\beta_1 - \dots - x_{im}\beta_m)^2 \end{aligned} \quad (2.2)$$

La estimación  $\hat{\beta}$  de  $\beta$  la llamaremos estimación MC, abreviación de mínimo-cuadrática, o LS del inglés *least squares*.

### Teorema 2.4.1

Toda estimación MC de  $\beta$  es solución de la ecuación

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{Y} \quad (2.3)$$

*Demostración:*

Si desarrollamos la suma de cuadrados  $\epsilon' \epsilon$  tenemos

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

y si derivamos matricialmente respecto a  $\beta$  resulta

$$\frac{\partial \mathbf{e}'\mathbf{e}}{\partial \beta} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta$$

De modo que, si igualamos a cero, obtenemos la ecuación enunciada en el teorema. ■

Las ecuaciones 2.3 reciben el nombre de *ecuaciones normales*.

Si el rango es máximo y  $r = m$ , entonces  $\mathbf{X}'\mathbf{X}$  tiene inversa y la única solución de las ecuaciones normales es

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Si  $r < m$  el sistema de ecuaciones 2.3 es indeterminado y su solución no es única. En estos casos, una posibilidad (ver Apéndice A) es considerar

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$$

donde  $\mathbf{A}^{-} = (\mathbf{X}'\mathbf{X})^{-}$  es una g-inversa de  $\mathbf{A} = \mathbf{X}'\mathbf{X}$ , es decir,  $\mathbf{A}^{-}$  verifica

$$\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$$

Entonces se puede demostrar que la solución general es

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{z}$$

siendo  $\mathbf{z}$  un vector paramétrico.

Ahora podemos definir la suma de cuadrados residual como

$$\text{SCR} = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\widehat{\beta})'(\mathbf{Y} - \mathbf{X}\widehat{\beta})$$

Como veremos, SCR entendido como un estadístico función de la muestra  $\mathbf{Y}$ , desempeña un papel fundamental en el Análisis de la Varianza.

El modelo lineal  $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ , bajo las hipótesis de Gauss-Markov, verifica

$$E(\mathbf{Y}) = \mathbf{X}\beta$$

### Teorema 2.4.2

Sea  $\Omega = \langle \mathbf{X} \rangle \subset \mathbb{R}^n$  el subespacio vectorial generado por las columnas de  $\mathbf{X}$  de dimensión  $\dim(\mathbf{X}) = r = \text{rango } \mathbf{X}$ .

Entonces se verifica:

- (i)  $E(\mathbf{Y}) \in \langle \mathbf{X} \rangle$
- (ii) Si  $\widehat{\beta}$  es una estimación MC, el vector de residuos  $\mathbf{e} = \mathbf{Y} - \mathbf{X}\widehat{\beta}$  es ortogonal a  $\langle \mathbf{X} \rangle$ .

*Demostración:*

En efecto,

- i) Si  $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(m)}$  son las columnas de  $\mathbf{X}$ , entonces

$$E(\mathbf{Y}) = \mathbf{x}_{(1)}\beta_1 + \dots + \mathbf{x}_{(m)}\beta_m \in \langle \mathbf{X} \rangle$$

- ii)  $\mathbf{X}'\mathbf{e} = \mathbf{X}'(\mathbf{Y} - \mathbf{X}\widehat{\beta}) = \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{0}$

■

**Teorema 2.4.3**

Para cualquier  $\widehat{\beta}$  solución MC de 2.3 se verifica que

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\beta} \quad \mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}} \quad \text{SCR} = (\mathbf{Y} - \mathbf{X}\widehat{\beta})'(\mathbf{Y} - \mathbf{X}\widehat{\beta})$$

son únicos.

Además

$$\text{SCR} = \mathbf{Y}'\mathbf{Y} - \widehat{\beta}'\mathbf{X}'\mathbf{Y} \quad (2.4)$$

*Demostración:*

Si desarrollamos la suma de cuadrados residual SCR resulta

$$\text{SCR} = \mathbf{Y}'\mathbf{Y} - \widehat{\beta}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\widehat{\beta} + \widehat{\beta}'\mathbf{X}'\mathbf{X}\widehat{\beta}$$

y como  $\mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{X}'\mathbf{Y}$ , obtenemos

$$\text{SCR} = \mathbf{Y}'\mathbf{Y} - 2\widehat{\beta}'\mathbf{X}'\mathbf{Y} + \widehat{\beta}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \widehat{\beta}'\mathbf{X}'\mathbf{Y}$$

Consideremos ahora los vectores  $\widehat{\mathbf{Y}}_1 = \mathbf{X}\widehat{\beta}_1$  y  $\widehat{\mathbf{Y}}_2 = \mathbf{X}\widehat{\beta}_2$ , donde  $\widehat{\beta}_1$  y  $\widehat{\beta}_2$  son dos soluciones MC. Entonces  $\widehat{\mathbf{Y}}_1$  y  $\widehat{\mathbf{Y}}_2$  pertenecen al subespacio  $\langle \mathbf{X} \rangle$  generado por las columnas de  $\mathbf{X}$  y su diferencia  $\widehat{\mathbf{Y}}_1 - \widehat{\mathbf{Y}}_2$  también. Por otra parte, observamos que

$$\mathbf{X}'(\widehat{\mathbf{Y}}_1 - \widehat{\mathbf{Y}}_2) = \mathbf{X}'\mathbf{X}\widehat{\beta}_1 - \mathbf{X}'\mathbf{X}\widehat{\beta}_2 = \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{Y} = \mathbf{0}$$

de modo que  $\widehat{\mathbf{Y}}_1 - \widehat{\mathbf{Y}}_2$  pertenece al ortogonal de  $\langle \mathbf{X} \rangle$ . Así pues, necesariamente  $\widehat{\mathbf{Y}}_1 - \widehat{\mathbf{Y}}_2 = \mathbf{0}$  y el vector de errores  $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}_1 = \mathbf{Y} - \widehat{\mathbf{Y}}_2$  es único.

En consecuencia, la suma de cuadrados de los errores SCR también es única. ■

**Interpretación geométrica**

El modelo teórico es

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon = \theta + \epsilon \quad \text{si } \theta = \mathbf{X}\beta$$

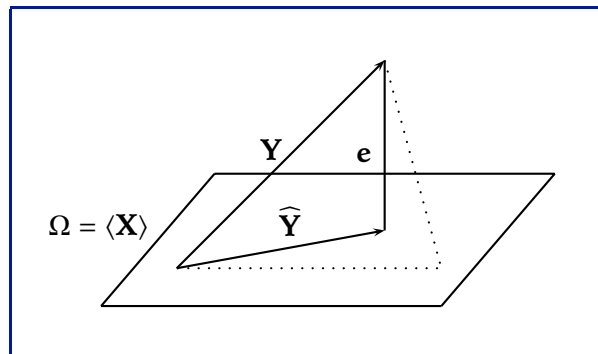
Entonces  $E(\mathbf{Y}) = \mathbf{X}\beta = \theta$  significa que el valor esperado de  $\mathbf{Y}$  pertenece al subespacio  $\Omega = \langle \mathbf{X} \rangle$  y para estimar los parámetros  $\beta$  debemos minimizar

$$\epsilon'\epsilon = \|\mathbf{Y} - \theta\|^2 \quad \text{con } \theta \in \Omega = \langle \mathbf{X} \rangle$$

Como el vector concreto de observaciones  $\mathbf{Y}$  se puede considerar un vector de  $\mathbb{R}^n$ , el problema anterior se puede resolver en términos geométricos. Así se sabe que cuando  $\theta \in \Omega$ ,  $\|\mathbf{Y} - \theta\|^2$  es mínimo para  $\theta = \widehat{\mathbf{Y}} = \mathbf{P}\mathbf{Y}$ , donde  $\mathbf{P}$  es la matriz de la proyección ortogonal en  $\Omega = \langle \mathbf{X} \rangle$  (ver Apéndice B). La estimación MC es equivalente a hallar la proyección ortogonal  $\widehat{\mathbf{Y}}$  de  $\mathbf{Y}$  sobre  $\langle \mathbf{X} \rangle$ , es decir, la norma euclídea de  $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}$  es mínima:

$$\text{SCR} = \mathbf{e}'\mathbf{e} = \|\mathbf{e}\|^2 = \|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2$$

Se comprende que cualquier otra proyección no ortogonal daría una solución menos adecuada.





Como  $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}$  es ortogonal a  $\Omega$ , se verifica que

$$\mathbf{X}'(\mathbf{Y} - \widehat{\mathbf{Y}}) = \mathbf{0} \quad \text{ó} \quad \mathbf{X}'\widehat{\mathbf{Y}} = \mathbf{X}'\mathbf{Y}$$

donde  $\widehat{\mathbf{Y}}$  está determinada por ser la única proyección ortogonal de  $\mathbf{Y}$  en  $\Omega$ . Cuando las columnas de  $\mathbf{X}$  son linealmente independientes, forman una base y existe un único vector  $\widehat{\boldsymbol{\beta}}$  tal que  $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$  de manera que

$$\mathbf{X}'\widehat{\mathbf{Y}} = \mathbf{X}'\mathbf{Y} \Rightarrow \mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$$

son las ecuaciones normales. En caso contrario, es decir, cuando las columnas de  $\mathbf{X}$  son dependientes no podemos concretar una solución única para los parámetros  $\boldsymbol{\beta}$ . Sin embargo todas las soluciones deben verificar la siguiente propiedad.

#### Teorema 2.4.4

$\widehat{\boldsymbol{\beta}}$  es una estimación MC de  $\boldsymbol{\beta}$  si y sólo si  $\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{Y}$ , donde  $\mathbf{P}$  es la proyección ortogonal en  $\Omega = \langle \mathbf{X} \rangle$

*Demostración:*

Una estimación  $\widehat{\boldsymbol{\beta}}$  de  $\boldsymbol{\beta}$  es MC si y sólo si

$$(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \min_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Sea  $\widetilde{\boldsymbol{\beta}}$  una estimación cualquiera de  $\boldsymbol{\beta}$ , entonces

$$\begin{aligned} (\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) &= (\mathbf{Y} - \mathbf{P}\mathbf{Y} + \mathbf{P}\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{P}\mathbf{Y} + \mathbf{P}\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) \\ &= (\mathbf{Y} - \mathbf{P}\mathbf{Y})'(\mathbf{Y} - \mathbf{P}\mathbf{Y}) + (\mathbf{Y} - \mathbf{P}\mathbf{Y})'(\mathbf{P}\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) \\ &\quad + (\mathbf{P}\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{P}\mathbf{Y}) + (\mathbf{P}\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})'(\mathbf{P}\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) \end{aligned}$$

Sin embargo

$$(\mathbf{Y} - \mathbf{P}\mathbf{Y})'(\mathbf{P}\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{P}\mathbf{Y} - \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{X}\widetilde{\boldsymbol{\beta}} = \mathbf{0}$$

ya que  $\mathbf{P}$  es idempotente y además  $\mathbf{P}\mathbf{X} = \mathbf{X}$ . De forma que

$$(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) = (\mathbf{Y} - \mathbf{P}\mathbf{Y})'(\mathbf{Y} - \mathbf{P}\mathbf{Y}) + (\mathbf{P}\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})'(\mathbf{P}\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})$$

donde ambos términos son positivos, el primero no depende de  $\widetilde{\boldsymbol{\beta}}$  y el segundo se minimiza si es cero, luego  $\mathbf{P}\mathbf{Y} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ . ■

En resumen y como ya hemos visto, la solución del problema se basa en la proyección ortogonal sobre el subespacio  $\Omega$  que garantiza la unicidad del vector de predicciones  $\widehat{\mathbf{Y}} = \mathbf{P}\mathbf{Y}$  y por ende del vector de residuos  $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}$  y de la suma de cuadrados de los residuos

$$\text{SCR} = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{P}\mathbf{Y})'(\mathbf{Y} - \mathbf{P}\mathbf{Y}) = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y}$$

ya que  $\mathbf{I} - \mathbf{P}$  es idempotente (ver Apéndice B).

