

Regresión por etapas

Un procedimiento estándar para buscar el “subconjunto óptimo” de variables ante la ausencia de ortogonalidad es una técnica denominada **regresión por etapas**, que se basa en el procedimiento de introducir en forma secuencial las variables al modelo, una por una. Dado un tamaño α predeterminado, la descripción de la rutina por etapas se entenderá mejor si primero se describen los métodos de **selección hacia delante** y **eliminación hacia atrás**.

La **selección hacia delante** se basa en el concepto de que las variables deben insertarse una por una hasta obtener una ecuación de regresión satisfactoria. El procedimiento es como sigue:

PASO 1. Elija la variable que proporcione la mayor suma de cuadrados de regresión cuando se ejecute la regresión lineal simple con y o, en forma equivalente, aquella que proporcione el mayor valor de R^2 . Esta variable inicial se llamará x_1 . Si x_1 es insignificante, el procedimiento se suspende.

PASO 2. Seleccione la variable que al ser integrada al modelo proporciona el mayor incremento de R^2 , en presencia de x_1 , sobre la R^2 encontrada en el paso 1. Ésta, por supuesto, es la variable x_j para la que

$$R(\beta_j | \beta_1) = R(\beta_1, \beta_j) - R(\beta_1)$$

es más grande. Dicha variable se llamará x_2 . Luego se ajusta el modelo de regresión con x_1 y x_2 , y se observa R^2 . Si x_2 es insignificante, el procedimiento se suspende.

PASO 3. Elija la variable x_j que proporciona el valor más grande de

$$R(\beta_j | \beta_1, \beta_2) = R(\beta_1, \beta_2, \beta_j) - R(\beta_1, \beta_2),$$

otra vez da como resultado el incremento mayor de R^2 sobre el que se obtuvo en el paso 2. A esta variable se le denomina x_3 , y ahora se tiene un modelo de regresión que incluye x_1 , x_2 y x_3 . Si x_3 es insignificante, el procedimiento se suspende.

Este proceso continúa hasta que la variable más reciente incluida ya no produce un incremento significativo en la regresión explicada. Tal incremento se puede determinar en cada paso utilizando adecuadamente una prueba F o una prueba t parciales. Por ejemplo, en el paso 2 el valor

$$f = \frac{R(\beta_2 | \beta_1)}{s^2}$$

se determina para probar la pertinencia de x_2 en el modelo. Aquí, el valor de s^2 es el cuadrado medio del error para el modelo que contiene las variables x_1 y x_2 . De manera similar, en el paso 3 la razón

$$f = \frac{R(\beta_3 | \beta_1, \beta_2)}{s^2}$$

prueba la pertinencia de x_3 en el modelo. Sin embargo, ahora el valor de s^2 es el cuadrado medio del error para el modelo que contiene las tres variables x_1 , x_2 y x_3 . Si en el paso 2, $f < f_\alpha(1, n - 3)$ para un nivel de significancia preseleccionado, x_2 no está incluida y el

proceso finaliza, lo que da como resultado una ecuación lineal simple que relaciona y y x_1 . Sin embargo, si $f > f_\alpha(1, n - 3)$, se avanza al paso 3. De nuevo, si en el paso 3, $f < f_\alpha(1, n - 4)$, entonces x_3 no se incluye y el proceso termina con la ecuación de la regresión apropiada que contiene las variables x_1 y x_2 .

La **eliminación hacia atrás** implica los mismos conceptos que la selección hacia delante, excepto que se comienza con todas las variables en el modelo. Por ejemplo, suponga que hay cinco variables en consideración. Los pasos son:

PASO 1. Ajuste una ecuación de regresión con las cinco variables incluidas en el modelo. Elija la variable que proporcione el valor más pequeño de la suma de cuadrados de regresión **ajustada para las demás**. Suponga que dicha variable es x_2 . Elimine x_2 del modelo si

$$f = \frac{R(\beta_2 \mid \beta_1, \beta_3, \beta_4, \beta_5)}{s^2}$$

es insignificante.

PASO 2. Ajuste una ecuación de regresión utilizando las variables restantes x_1, x_3, x_4 y x_5 , y repita el paso 1. Suponga que esta vez elige la variable x_5 . Nuevamente, si

$$f = \frac{R(\beta_5 \mid \beta_1, \beta_3, \beta_4)}{s^2}$$

es insignificante, se retira del modelo la variable x_5 . En cada paso la s^2 que se usa en la prueba F es el cuadrado medio del error para el modelo de regresión en esa etapa.

Este proceso se repite hasta que en algún paso la variable con la suma de cuadrados de regresión ajustada más pequeña produce un valor f significativo a un nivel de significancia predeterminado.

La **regresión por etapas** se lleva a cabo con una modificación ligera pero importante del procedimiento de selección hacia delante. La modificación requiere efectuar más pruebas en cada etapa para garantizar la eficacia continuada de las variables que se hubieran incluido en el modelo durante alguna etapa anterior. Esto representa una mejoría sobre la selección hacia delante, ya que es muy posible que una variable que haya entrado a la ecuación de regresión en una etapa temprana resulte poco importante o redundante debido a las relaciones que existen entre ella y las otras variables que se incluyeron en etapas posteriores. Por lo tanto, en la etapa en que se incluyó una variable nueva a la ecuación de regresión mediante un incremento significativo de R^2 , según lo determina la prueba F , todas las variables que ya estén en el modelo se someten a pruebas F (o bien, a pruebas t) a la luz de esta nueva variable, y si no muestran un valor f significativo, se eliminan. El procedimiento continúa hasta que se alcance una etapa donde ya no sea posible insertar ni eliminar variables adicionales. Este procedimiento por etapas se ilustra con el siguiente ejemplo.

Ejemplo 12.11: Utilice las técnicas de regresión por etapas y calcule un modelo de regresión lineal adecuado para predecir la talla de los bebés cuyos datos se presentan en la tabla 12.8.

Solución: **PASO 1.** Se considera cada variable por separado y se ajustan cuatro ecuaciones individuales de regresión lineal simple. Se calculan las siguientes sumas de cua-

drados de regresión pertinentes:

$$\begin{aligned} R(\beta_1) &= 288.1468, & R(\beta_2) &= 215.3013, \\ R(\beta_3) &= 186.1065, & R(\beta_4) &= 100.8594. \end{aligned}$$

Es evidente que la variable x_1 proporciona la suma de cuadrados de regresión más elevada. El cuadrado medio del error para la ecuación que implica sólo x_1 es $s^2 = 4.7276$, y como

$$f = \frac{R(\beta_1)}{s^2} = \frac{288.1468