La solución para los parámetros  $\boldsymbol{\beta}$  debe salir de las ecuaciones normales o de la ecuación  $\mathbf{X}\boldsymbol{\beta} = \mathbf{P}\mathbf{Y}$  y sólo es única cuando el rango de la matriz  $\mathbf{X}$  es máximo.

#### Ejemplo 2.4.1

Consideremos el modelo lineal con  $n = 3$ ,  $m = 1$  y  $r = 1$

$$\begin{aligned} y_1 &= \theta + \epsilon_1 \\ y_2 &= 2\theta + \epsilon_2 \\ y_3 &= -\theta + \epsilon_3 \end{aligned}$$

que en expresión matricial escribimos

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \theta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}$$

de modo que  $\mathbf{X}' = (1, 2, -1)$ .

Las ecuaciones normales son

$$\begin{pmatrix} 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \theta = \begin{pmatrix} 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

es decir

$$6\theta = y_1 + 2y_2 - y_3$$

y la estimación MC de  $\theta$  es  $\widehat{\theta} = (y_1 + 2y_2 - y_3)/6$ .

La suma de cuadrados residual es

$$\text{SCR} = \mathbf{Y}'\mathbf{Y} - \widehat{\theta}\mathbf{X}'\mathbf{Y} = y_1^2 + y_2^2 + y_3^2 - (y_1 + 2y_2 - y_3)^2/6$$

### Ejemplo 2.4.2

Supongamos que se desea pesar tres objetos cuyos pesos exactos son  $\beta_1$ ,  $\beta_2$  y  $\beta_3$ . Se dispone de una balanza de platillos con un error de pesada que podemos considerar con distribución  $N(0, \sigma)$ . Un artificio para mejorar la precisión y ahorrar pesadas consiste en repartir los objetos en uno o en los dos platillos y anotar las sumas o diferencias de pesos:

$$x_1\beta_1 + x_2\beta_2 + x_3\beta_3 = y$$

donde  $y$  es el peso observado y  $x_i = 0, 1, -1$ .

Consideremos las siguientes pesadas:

$$\begin{aligned} \beta_1 + \beta_2 + \beta_3 &= 5.53 \\ \beta_1 - \beta_2 + \beta_3 &= 1.72 \\ \beta_1 + \beta_2 - \beta_3 &= 0.64 \\ \beta_1 + \beta_2 + \beta_3 &= 5.48 \\ \beta_1 - \beta_2 + \beta_3 &= 1.70 \end{aligned}$$

A partir de estos datos, las ecuaciones normales son

$$\begin{cases} 5\beta_1 + \beta_2 + 3\beta_3 &= 15.07 \\ \beta_1 + 5\beta_2 - \beta_3 &= 8.23 \\ 3\beta_1 - \beta_2 + 5\beta_3 &= 13.79 \end{cases}$$

La estimación de los parámetros proporciona

$$\widehat{\beta}_1 = 1.175 \quad \widehat{\beta}_2 = 1.898 \quad \widehat{\beta}_3 = 2.433$$

y la suma de cuadrados residual es

$$\text{SCR} = (5.53 - (\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_3))^2 + \dots = 0.00145$$

## 2.5. Estimación de la varianza

La varianza de los errores del modelo lineal

$$\sigma^2 = \text{var}(\epsilon_i) = \text{var}(y_i) \quad i = 1, \dots, n$$

es otro parámetro que debe ser estimado a partir de las observaciones de  $y_1, \dots, y_n$ .

### Teorema 2.5.1

Sea  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  el modelo lineal con las hipótesis impuestas en la sección 2.3. Entonces el estadístico<sup>1</sup>

$$\widehat{\sigma}^2 = \text{ECM} = \text{SCR}/(n - r)$$

es un estimador insesgado de la varianza  $\sigma^2$ . En este estadístico SCR es la suma de cuadrados residual,  $n$  el número total de observaciones y  $r$  el rango del diseño.

#### Demostración 1:

Las columnas  $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(m)}$  de la matriz de diseño  $\mathbf{X}$  generan el subespacio de dimensión  $r$  que escribimos

$$\langle \mathbf{X} \rangle = \langle \mathbf{x}_{(1)}, \dots, \mathbf{x}_{(m)} \rangle$$

Sea ahora  $\mathbf{V}$  una matriz ortogonal, es decir, tal que  $\mathbf{V}\mathbf{V}' = \mathbf{V}'\mathbf{V} = \mathbf{I}_n$ , cuyas columnas  $\mathbf{v}_{(1)}, \dots, \mathbf{v}_{(r)}, \mathbf{v}_{(r+1)}, \dots, \mathbf{v}_{(n)}$  forman una base ortogonal de  $\mathbb{R}^n$ . Es posible construir  $\mathbf{V}$  de modo que las  $r$  primeras columnas generen el subespacio  $\langle \mathbf{X} \rangle$

$$\langle \mathbf{X} \rangle = \langle \mathbf{v}_{(1)}, \dots, \mathbf{v}_{(r)} \rangle$$

Por otra parte,  $\mathbf{Y} = (y_1, \dots, y_n)'$  es un vector aleatorio de  $\mathbb{R}^n$  que, mediante  $\mathbf{V}$ , transformamos en  $\mathbf{Z} = (z_1, \dots, z_n)' = \mathbf{V}'\mathbf{Y}$

$$z_i = v_{1i}y_1 + \dots + v_{ni}y_n \quad i = 1, \dots, n$$

Para las variables transformadas se verifica que

$$E(z_i) = \sum_{h=1}^n v_{hi}E(y_h) = \mathbf{v}_{(i)}'\mathbf{X}\boldsymbol{\beta} = \begin{cases} \eta_i & \text{si } i \leq r \\ 0 & \text{si } i > r \end{cases}$$

pues  $\mathbf{X}\boldsymbol{\beta} \in \langle \mathbf{X} \rangle$  que es ortogonal a  $\mathbf{v}_{(i)}$  para  $i > r$ .

Sea  $\widehat{\boldsymbol{\beta}}$  una estimación MC. Entonces

$$\mathbf{Y} = \mathbf{X}\widehat{\boldsymbol{\beta}} + (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{e}$$

donde obviamente  $\mathbf{X}\widehat{\boldsymbol{\beta}} \in \langle \mathbf{X} \rangle$  y como sabemos  $\mathbf{e} \in \langle \mathbf{X} \rangle^\perp$ , de manera que la transformación ortogonal  $\mathbf{V}'$  aplicada sobre  $\mathbf{e}$  proporciona

$$\mathbf{V}'\mathbf{e} = (0, \dots, 0, z_{r+1}, \dots, z_n)'$$

Luego, en función de las variables  $z_i$  tenemos

$$\text{SCR} = \mathbf{e}'\mathbf{e} = (\mathbf{V}'\mathbf{e})'\mathbf{V}'\mathbf{e} = \sum_{i=r+1}^n z_i^2$$

Además, por ser una transformación ortogonal, las variables  $z_1, \dots, z_n$  siguen siendo incorrelacionadas y de varianza  $\sigma^2$ . Así pues

$$E(z_i) = 0 \quad E(z_i^2) = \text{var}(z_i) = \text{var}(y_i) = \sigma^2$$

1. En muchos de los libros clásicos escritos en inglés este estadístico se llama MSE, siglas de *mean square error*.

y por lo tanto

$$E(\text{SCR}) = \sum_{i=r+1}^n E(z_i^2) = (n-r)\sigma^2$$

La expresión

$$\text{SCR} = z_{r+1}^2 + \cdots + z_n^2 \quad (2.5)$$

se llama *forma canónica* de la suma de cuadrados residual del modelo lineal bajo las hipótesis de Gauss-Markov. ■

### Demostración 2:

Se puede hacer una demostración mucho más directa a partir de la propiedad 2 explicada en el Apéndice C1 de Estadística Multivariante:

Para un vector aleatorio  $\mathbf{Y}$  con esperanza  $E(\mathbf{Y}) = \boldsymbol{\mu}$  y matriz de varianzas y covarianzas  $\text{var}(\mathbf{Y}) = \mathbf{V}$ , se tiene que

$$E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = \text{tr}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$$

donde  $\mathbf{A}$  es una matriz constante.

En nuestro caso  $E(\mathbf{Y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$  y  $\text{var}(\mathbf{Y}) = \mathbf{V} = \sigma^2\mathbf{I}$ , de forma que

$$\begin{aligned} E(\text{SCR}) &= E(\mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y}) = \text{tr}(\sigma^2(\mathbf{I} - \mathbf{P})) + \boldsymbol{\beta}'\mathbf{X}'(\mathbf{I} - \mathbf{P})\mathbf{X}\boldsymbol{\beta} \\ &= \sigma^2 \text{tr}(\mathbf{I} - \mathbf{P}) \\ &= \sigma^2 \text{rg}(\mathbf{I} - \mathbf{P}) = \sigma^2(n-r) \end{aligned}$$

gracias a las propiedades de la matriz  $\mathbf{I} - \mathbf{P}$ . ■

## 2.6. Distribuciones de los estimadores

Vamos ahora a establecer algunas propiedades de los estimadores MC para un modelo de rango máximo.

Si asumimos que los errores son insesgados  $E(\boldsymbol{\epsilon}) = \mathbf{0}$ , que es la primera condición de Gauss-Markov, entonces  $\widehat{\boldsymbol{\beta}}$  es un estimador insesgado de  $\boldsymbol{\beta}$

$$E(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

Si asumimos además que los errores  $\epsilon_i$  son incorrelacionados y con la misma varianza, es decir  $\text{var}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$ , resulta que

$$\text{var}(\mathbf{Y}) = \text{var}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \text{var}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$$

ya que  $\mathbf{X}\boldsymbol{\beta}$  no es aleatorio y en consecuencia

$$\begin{aligned} \text{var}(\widehat{\boldsymbol{\beta}}) &= \text{var}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{var}(\mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Veamos a continuación algunos resultados acerca de la distribución de  $\widehat{\boldsymbol{\beta}}$  y SCR bajo las hipótesis del modelo lineal normal en el caso de rango máximo.

### Teorema 2.6.1

Sea  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$  con rango  $\mathbf{X} = m$ . Entonces se verifican las siguientes propiedades:

- i) La estimación MC de  $\boldsymbol{\beta}$  coincide con la estimación de la máxima verosimilitud. Además es insesgada y de mínima varianza.

- ii)  $\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$
- iii)  $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma^2 \sim \chi_m^2$
- iv)  $\widehat{\boldsymbol{\beta}}$  es independiente de SCR
- v)  $\text{SCR}/\sigma^2 \sim \chi_{n-m}^2$

*Demostración:*

- i) La función de verosimilitud es

$$L(\mathbf{Y}; \boldsymbol{\beta}, \sigma^2) = (\sqrt{2\pi\sigma^2})^{-n} \exp \left[ -\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right]$$

de modo que el mínimo de  $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$  es el máximo de  $L$ .

Ya hemos visto que  $\widehat{\boldsymbol{\beta}}$  es insesgado y además, cada  $\widehat{\beta}_i$  es un estimador lineal de varianza mínima de  $\beta_i$ , ya que es centrado y de máxima verosimilitud, luego suficiente. Se llegará a la misma conclusión como consecuencia del Teorema 3.2.1.

Por otra parte, si sustituimos  $\boldsymbol{\beta}$  por  $\widehat{\boldsymbol{\beta}}$  en la función de verosimilitud y derivamos respecto a  $\sigma^2$  resulta que el estimador de máxima verosimilitud de la varianza es

$$\widehat{\sigma}_{MV}^2 = \text{SCR}/n$$

Este estimador es sesgado y en la práctica no se utiliza, ya que disponemos del estimador insesgado propuesto en el apartado anterior. Además, bajo ciertas condiciones generales se puede probar que  $\widehat{\sigma}^2 = \text{SCR}/(n - m)$  es un estimador de varianza mínima de  $\sigma^2$  (véase Seber [66, pág. 52]).

- ii) Como  $\widehat{\boldsymbol{\beta}} = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$ ,  $\widehat{\boldsymbol{\beta}}$  es combinación lineal de una normal y, por tanto, tiene distribución normal multivariante con matriz de varianzas-covarianzas

$$(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

- iii) Es consecuencia de las propiedades de la normal multivariante del apartado anterior ya que

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma^2 = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\text{var}(\widehat{\boldsymbol{\beta}})^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi_m^2$$

- iv) Si calculamos la matriz de covarianzas entre  $\widehat{\boldsymbol{\beta}}$  i  $\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$  tenemos

$$\begin{aligned} \text{cov}(\widehat{\boldsymbol{\beta}}, \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) &= \text{cov}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, (\mathbf{I} - \mathbf{P})\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{var}(\mathbf{Y})(\mathbf{I} - \mathbf{P})' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{P}) = \mathbf{0} \end{aligned}$$

de modo que efectivamente  $\widehat{\boldsymbol{\beta}}$  es independiente de  $(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})$ , ya que la incorrelación entre normales multivariantes implica su independencia.

Este resultado se ampliará en el Teorema 3.4.1.

- v) Aplicando la ecuación 2.5

$$\text{SCR}/\sigma^2 = (z_{m+1}/\sigma)^2 + \cdots + (z_n/\sigma)^2$$

obtenemos una suma de cuadrados de  $n - m$  variables normales independientes, es decir, una distribución  $\chi_{n-m}^2$ .

■

### Ejemplo 2.6.1

La distribución de  $\widehat{\theta}$  del ejemplo 2.4.1 es  $N(\theta, \sigma/\sqrt{6})$

$$\begin{aligned} E(\widehat{\theta}) &= E((y_1 + 2y_2 - y_3)/6) = (1/6)(\theta + 4\theta + \theta) = \theta \\ \text{var}(\widehat{\theta}) &= (\sigma^2 + 4\sigma^2 + \sigma^2)/6^2 = \sigma^2/6 \end{aligned}$$

La distribución de  $\text{SCR}/\sigma^2$  es  $\chi^2_2$ , siendo

$$\text{SCR} = (y_1 - \widehat{\theta})^2 + (y_2 - 2\widehat{\theta})^2 + (y_3 + \widehat{\theta})^2$$

### Ejemplo 2.6.2

La estimación de la varianza del error  $\sigma^2$  en el ejemplo 2.4.2 es

$$\widehat{\sigma}^2 = 0.00145/(5 - 3) = 0.725 \times 10^{-3}$$

Observemos que el número de pesadas necesarias para obtener la misma precisión sería mayor si pesáramos cada objeto individualmente.

## 2.7. Matriz de diseño reducida

Supongamos que varias observaciones  $y_i$  han sido obtenidas bajo las mismas condiciones experimentales. Para estas observaciones, el modelo que liga  $y_i$  con las  $\beta$  es el mismo, lo que se traduce en que las filas de la matriz de diseño correspondientes están repetidas. Para evitar la redundancia que esto supone nos será muy útil, a efectos teóricos y de cálculo, introducir el concepto de matriz de diseño reducida.

### Definición 2.7.1

Dado el modelo lineal  $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ , llamaremos matriz de diseño reducida  $\mathbf{X}_R$  a la matriz  $k \times m$  obtenida tomando las  $k$  filas distintas de la matriz de diseño original  $\mathbf{X}$ . Diremos entonces que  $k$  es el número de condiciones experimentales.

Las matrices de diseño original o ampliada y reducida las indicaremos por  $\mathbf{X}$  y  $\mathbf{X}_R$  respectivamente, cuando convenga distinguir una de otra.

Si la fila  $i$ -ésima de  $\mathbf{X}_R$  está repetida  $n_i$  veces en  $\mathbf{X}$ , significa que se han obtenido  $n_i$  réplicas de la variable observable bajo la  $i$ -ésima condición experimental. Si estos números de réplicas son  $n_1, n_2, \dots, n_k$ , entonces

$$n = n_1 + n_2 + \dots + n_k$$

Además de la matriz reducida  $\mathbf{X}_R$ , utilizaremos también la matriz diagonal

$$\mathbf{D} = \text{diag}(n_1, n_2, \dots, n_k)$$

y el vector de medias

$$\bar{\mathbf{Y}} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k)'$$

donde cada  $\bar{y}_i$  es la media de las réplicas bajo la condición experimental  $i$ .

En una experiencia bajo la cual todas las observaciones han sido tomadas en condiciones experimentales distintas (caso de una sola observación por casilla), entonces

$$\mathbf{X}_R = \mathbf{X} \quad \bar{\mathbf{Y}} = \mathbf{Y} \quad \mathbf{D} = \mathbf{I} \quad n_i = 1$$

Como veremos más adelante (ver sección 11.7), la utilización de  $\mathbf{X}_R$ ,  $\mathbf{D}$  e  $\bar{\mathbf{Y}}$  nos permitirá abordar diseños no balanceados y el caso de observaciones faltantes.

**Teorema 2.7.1**

La solución de las ecuaciones normales y la suma de cuadrados residual en términos de la matriz de diseño reducida  $\mathbf{X}_R$ , de  $\mathbf{D}$  e  $\bar{\mathbf{Y}}$  es

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_R' \mathbf{D} \mathbf{X}_R)^{-1} \mathbf{X}_R' \mathbf{D} \bar{\mathbf{Y}}$$

$$\text{SCR} = \mathbf{Y}' \mathbf{Y} - \hat{\boldsymbol{\beta}}' \mathbf{X}_R' \mathbf{D} \bar{\mathbf{Y}}$$

*Demostración:*

Sea  $\mathbf{M}$  una matriz  $n \times k$  de forma que cada columna  $i$  es

$$(\underbrace{0, \dots, 0}_{n'}, \underbrace{1, \dots, 1}_{n_i}, \underbrace{0, \dots, 0}_{n''})'$$

donde  $k$  es el número de condiciones experimentales (número de filas distintas de  $\mathbf{X}$ ),  $n_i$  el número de réplicas bajo la condición  $i$ , y además

$$n' = n_1 + \dots + n_{i-1} \quad n'' = n_{i+1} + \dots + n_k$$

Se verifica

$$\mathbf{M}' \mathbf{Y} = \mathbf{D} \bar{\mathbf{Y}} \quad \mathbf{M} \mathbf{X}_R = \mathbf{X} \quad \mathbf{M}' \mathbf{M} = \mathbf{D} \quad \mathbf{X}' \mathbf{Y} = \mathbf{X}_R' \mathbf{M}' \mathbf{Y} = \mathbf{X}_R' \mathbf{D} \bar{\mathbf{Y}}$$

de donde se siguen inmediatamente las fórmulas del teorema. ■

**Ejemplo 2.7.1**

Con los datos del ejemplo 2.4.2

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \quad \mathbf{Y} = \begin{pmatrix} 5.53 \\ 1.72 \\ 0.64 \\ 5.48 \\ 1.70 \end{pmatrix}$$

Agrupando las filas 1, 4 y 2, 5 obtenemos

$$\mathbf{X}_R = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

donde  $n_1 = n_2 = 2$ ,  $n_3 = 1$ ,  $k = 3$ .

$$\bar{\mathbf{Y}} = \begin{pmatrix} (5.53 + 5.48)/2 \\ (1.72 + 1.70)/2 \\ 0.64 \end{pmatrix} = \begin{pmatrix} 5.505 \\ 1.710 \\ 0.640 \end{pmatrix}$$

La matriz  $\mathbf{M}$  es

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

### Ejemplo 2.7.2

Consideremos el modelo

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

correspondiente al diseño de dos factores sin interacción.

Supongamos que el primer factor tiene 2 niveles y el segundo tiene 3 niveles, y que los números de réplicas son

$$n_{11} = 2 \quad n_{21} = 1 \quad n_{12} = 3 \quad n_{22} = 3 \quad n_{13} = 5 \quad n_{23} = 4$$

La matriz de diseño reducida es

$\mu$	$\alpha_1$	$\alpha_2$	$\beta_1$	$\beta_2$	$\beta_3$
1	1	0	1	0	0
1	0	1	1	0	0
1	1	0	0	1	0
1	0	1	0	1	0
1	1	0	0	0	1
1	0	1	0	0	1

Sin embargo, la matriz de diseño ampliada tiene 6 columnas y  $\sum n_{ij} = 18$  filas.

## 2.8. Matrices de diseño de rango no máximo

Cuando el modelo lineal corresponde al análisis de los datos de un diseño experimental, la matriz  $\mathbf{X}$  tiene todos sus elementos con valores 0 ó 1 y sus columnas acostumbran a ser linealmente dependientes. Ya sabemos que en este caso es posible hallar el estimador MC de  $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}$  pero, por desgracia, hay múltiples estimaciones de los parámetros  $\boldsymbol{\beta}$  que más bien podemos considerar como soluciones  $\hat{\boldsymbol{\beta}}$  de las ecuaciones normales. En todo caso y como veremos en el próximo capítulo estamos interesados en concretar una estimación de los parámetros  $\boldsymbol{\beta}$  aunque no sea única. A continuación se comentan algunos métodos para hallar una solución  $\hat{\boldsymbol{\beta}}$  o para hallar la SCR directamente.

### 2.8.1. Reducción a un modelo de rango máximo

Sea  $\mathbf{X}_1$  la matriz  $n \times r$  con las  $r = \text{rg } \mathbf{X}$  columnas linealmente independientes de la matriz de diseño  $\mathbf{X}$ , entonces  $\mathbf{P} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$  de forma que

$$\text{SCR} = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\alpha}}'\mathbf{X}_1'\mathbf{Y}$$

donde  $\hat{\boldsymbol{\alpha}} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$  es la solución del modelo  $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\alpha} + \boldsymbol{\epsilon}$  de rango máximo.

Podemos asumir, sin pérdida de generalidad, que  $\mathbf{X}_1$  está formada por las  $r$  primeras filas de  $\mathbf{X}$  de manera que  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ . Entonces  $\mathbf{X}_2 = \mathbf{X}_1\mathbf{F}$  ya que las columnas de  $\mathbf{X}_2$  son linealmente dependientes de las de  $\mathbf{X}_1$  y, por tanto,  $\mathbf{X} = \mathbf{X}_1(\mathbf{I}, \mathbf{F})$ . Así, éste es un caso especial de una factorización más general del tipo

$$\mathbf{X} = \mathbf{KL}$$

donde  $\mathbf{K}$  es  $n \times r$  de rango  $r$ , y  $\mathbf{L}$  es  $r \times m$  de rango  $r$ . Entonces podemos escribir

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{KL}\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\alpha}$$

y estimar  $\boldsymbol{\alpha}$ .



### 2.8.2. Imposición de restricciones

Este método consiste en imponer un conjunto de restricciones del tipo  $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$  para evitar la indeterminación de  $\boldsymbol{\beta}$ . Las restricciones apropiadas, llamadas identificables, son aquellas que, para cada  $\boldsymbol{\theta} \in \Omega = \langle \mathbf{X} \rangle$ , existe un único  $\boldsymbol{\beta}$  que satisface  $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}$  y  $\mathbf{0} = \mathbf{H}\boldsymbol{\beta}$ , es decir, que satisface

$$\begin{pmatrix} \boldsymbol{\theta} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{H} \end{pmatrix} \boldsymbol{\beta} = \mathbf{G}\boldsymbol{\beta}$$

La solución es simple. Debemos elegir como filas de  $\mathbf{H}$  un conjunto de  $m - r$  vectores de longitud  $m$  linealmente independientes que sean también linealmente independientes de las filas de  $\mathbf{X}$ . Entonces la matriz  $\mathbf{G}$  de orden  $(n + m - r) \times m$  tendrá rango  $m$  de modo que  $\mathbf{G}'\mathbf{G} = \mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H}$  es  $m \times m$  de rango  $m$  y en consecuencia tiene inversa. Luego hemos salvado la deficiencia en el rango de  $\mathbf{X}'\mathbf{X}$  introduciendo la matriz  $\mathbf{H}'\mathbf{H}$ .

Así pues, si añadimos  $\mathbf{H}'\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$  a las ecuaciones normales tenemos

$$\mathbf{G}'\mathbf{G}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$$

cuya solución es  $\widehat{\boldsymbol{\beta}} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{X}'\mathbf{Y}$ . Se puede ver, a partir de  $\widehat{\boldsymbol{\theta}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{Y}$ , que  $\mathbf{P} = \mathbf{X}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{X}'$  ya que  $\mathbf{P}$  es única.

La demostración de todos los detalles aquí expuestos puede verse en Seber [66, pág. 74].

Es interesante comprobar que, si  $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$ , entonces

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}) &= (\mathbf{G}'\mathbf{G})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= (\mathbf{G}'\mathbf{G})^{-1}(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})\boldsymbol{\beta} = \boldsymbol{\beta} \end{aligned}$$

de modo que  $\widehat{\boldsymbol{\beta}}$  es un estimador insesgado de  $\boldsymbol{\beta}$ .

Este método es particularmente útil en los modelos de análisis de la varianza para los que  $\mathbf{H}$  se halla con mucha facilidad.

#### Ejemplo 2.8.1

Consideremos el modelo correspondiente al diseño de un factor con, por ejemplo, 3 niveles

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i = 1, 2, 3 \quad j = 1, \dots, n_i$$

entonces, tenemos  $m = 4$  y una matriz de diseño de rango 3. La estimación de los parámetros resulta indeterminada.

Sin embargo, si añadimos la restricción  $\sum \alpha_i = 0$ , es decir, si hacemos  $\mathbf{H} = (0, 1, 1, 1)$ , el sistema conjunto es de rango 4 y podemos determinar una solución o calcular la suma de cuadrados residual.

## 2.9. Ejercicios

### Ejercicio 2.1

Una variable  $Y$  toma los valores  $y_1, y_2$  y  $y_3$  en función de otra variable  $X$  con los valores  $x_1, x_2$  y  $x_3$ . Determinar cuales de los siguientes modelos son lineales y encontrar, en su caso, la matriz de diseño para  $x_1 = 1, x_2 = 2$  y  $x_3 = 3$ .

a)  $y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i^2 - 1) + \epsilon_i$

b)  $y_i = \beta_0 + \beta_1 x_i + \beta_2 e^{x_i} + \epsilon_i$

c)  $y_i = \beta_1 x_i (\beta_2 \tan(x_i)) + \epsilon_i$

**Ejercicio 2.2**

Dado el modelo lineal

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \theta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

hallar la estimación MC de  $\theta$  y la suma de cuadrados residual.

**Ejercicio 2.3**

Si  $\widehat{\boldsymbol{\beta}}$  es una estimación MC, probar que

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

**Ejercicio 2.4**

Cuatro objetos cuyos pesos exactos son  $\beta_1, \beta_2, \beta_3$  y  $\beta_4$  han sido pesados en una balanza de platillos de acuerdo con el siguiente esquema:

$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	peso
1	1	1	1	9.2
1	-1	1	1	8.3
1	0	0	1	5.4
1	0	0	-1	-1.6
1	0	1	1	8.7
1	1	-1	1	3.5

Hallar las estimaciones de cada  $\beta_i$  y de la varianza del error.

**Ejercicio 2.5**

Sea  $\widehat{\boldsymbol{\beta}}$  la estimación MC de  $\boldsymbol{\beta}$ . Si  $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{Y}$ , probar que la matriz  $\mathbf{P}$  verifica

$$\mathbf{P}^2 = \mathbf{P} \quad (\mathbf{I} - \mathbf{P})^2 = \mathbf{I} - \mathbf{P}$$

**Ejercicio 2.6**

La matriz de diseño reducida de un modelo lineal normal es

$$\mathbf{X}_R = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Se sabe además que

$$\bar{y}_1 = 10 \quad \bar{y}_2 = 12 \quad \bar{y}_3 = 17 \quad n_1 = n_2 = n_3 = 10$$

$$s_1^2 = \frac{1}{n_1} \sum (y_{i1} - \bar{y}_1)^2 = 2.8 \quad s_2^2 = 4.2 \quad s_3^2 = 4.0$$

Se pide:

- Hallar la expresión general de las estimaciones MC de los parámetros  $\boldsymbol{\beta}$ .
- Calcular SCR. Estimar la varianza del diseño  $\sigma^2$ .
- Estudiar si la hipótesis nula  $H_0 : \sigma^2 = 3$  puede ser aceptada.

**Ejercicio 2.7**

Consideremos el modelo lineal

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \epsilon_i \quad i = 1, \dots, n$$

Sean  $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_m$  las estimaciones MC de los parámetros y sea

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_m x_{im} \quad i = 1, \dots, n$$

Probar que

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



## Funciones paramétricas estimables

### 3.1. Introducción

En los modelos lineales, además de la estimación de los parámetros  $\beta_i$  y de  $\sigma^2$ , interesa también la estimación de ciertas funciones lineales de los parámetros. Como vamos a ver, esto es especialmente necesario cuando los parámetros carecen de una estimación única.

#### Definición 3.1.1

Llamaremos función paramétrica a toda función lineal  $\psi$  de los parámetros

$$\psi = a_1\beta_1 + \cdots + a_m\beta_m = \mathbf{a}'\boldsymbol{\beta}$$

y diremos que una función paramétrica  $\psi$  es estimable si existe un estadístico  $\widehat{\psi}$ , combinación lineal de las observaciones  $y_1, \dots, y_n$

$$\widehat{\psi} = b_1y_1 + \cdots + b_ny_n = \mathbf{b}'\mathbf{Y}$$

tal que

$$E(\widehat{\psi}) = \psi$$

es decir,  $\widehat{\psi}$  es estimador lineal insesgado de  $\psi$ .

Estas funciones paramétricas tienen la siguiente caracterización

#### Teorema 3.1.1

Sea  $\psi = \mathbf{a}'\boldsymbol{\beta}$  una función paramétrica estimable asociada al modelo lineal  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ . Se verifica:

- i)  $\psi$  es estimable si y sólo si el vector fila  $\mathbf{a}'$  es combinación lineal de las filas de  $\mathbf{X}$ .
- ii) Si  $\psi_1, \dots, \psi_q$  son funciones paramétricas estimables, entonces la combinación lineal  $\psi = c_1\psi_1 + \cdots + c_q\psi_q$  es también función paramétrica estimable.
- iii) El número máximo de funciones paramétricas estimables linealmente independientes es  $r = \text{rango}(\mathbf{X})$ .

*Demostración:*

- i) Sea  $\widehat{\psi} = \mathbf{b}'\mathbf{Y}$  tal que  $E(\widehat{\psi}) = \psi$ . Entonces

$$\mathbf{a}'\boldsymbol{\beta} = E(\mathbf{b}'\mathbf{Y}) = \mathbf{b}'E(\mathbf{Y}) = \mathbf{b}'\mathbf{X}\boldsymbol{\beta}$$

cualquiera que sea  $\boldsymbol{\beta}$ , luego

$$\mathbf{a}' = \mathbf{b}'\mathbf{X}$$

lo que nos dice que  $\mathbf{a}'$  es combinación lineal de las filas de la matriz de diseño  $\mathbf{X}$ .

Recíprocamente, si suponemos que  $\mathbf{b}'\mathbf{X} = \mathbf{a}'$ , entonces basta tomar  $\widehat{\psi} = \mathbf{b}'\mathbf{Y}$  como estimador lineal insesgado de  $\psi$ .

ii) y iii) para el lector (ver ejercicio 3.4) ■

### Observaciones:

- 1) Si rango  $\mathbf{X} = m$ , entonces todos los parámetros  $\beta_i$  y todas las funciones paramétricas  $\psi$  son estimables, pues el subespacio generado por las filas de  $\mathbf{X}$  coincide con  $\mathbb{R}^m$ .
- 2) Si rango  $\mathbf{X} < m$ , pueden construirse funciones paramétricas que no son estimables.
- 3) Una caracterización algebraica de que  $\psi = \mathbf{a}'\boldsymbol{\beta}$  es estimable viene dada por la identidad

$$\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{a}'$$

donde  $(\mathbf{X}'\mathbf{X})^{-}$  representa una g-inversa de  $\mathbf{X}'\mathbf{X}$ .

En efecto, consideremos las matrices

$$\mathbf{S} = \mathbf{X}'\mathbf{X} \quad \mathbf{S}^{-} = (\mathbf{X}'\mathbf{X})^{-} \quad \mathbf{H} = \mathbf{S}^{-}\mathbf{S}$$

entonces se comprueba fácilmente que

$$\mathbf{H}^2 = \mathbf{H} \quad \mathbf{SH} = \mathbf{S}$$

Puesto que  $\mathbf{H}$  es idempotente

$$\text{rango } \mathbf{H} = \text{traza } \mathbf{H} = \text{rango } \mathbf{S} = \text{rango } \mathbf{X} = r$$

Por otra parte tenemos

$$\begin{aligned} \mathbf{0} &= \mathbf{S} - \mathbf{SH} = (\mathbf{I}_m - \mathbf{H})'(\mathbf{S} - \mathbf{SH}) = (\mathbf{I}_m - \mathbf{H})'(\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{XH}) \\ &= (\mathbf{I}_m - \mathbf{H})'(\mathbf{X}'(\mathbf{X} - \mathbf{XH})) = (\mathbf{X} - \mathbf{XH})'(\mathbf{X} - \mathbf{XH}) \end{aligned}$$

luego

$$\mathbf{X} = \mathbf{XH}$$

Entonces, si  $\psi = \mathbf{a}'\boldsymbol{\beta}$  es estimable,  $\mathbf{a}' = \mathbf{b}'\mathbf{X}$  y

$$\mathbf{a}'\mathbf{H} = \mathbf{b}'\mathbf{XH} = \mathbf{b}'\mathbf{X} = \mathbf{a}'$$

Recíprocamente, si  $\mathbf{a}'\mathbf{H} = \mathbf{a}'$ , resulta que

$$\mathbf{a}' = \mathbf{a}'\mathbf{S}^{-}\mathbf{S} = (\mathbf{a}'\mathbf{S}^{-}\mathbf{X}')\mathbf{X} = \mathbf{b}'\mathbf{X}$$

siendo  $\mathbf{b}' = \mathbf{a}'\mathbf{S}^{-}\mathbf{X}'$ .

## 3.2. Teorema de Gauss-Markov

Vamos a ver en primer lugar que, cuando el rango de la matriz de diseño no es máximo y, por tanto, la estimación MC de los parámetros no es única, la estimación de cualquier función paramétrica estimable utilizando cualquiera de los estimadores MC sí es única.

**Teorema 3.2.1**

Si  $\psi = \mathbf{a}'\boldsymbol{\beta}$  una función paramétrica estimable y  $\widehat{\boldsymbol{\beta}}$  es un estimador MC de  $\boldsymbol{\beta}$ , entonces el estimador  $\widehat{\psi} = \mathbf{a}'\widehat{\boldsymbol{\beta}}$  de  $\psi$  es único.

*Demostración:*

Si  $\psi$  es una función paramétrica estimable, tiene un estimador lineal insesgado  $\widetilde{\mathbf{b}}'\mathbf{Y}$ , donde  $\widetilde{\mathbf{b}}$  es un vector  $n \times 1$ . Consideremos el subespacio  $\Omega = \langle \mathbf{X} \rangle$  de  $\mathbb{R}^n$  generado por las columnas de  $\mathbf{X}$ . El vector  $\widetilde{\mathbf{b}}$  se puede descomponer de forma única

$$\widetilde{\mathbf{b}} = \widetilde{\mathbf{b}} + \mathbf{c} \quad \widetilde{\mathbf{b}} \in \Omega \quad \mathbf{c} \perp \Omega$$

de modo que  $\mathbf{c}$  es ortogonal a todo vector de  $\Omega$ .

Consideremos ahora el estimador lineal  $\widetilde{\mathbf{b}}'\mathbf{Y}$  y veamos que es insesgado y que su valor es único. Sabemos que  $\widetilde{\mathbf{b}}'\mathbf{Y}$  es insesgado

$$\psi = \mathbf{a}'\boldsymbol{\beta} = E(\mathbf{b}'\mathbf{Y}) = E(\widetilde{\mathbf{b}}'\mathbf{Y}) + E(\mathbf{c}'\mathbf{Y}) = E(\widetilde{\mathbf{b}}'\mathbf{Y}) = \widetilde{\mathbf{b}}'\mathbf{X}\boldsymbol{\beta} \quad (3.1)$$

luego  $E(\widetilde{\mathbf{b}}'\mathbf{Y}) = \mathbf{a}'\boldsymbol{\beta}$ , pues

$$E(\mathbf{c}'\mathbf{Y}) = \mathbf{c}'E(\mathbf{Y}) = \mathbf{c}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0}\boldsymbol{\beta} = \mathbf{0}$$

Supongamos que  $\mathbf{b}^{*'}\mathbf{Y}$  es otro estimador insesgado para  $\psi$  y  $\mathbf{b}^* \in \Omega$ . Entonces

$$\mathbf{0} = E(\widetilde{\mathbf{b}}'\mathbf{Y}) - E(\mathbf{b}^{*'}\mathbf{Y}) = (\widetilde{\mathbf{b}} - \mathbf{b}^*)'\mathbf{X}\boldsymbol{\beta}$$

luego

$$(\widetilde{\mathbf{b}} - \mathbf{b}^*)'\mathbf{X} = \mathbf{0}$$

lo que quiere decir que  $(\widetilde{\mathbf{b}} - \mathbf{b}^*)$  es ortogonal a  $\Omega$ . Como también pertenece a  $\Omega$ , debe ser  $\widetilde{\mathbf{b}} - \mathbf{b}^* = \mathbf{0}$ , es decir,  $\widetilde{\mathbf{b}} = \mathbf{b}^*$ .

Por último, sabemos que para cualquier estimador MC de  $\boldsymbol{\beta}$   $\mathbf{e} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$  es ortogonal a  $\Omega$ , de manera que

$$\mathbf{0} = \widetilde{\mathbf{b}}'\mathbf{e} = \widetilde{\mathbf{b}}'\mathbf{Y} - \widetilde{\mathbf{b}}'\mathbf{X}\widehat{\boldsymbol{\beta}}$$

y así  $\widetilde{\mathbf{b}}'\mathbf{Y} = \widetilde{\mathbf{b}}'\mathbf{X}\widehat{\boldsymbol{\beta}}$ . Además, por 3.1 sabemos que  $\widetilde{\mathbf{b}}'\mathbf{X} = \mathbf{b}'\mathbf{X} = \mathbf{a}'$ , luego

$$\widetilde{\mathbf{b}}'\mathbf{Y} = \mathbf{a}'\widehat{\boldsymbol{\beta}}$$

para cualquier  $\widehat{\boldsymbol{\beta}}$ . ■

A continuación se demuestra la principal ventaja de la utilización de los estimadores MC.

**Teorema 3.2.2 (Gauss-Markov)**

Si  $\psi = \mathbf{a}'\boldsymbol{\beta}$  una función paramétrica estimable y  $\widehat{\boldsymbol{\beta}}$  es un estimador MC de  $\boldsymbol{\beta}$ , entonces  $\widehat{\psi} = \mathbf{a}'\widehat{\boldsymbol{\beta}}$  es el estimador de varianza mínima<sup>1</sup> en la clase de los estimadores lineales insesgados de  $\psi$ .

*Demostración:*

Con la notación

$$\|\mathbf{b}\|^2 = b_1^2 + \cdots + b_n^2$$

tenemos que

$$\text{var}(\mathbf{b}'\mathbf{Y}) = b_1^2\sigma^2 + \cdots + b_n^2\sigma^2 = \|\mathbf{b}\|^2\sigma^2$$

1. BLUE: best linear unbiased estimate

Si consideramos la descomposición de cualquier estimador insesgado de  $\psi$  que hemos utilizado en el teorema anterior y dado que

$$\|\mathbf{b}\|^2 = \|\widetilde{\mathbf{b}}\|^2 + \|\mathbf{c}\|^2$$

resulta

$$\text{var}(\mathbf{a}'\widehat{\boldsymbol{\beta}}) = \text{var}(\widetilde{\mathbf{b}}'\mathbf{Y}) = \|\widetilde{\mathbf{b}}\|^2\sigma^2 \leq (\|\widetilde{\mathbf{b}}\|^2 + \|\mathbf{c}\|^2)\sigma^2 = \text{var}(\mathbf{b}'\mathbf{Y})$$

### Observaciones:

- 1) Estos resultados son válidos incluso para un modelo lineal sin la hipótesis de normalidad.
- 2) La estimación con varianza mínima es

$$\widehat{\psi} = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- 3) Como la varianza de  $\mathbf{b}'\mathbf{Y}$  es  $\mathbf{b}'\mathbf{b}\sigma^2$ , resulta que la varianza mínima es

$$\text{var}(\widehat{\psi}) = \text{var}(\mathbf{a}'\widehat{\boldsymbol{\beta}}) = \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}$$

- 4) Utilizando la matriz de diseño reducida tenemos

$$\widehat{\psi} = \mathbf{a}'(\mathbf{X}'_R\mathbf{D}\mathbf{X}_R)^{-1}\mathbf{X}'_R\mathbf{D}\bar{\mathbf{Y}}$$

$$\text{var}(\widehat{\psi}) = \sigma^2\mathbf{a}'(\mathbf{X}'_R\mathbf{D}\mathbf{X}_R)^{-1}\mathbf{a}$$

De aquí deducimos que  $\widehat{\psi}$  es combinación lineal de las medias de las  $k$  condiciones experimentales

$$\widehat{\psi} = c_1\bar{Y}_1 + \dots + c_k\bar{Y}_k = \mathbf{c}'\bar{\mathbf{Y}}$$

donde  $\mathbf{c} = (c_1, \dots, c_k)'$  es

$$\mathbf{c} = \mathbf{D}\mathbf{X}_R(\mathbf{X}'_R\mathbf{D}\mathbf{X}_R)^{-1}\mathbf{a}$$

Entonces

$$\text{var}(\widehat{\psi}) = \left( \sum_{i=1}^k c_i^2/n_i \right) \sigma^2 = \delta^2\sigma^2$$

Por otra parte, todo estimador lineal insesgado  $\widehat{\psi} = \mathbf{b}'\mathbf{Y}$  de  $\psi = \mathbf{a}'\boldsymbol{\beta}$  se descompone como hemos visto en

$$\mathbf{b}'\mathbf{Y} = \widetilde{\mathbf{b}}'\mathbf{Y} + \mathbf{c}'\mathbf{Y}$$

Diremos que  $\widetilde{\mathbf{b}}'\mathbf{Y}$  (donde  $\widetilde{\mathbf{b}}$  es único) pertenece al *espacio estimación* y que  $\mathbf{c}'\mathbf{Y}$  pertenece al *espacio error*.

Más explícitamente, la descomposición de  $\mathbf{b}'$  es

$$\mathbf{b}' = \mathbf{b}'\mathbf{P} + \mathbf{b}'(\mathbf{I} - \mathbf{P})$$

siendo  $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  la matriz del operador que proyecta  $\mathbf{b}$  en  $\Omega = \langle \mathbf{X} \rangle$  (ver Apéndice B). El vector proyectado es  $\widetilde{\mathbf{b}} = \mathbf{b}'\mathbf{P}$ . Asimismo,  $\mathbf{I} - \mathbf{P}$  es otro operador que proyecta  $\mathbf{b}$  en el espacio ortogonal a  $\Omega$ . La proyección es  $\mathbf{c}' = \mathbf{b}'(\mathbf{I} - \mathbf{P})$ . Como  $\widetilde{\mathbf{b}}'\mathbf{c} = \mathbf{0}$ , se verifica

$$\text{cov}(\widetilde{\mathbf{b}}'\mathbf{Y}, \mathbf{c}'\mathbf{Y}) = 0$$

Así pues, todo estimador lineal insesgado  $\mathbf{b}'\mathbf{Y}$  se descompone en

$$\mathbf{b}'\mathbf{Y} = \mathbf{b}'\mathbf{P}\mathbf{Y} + \mathbf{b}'(\mathbf{I} - \mathbf{P})\mathbf{Y}$$



donde  $\mathbf{b}'\mathbf{P}\mathbf{Y}$  es el estimador de Gauss-Markov, mientras que  $\mathbf{b}'(\mathbf{I} - \mathbf{P})\mathbf{Y}$  tiene esperanza cero y provoca un aumento de la varianza mínima del mejor estimador  $\widehat{\psi} = \mathbf{b}'\mathbf{P}\mathbf{Y}$ .

Finalmente, observemos que

$$\begin{aligned}\widehat{\psi} &= \mathbf{b}'\mathbf{P}\mathbf{Y} = \mathbf{b}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{b}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \\ &= \mathbf{b}'\mathbf{X}\mathbf{H}\widehat{\boldsymbol{\beta}} = \mathbf{a}'\widehat{\boldsymbol{\beta}}\end{aligned}\quad (3.2)$$

Siendo  $\mathbf{H} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}$ , que verifica  $\mathbf{X}\mathbf{H} = \mathbf{X}$ , y siendo  $\mathbf{a}' = \mathbf{b}'\mathbf{X}$ .

El aspecto geométrico de las estimaciones se puede resumir en el hecho que el espacio muestral  $\mathbb{R}^n$  al que pertenece el vector de observaciones  $\mathbf{Y}$ , se descompone en

$$\mathbb{R}^n = \Omega + \Omega^\perp$$

donde  $\Omega$  representa el *espacio estimación*. Toda estimación de los parámetros de regresión está ligada a  $\Omega$ . Toda estimación de la varianza del modelo está ligada al *espacio error*  $\Omega^\perp$ . Ambos espacios son ortogonales y bajo el modelo lineal normal, como veremos más adelante, ambas clases de estimaciones son estocásticamente independientes.

### Ejemplo 3.2.1

Sea  $y_1, \dots, y_n$  una muestra aleatoria simple procedente de una población  $N(\mu, \sigma)$ . El modelo lineal asociado es

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \mu + \boldsymbol{\epsilon}$$

El estimador MC de  $\mu$  es  $\widehat{\mu} = (1/n) \sum y_i$  que también es de Gauss-Markov (centrado y de varianza mínima).

En este caso  $\mathbb{R}^n = \Omega + \Omega^\perp$ , siendo

$$\begin{aligned}\Omega &= \langle (1, \dots, 1)' \rangle \\ \Omega^\perp &= \{(x_1, \dots, x_n)' \mid \sum x_i = 0\}\end{aligned}$$

Sea  $\mathbf{a}'\mathbf{Y} = \sum a_i y_i$  otro estimador centrado de  $\mu$ . Entonces  $E(\mathbf{a}'\mathbf{Y}) = \mu$  implica  $\sum a_i = 1$ . Luego se verifica  $\mathbf{a} = \widetilde{\mathbf{a}} + \mathbf{b}$ , es decir,

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix} + \begin{pmatrix} a_1 - 1/n \\ \vdots \\ a_n - 1/n \end{pmatrix}$$

con  $\widetilde{\mathbf{a}} \in \Omega$ ,  $\mathbf{b} \in \Omega^\perp$ . Es fácil ver que  $\widetilde{\mathbf{a}}'\mathbf{b} = 0$ . Además

$$\sum a_i y_i = (1/n) \sum y_i + \sum (a_i - 1/n) y_i$$

El primer término es estimador centrado y de varianza mínima  $\sigma^2/n$ . El segundo término verifica

$$\begin{aligned}E\left(\sum (a_i - 1/n) y_i\right) &= 0 \\ \text{cov}\left(1/n \sum y_i, \sum (a_i - 1/n) y_i\right) &= 0\end{aligned}$$

La matriz del operador que proyecta  $\mathbf{a}$  en  $\Omega$  es

$$\mathbf{P} = 1/n \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} (1, \dots, 1) = \begin{pmatrix} 1/n & \dots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \dots & 1/n \end{pmatrix}$$

siendo fácil ver que

$$\begin{aligned}\mathbf{a}'\mathbf{P} &= (1/n, \dots, 1/n) \\ \mathbf{a}'(\mathbf{I} - \mathbf{P}) &= (a_1 - 1/n, \dots, a_n - 1/n)\end{aligned}$$

### Ejemplo 3.2.2

Ver especialmente el final del ejemplo 5.3.2.

## 3.3. Varianza de la estimación y multicolinealidad

Sabemos que  $\mathbf{a}'\boldsymbol{\beta}$  se dice estimable si tiene un estimador lineal insesgado  $\mathbf{b}'\mathbf{Y}$  o, equivalentemente, cuando  $\mathbf{a} = \mathbf{X}'\mathbf{b}$ . Es decir, cuando  $\mathbf{a}$  es combinación lineal de las filas de la matriz  $\mathbf{X}$ .

### Teorema 3.3.1

La función paramétrica  $\mathbf{a}'\boldsymbol{\beta}$  es estimable si y sólo si

$$a \in \langle \mathbf{X}' \rangle = \langle \mathbf{X}'\mathbf{X} \rangle$$

#### Demostración:

Como sabemos, la función paramétrica  $\mathbf{a}'\boldsymbol{\beta}$  es estimable si y sólo si  $\mathbf{a}$  es combinación lineal de las filas de  $\mathbf{X}$ , es decir, cuando  $a \in \langle \mathbf{X}' \rangle$ . De modo que sólo queda probar que

$$\langle \mathbf{X}' \rangle = \langle \mathbf{X}'\mathbf{X} \rangle$$

Pero  $\mathbf{X}'\mathbf{X}\mathbf{c} = \mathbf{X}'\mathbf{d}$  para  $\mathbf{d} = \mathbf{X}\mathbf{c}$ , de forma que  $\langle \mathbf{X}'\mathbf{X} \rangle \subset \langle \mathbf{X}' \rangle$ . Además, las dimensiones de ambos subespacios son iguales ya que  $\text{rg } \mathbf{X}' = \text{rg } \mathbf{X}'\mathbf{X}$ , de donde se deduce la igualdad.

Los detalles pueden verse en Seber [66, pág. 385]. ■

En el apartado anterior hemos demostrado que para una función paramétrica estimable  $\mathbf{a}'\boldsymbol{\beta}$ , su estimador MC  $\widehat{\mathbf{a}'\boldsymbol{\beta}}$  es el de mínima varianza. Pero, ¿cuanto vale esta varianza?

Supongamos que  $\mathbf{X}'\mathbf{X}$  tiene como valores propios  $\lambda_1, \dots, \lambda_r$  todos positivos no nulos asociados a los correspondientes vectores propios ortonormales  $\mathbf{v}_1, \dots, \mathbf{v}_r$ , es decir

$$\mathbf{X}'\mathbf{X}\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad i = 1, \dots, r$$

y tales que  $\mathbf{v}_i'\mathbf{v}_j = \delta_{ij}$ .

Si  $\mathbf{a}'\boldsymbol{\beta}$  es estimable, entonces  $\mathbf{a} \in \langle \mathbf{X}'\mathbf{X} \rangle$  y este subespacio está generado por los vectores propios. Así pues,  $\mathbf{a}$  se puede expresar en la forma

$$\mathbf{a} = \sum_{i=1}^r c_i \mathbf{v}_i$$

Entonces

$$\begin{aligned}\text{var}(\widehat{\mathbf{a}'\boldsymbol{\beta}}) &= \text{var}\left(\sum_i c_i \widehat{\mathbf{v}_i'\boldsymbol{\beta}}\right) \\ &= \sum_i c_i^2 \text{var}(\widehat{\mathbf{v}_i'\boldsymbol{\beta}}) \\ &= \sigma^2 \sum_i c_i^2 \lambda_i^{-1}\end{aligned}$$

ya que

$$\begin{aligned}
 \text{cov}(\mathbf{v}_i' \hat{\boldsymbol{\beta}}, \mathbf{v}_j' \hat{\boldsymbol{\beta}}) &= \lambda_i^{-1} \lambda_j^{-1} \text{cov}(\mathbf{v}_i' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}, \mathbf{v}_j' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}) \\
 &= (\lambda_i \lambda_j)^{-1} \text{cov}(\mathbf{v}_i' \mathbf{X}' \mathbf{Y}, \mathbf{v}_j' \mathbf{X}' \mathbf{Y}) \\
 &= (\lambda_i \lambda_j)^{-1} \sigma^2 \mathbf{v}_i' \mathbf{X}' \mathbf{X} \mathbf{v}_j \\
 &= (\lambda_i \lambda_j)^{-1} \sigma^2 \lambda_j \mathbf{v}_i' \mathbf{v}_j \\
 &= \sigma^2 \lambda_i^{-1} \delta_{ij}
 \end{aligned}$$

Silvey[68] concluyó que es posible una estimación relativamente precisa en las direcciones de los vectores propios de  $\mathbf{X}'\mathbf{X}$  correspondientes a los mayores valores propios, mientras que se obtienen unas estimaciones relativamente imprecisas (poco eficientes) en las direcciones correspondientes a los valores propios más pequeños.

Supongamos que  $\mathbf{X}$  tiene rango máximo pero que sus columnas están cerca de ser linealmente dependientes. Entonces  $\mathbf{X}'\mathbf{X}$  está cerca de ser singular (no inversible), en el sentido que uno o varios de sus valores propios no nulos son excesivamente pequeños, casi despreciables, y por lo que hemos visto las estimaciones en algunas direcciones serán muy imprecisas.

La presencia de relaciones quasi lineales entre las variables regresoras se conoce en Econometría con el nombre de *multicolinealidad*, cuya forma más extrema se presenta cuando la matriz de datos  $\mathbf{X}$  no tiene rango máximo. Este grave problema debe ser detectado previamente a la estimación y se puede corregir de varias formas (ver sección 8.5).

Una solución teórica consiste en minimizar o incluso erradicar la multicolinealidad, mediante la incorporación de nuevas observaciones en las direcciones de los vectores propios con valores propios demasiado pequeños (o cero).

Supongamos que una nueva observación se añade al modelo  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  y resulta

$$\begin{aligned}
 \begin{pmatrix} \mathbf{Y} \\ Y_{n+1} \end{pmatrix} &= \begin{pmatrix} \mathbf{X} \\ \mathbf{x}_{n+1}' \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\epsilon} \\ \epsilon_{n+1} \end{pmatrix} \\
 &= \mathbf{X}_* \boldsymbol{\beta} + \boldsymbol{\epsilon}_*
 \end{aligned}$$

donde  $\mathbf{x}_{n+1} = c\mathbf{v}$ , donde  $\mathbf{v}$  es un vector propio normalizado de  $\mathbf{X}'\mathbf{X}$  correspondiente a un valor propio  $\lambda$ . Entonces se puede probar que  $\mathbf{v}$  es también un vector propio de  $\mathbf{X}_*'\mathbf{X}_*$  correspondiente al valor propio  $\lambda + c^2$ . Y de esta forma Sylvey propuso un análisis para la elección de las direcciones en las que es conveniente elegir nuevas observaciones para mejorar la precisión de las estimaciones de un  $\mathbf{a}'\boldsymbol{\beta}$  particular.

### 3.4. Sistemas de funciones paramétricas estimables

Consideremos un sistema de funciones paramétricas estimables

$$\psi_1 = \mathbf{a}_1' \boldsymbol{\beta}, \dots, \psi_q = \mathbf{a}_q' \boldsymbol{\beta}$$

sobre el mismo modelo lineal normal y donde los vectores  $\mathbf{a}_1, \dots, \mathbf{a}_q$  ( $q \leq r = \text{rango } \mathbf{X}$ ) son linealmente independientes. Para cada una, tenemos las correspondientes estimaciones de Gauss-Markov

$$\widehat{\psi}_i = \mathbf{a}_i' \widehat{\boldsymbol{\beta}} \quad i = 1, \dots, q$$

que podemos condensar matricialmente en la forma

$$\widehat{\boldsymbol{\psi}} = (\widehat{\psi}_1, \dots, \widehat{\psi}_q)' = \mathbf{A} \widehat{\boldsymbol{\beta}}$$

donde

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_q' \end{pmatrix}$$

Con esta matriz,  $\widehat{\boldsymbol{\psi}}$  es el conjunto de estimadores MC del sistema de funciones paramétricas  $\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\beta}$ .

### Teorema 3.4.1

Bajo el modelo lineal normal, el conjunto de estimadores  $\widehat{\boldsymbol{\psi}} = \mathbf{A}\widehat{\boldsymbol{\beta}}$  del sistema de funciones paramétricas  $\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\beta}$  verifica:

- i)  $\widehat{\boldsymbol{\psi}}$  sigue la distribución normal multivariante

$$\widehat{\boldsymbol{\psi}} \sim N_q(\boldsymbol{\psi}, \boldsymbol{\Sigma}_{\boldsymbol{\psi}})$$

donde  $\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\beta}$  es el vector de medias y

$$\boldsymbol{\Sigma}_{\boldsymbol{\psi}} = \sigma^2 \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'$$

es la matriz de varianzas-covarianzas.

- ii) La estimación MC de toda función paramétrica estimable es estocásticamente independiente de la suma de cuadrados residual

$$\text{SCR} = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})$$

En particular,  $\widehat{\boldsymbol{\psi}} = \mathbf{A}\widehat{\boldsymbol{\beta}}$  es estocásticamente independiente de SCR.

*Demostración:*

- i) Es consecuencia de que  $\widehat{\boldsymbol{\psi}}$  es una combinación lineal de variables normales independientes:

$$\widehat{\boldsymbol{\psi}}_i = \mathbf{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

luego si

$$\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{C}$$

sabemos que  $E(\widehat{\boldsymbol{\psi}}) = \boldsymbol{\psi}$  y la matriz de covarianzas de  $\mathbf{CY}$  es  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{CC}'$ , de manera que

$$\boldsymbol{\Sigma}_{\boldsymbol{\psi}} = \sigma^2 \mathbf{CC}' = \sigma^2 \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}' = \sigma^2 \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'$$

- ii) Como en el teorema 2.5.1, consideremos la transformación ortogonal

$$\mathbf{Z} = \mathbf{V}'\mathbf{Y}$$

donde las primeras  $r$  columnas de la matriz ortogonal  $\mathbf{V}$  generan el subespacio  $\Omega = \langle \mathbf{X} \rangle$ . Entonces las variables  $z_1, \dots, z_n$  son normales e independientes, y toda estimación de Gauss-Markov es una combinación lineal de

$$z_1, \dots, z_r$$

puesto que pertenece al *espacio estimación*. Sin embargo, la suma de cuadrados residual es

$$\text{SCR} = z_{r+1}^2 + \dots + z_n^2$$

y, por tanto, será estocásticamente independiente de cualquier estimación  $\widehat{\boldsymbol{\psi}}_i = \mathbf{a}_i'\widehat{\boldsymbol{\beta}}$ .

Esto mismo se puede deducir de la expresión 3.2 ya que  $\widehat{\boldsymbol{\psi}} = \mathbf{B}\mathbf{P}\mathbf{Y}$ , mientras que

$$\text{SCR} = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y} = ((\mathbf{I} - \mathbf{P})\mathbf{Y})'(\mathbf{I} - \mathbf{P})\mathbf{Y}$$

donde  $(\mathbf{I} - \mathbf{P})\mathbf{Y}$  pertenece al espacio ortogonal de  $\Omega$ .



**Teorema 3.4.2**

La distribución de  $U = (\widehat{\mathbf{A}\beta} - \mathbf{A}\beta)'(\sigma^2 \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1}(\widehat{\mathbf{A}\beta} - \mathbf{A}\beta)$  es una  $\chi_q^2$ .  
Además,  $U$  es estocásticamente independiente de  $\text{SCR}/\sigma^2$  cuya distribución es  $\chi_{n-r}^2$ .

*Demostración:*

Es consecuencia de las propiedades de la distribución normal multivariante y de los teoremas 2.5.1 y 3.4.1. ■

Dos resultados importantes que se deducen de los teoremas anteriores son:

- a) Para el modelo lineal normal y el sistema de  $q$  funciones paramétricas estimables  $\psi = \mathbf{A}\beta$  se verifica que la distribución de

$$F = \frac{(\widehat{\mathbf{A}\beta} - \mathbf{A}\beta)'(\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1}(\widehat{\mathbf{A}\beta} - \mathbf{A}\beta)/q}{\text{SCR}/(n-r)} \quad (3.3)$$

es una  $F$  con  $q$  y  $n-r$  grados de libertad, ya que se trata de un cociente de dos  $\chi^2$  independientes divididas por sus grados de libertad respectivos. Observemos la desaparición del parámetro  $\sigma^2$  desconocido.

- b) En el caso  $q = 1$ , si  $\widehat{\psi}$  es la estimación de Gauss-Markov de  $\psi$ , entonces  $\widehat{\psi} \sim N(\psi, \sigma_{\widehat{\psi}}^2)$ , siendo

$$\sigma_{\widehat{\psi}}^2 = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\sigma^2 = \delta^2\sigma^2$$

luego la distribución de

$$t = \frac{\widehat{\psi} - \psi}{\sqrt{\delta^2\text{SCR}}} \sqrt{n-r} \quad (3.4)$$

es la de una  $t$  de Student con  $n-r$  grados de libertad. Este resultado se puede establecer directamente o a partir de 3.3 ya que  $F_{1,n-r} = t_{n-r}^2$ .

**3.5. Intervalos de confianza**

Consideremos una función paramétrica estimable  $\psi = \mathbf{a}'\beta$ , su estimación MC  $\widehat{\psi} = \mathbf{a}'\widehat{\beta}$  y sea  $t_\alpha$  tal que

$$P(-t_\alpha < t < t_\alpha) = 1 - \alpha$$

para una distribución  $t$  de Student con  $n-r$  grados de libertad. Entonces, de la distribución 3.4 deducimos que

$$P\left(-t_\alpha < \frac{\widehat{\psi} - \psi}{\sqrt{\delta^2\text{SCR}}} \sqrt{n-r} < t_\alpha\right) = 1 - \alpha$$

y despejando obtenemos

$$P\left(\widehat{\psi} - t_\alpha \sqrt{\frac{\delta^2\text{SCR}}{n-r}} < \psi < \widehat{\psi} + t_\alpha \sqrt{\frac{\delta^2\text{SCR}}{n-r}}\right) = 1 - \alpha$$

Por lo tanto

$$\widehat{\psi} - t_\alpha \sqrt{\frac{\delta^2\text{SCR}}{n-r}} < \psi < \widehat{\psi} + t_\alpha \sqrt{\frac{\delta^2\text{SCR}}{n-r}}$$

es decir

$$\mathbf{a}'\widehat{\beta} \pm t_\alpha [\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\widehat{\sigma}^2]^{1/2} \quad (3.5)$$

es un intervalo de confianza para la función paramétrica estimable  $\psi = \mathbf{a}'\beta$ , con coeficiente de confianza  $1 - \alpha$ .

Por otra parte, como  $SCR/\sigma^2$  sigue una  $\chi^2_{n-r}$  tenemos

$$P(a < SCR/\sigma^2 < b) = 1 - \alpha$$

donde  $a$  y  $b$  son tales que

$$P(\chi^2_{n-r} \leq a) = \alpha/2 \quad P(\chi^2_{n-r} > b) = \alpha/2$$

Deducimos entonces que

$$P\left(\frac{SCR}{b} < \sigma^2 < \frac{SCR}{a}\right) = 1 - \alpha \quad (3.6)$$

define un intervalo de confianza para la varianza  $\sigma^2$  del modelo lineal normal, con coeficiente de confianza  $1 - \alpha$ .

### 3.6. Ejercicios

#### Ejercicio 3.1

Sea  $\psi$  una función paramétrica estimable y  $\widehat{\psi}_1, \widehat{\psi}_2$  dos estimadores insesgados, estocásticamente independientes, de varianzas  $\sigma_1^2$  y  $\sigma_2^2$ . Hallar la combinación lineal de  $\widehat{\psi}_1, \widehat{\psi}_2$  cuya varianza es mínima y además es insesgado.

#### Ejercicio 3.2

En un modelo lineal, la matriz de diseño es

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Hallar la expresión general de las funciones paramétricas estimables.

#### Ejercicio 3.3

Probar que

$$\widehat{\psi} = \mathbf{b}'\mathbf{Y} \quad E(\widehat{\psi}) = \psi = \mathbf{a}'\boldsymbol{\beta}$$

siendo  $\mathbf{b}$  combinación lineal de las columnas de  $\mathbf{X}$ , implica que  $\mathbf{a}$  es combinación lineal de las filas de  $\mathbf{X}$ .

#### Ejercicio 3.4

Probar que toda combinación lineal de funciones paramétricas estimables es también función paramétrica estimable y que  $r = \text{rg } \mathbf{X}$  es el número máximo de funciones linealmente independientes.

#### Ejercicio 3.5

Si  $\widehat{\psi}$  es la estimación de Gauss-Markov, probar que la expresión

$$\widehat{\psi} = c_1 \bar{y}_1 + \cdots + c_k \bar{y}_k$$

función de las medias de las condiciones experimentales, es única.

**Ejercicio 3.6**

La matriz de diseño reducida correspondiente a un modelo lineal normal es

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

Se sabe además que

$$\begin{aligned} \bar{y}_1 &= 11 & \bar{y}_2 &= 10 & \bar{y}_3 &= 15 \\ n_1 &= n_2 = n_3 = 10 \\ s_1^2 &= (1/n_1) \sum_{i=1}^{n_1} (y_i - \bar{y}_1)^2 = 4.5 \\ s_2^2 &= 6.0 & s_3^2 &= 4.3 \end{aligned}$$

Se pide

- 1) Hallar la expresión general de las estimaciones MC de  $\beta$ .
- 2) Calcular SCR. ¿Se ajustan los datos al modelo definido por  $\mathbf{X}$ ? (nivel de significación 0.05)
- 3) Dada la función paramétrica estimable

$$\psi = \beta_1 + \beta_3$$

contrastar la hipótesis  $H_0 : \psi = 3$  en los casos:

- a)  $\sigma^2$  varianza del diseño desconocida
- b)  $\sigma^2 = 5$  varianza del diseño conocida

(nivel de significación 0.05)

- 4) Hallar la función paramétrica estimable  $\psi$  tal que

$$\widehat{\psi} = c_1 \bar{y}_1 + c_2 \bar{y}_2 + c_3 \bar{y}_3$$

verifica  $c_1^2 + c_2^2 + c_3^2 = 1$  y además  $\widehat{\psi}$  es máximo.

**Ejercicio 3.7**

Consideremos el modelo lineal

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 + \epsilon_1 \\ y_2 &= \beta_1 + \beta_3 + \epsilon_2 \\ y_3 &= \beta_1 + \beta_2 + \epsilon_3 \end{aligned}$$

Se pide:

- 1) ¿Es la función paramétrica

$$\psi = \beta_1 + \beta_2 + \beta_3$$

estimable?

- 2) Probar que toda función paramétrica

$$\psi = a_1 \beta_1 + a_2 \beta_2 + a_3 \beta_3$$

es estimable si y sólo si  $a_1 = a_2 + a_3$ .

**Ejercicio 3.8**

Consideremos el modelo lineal

$$y_1 = \mu + \alpha_1 + \beta_1 + \epsilon_1$$

$$y_2 = \mu + \alpha_1 + \beta_2 + \epsilon_2$$

$$y_3 = \mu + \alpha_2 + \beta_1 + \epsilon_3$$

$$y_4 = \mu + \alpha_2 + \beta_2 + \epsilon_4$$

$$y_5 = \mu + \alpha_3 + \beta_1 + \epsilon_5$$

$$y_6 = \mu + \alpha_3 + \beta_2 + \epsilon_6$$

- (a) ¿Cuando es  $\lambda_0\mu + \lambda_1\alpha_1 + \lambda_2\alpha_2 + \lambda_3\alpha_3 + \lambda_4\beta_1 + \lambda_5\beta_2$  estimable?
- (b) ¿Es  $\alpha_1 + \alpha_2$  estimable?
- (c) ¿Es  $\beta_1 - \beta_2$  estimable?
- (d) ¿Es  $\mu + \alpha_1$  estimable?
- (e) ¿Es  $6\mu + 2\alpha_1 + 2\alpha_2 + 2\alpha_3 + 3\beta_1 + 3\beta_2$  estimable?
- (f) ¿Es  $\alpha_1 - 2\alpha_2 + \alpha_3$  estimable?
- (g) Hallar la covarianza entre los estimadores lineales MC de las funciones paramétricas  $\beta_1 - \beta_2$  y  $\alpha_1 - \alpha_2$ , si éstas son estimables.
- (h) Hallar la dimensión del espacio paramétrico.
- (i) Obtener una expresión del espacio de los errores.

**Ejercicio 3.9**

Cuatro objetos  $A, B, C, D$  están involucrados en un experimento de pesado. Todos reunidos pesan  $y_1$  gramos. Cuando  $A$  y  $C$  se ponen en el plato izquierdo de la balanza y  $B$  y  $D$  se ponen en el plato derecho, un peso de  $y_2$  gramos es necesario en el plato derecho para equilibrar la balanza. Con  $A$  y  $B$  en el plato izquierdo y  $C, D$  en el plato derecho,  $y_3$  gramos son necesarios en el plato derecho y, finalmente, con  $A, D$  en el plato izquierdo y  $B, C$  en el plato derecho,  $y_4$  gramos son necesarios en la derecha para equilibrar. Si las observaciones  $y_1, y_2, y_3, y_4$  son todas con errores incorrelacionados y con varianza común  $\sigma^2$ , obtener la estimación BLUE del peso total de los cuatro objetos y su varianza.

**Ejercicio 3.10**

Un transportista realiza diversos trayectos entre tres poblaciones  $A, B$  y  $C$ . En cuatro días consecutivos ha hecho los recorridos que muestra la siguiente tabla:

trayecto	km
$A \rightarrow B \rightarrow A \rightarrow C$	533
$C \rightarrow A \rightarrow C \rightarrow B$	583
$B \rightarrow C \rightarrow A \rightarrow C \rightarrow A \rightarrow B \rightarrow A$	1111
$A \rightarrow B \rightarrow A \rightarrow C \rightarrow A \rightarrow B \rightarrow A$	1069

donde el kilometraje es, por diversas causas, aproximado.

- (a) Proponer un modelo lineal, con la matriz de diseño y las hipótesis necesarias, para estimar las distancias kilométricas entre las tres poblaciones.

Con los datos proporcionados, ¿es posible estimar las distancias entre las tres poblaciones? ¿Cuales son las distancias o funciones paramétricas estimables (fpe) en este modelo?



- (b) ¿Se puede estimar el kilometraje del trayecto  $M_{BC} \rightarrow B \rightarrow A \rightarrow C \rightarrow M_{AC}$ , donde  $M_{IJ}$  es el punto medio entre dos poblaciones? ¿Es una buena estimación? ¿Cual es el *error* de esta estimación?

### Ejercicio 3.11

Con el modelo lineal

$$y_1 = \theta_1 + \theta_5 + \epsilon_1$$

$$y_2 = \theta_2 + \theta_5 + \epsilon_2$$

$$y_3 = \theta_3 + \theta_6 + \epsilon_3$$

$$y_4 = \theta_4 + \theta_6 + \epsilon_4$$

$$y_5 = \theta_1 + \theta_7 + \epsilon_5$$

$$y_6 = \theta_3 + \theta_7 + \epsilon_6$$

$$y_7 = \theta_2 + \theta_8 + \epsilon_7$$

$$y_8 = \theta_4 + \theta_8 + \epsilon_8$$

contestar las siguientes preguntas:

- (a) ¿Cuántas funciones paramétricas son estimables? Obtener el conjunto completo de todas ellas.
- (b) Probar que  $\theta_1 - \theta_2$  es estimable. Calcular su estimador lineal MC y su varianza.
- (c) Probar que  $\theta_1 + \theta_2$  no es estimable.
- (d) Hallar cuatro estimadores insesgados diferentes de  $\theta_1 - \theta_2$  y calcular sus varianzas. Compararlas con la varianza del estimador MC.
- (e) Hallar un estimador insesgado de la varianza de los errores  $\sigma^2$ .

### Ejercicio 3.12

Diremos que el estimador lineal  $\mathbf{b}'\mathbf{Y}$  pertenece al espacio error si  $E(\mathbf{b}'\mathbf{Y}) = \mathbf{0}$ . Probar que la covarianza entre  $\mathbf{b}'\mathbf{Y}$  y todo estimador de Gauss-Markov  $\widehat{\psi} = \mathbf{a}'\boldsymbol{\beta}$  es siempre cero.

### Ejercicio 3.13

Consideremos el modelo lineal normal  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , siendo  $\text{rg } \mathbf{X} = r$ . Sea  $\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}'$  una descomposición en valores singulares de  $\mathbf{X}$ . Se pide:

- 1) Expresar la estimación MC de  $\boldsymbol{\beta}$  en términos de  $\mathbf{U}$ ,  $\boldsymbol{\Delta}$ ,  $\mathbf{V}$  y  $\mathbf{Y}$ .
- 2) Sea  $\psi = \mathbf{a}'\boldsymbol{\beta}$  una función paramétrica. Probar que  $\psi$  es estimable si y sólo si se verifica

$$\mathbf{a}' = \mathbf{b}'\mathbf{V}'$$

para algún vector  $\mathbf{b}$ .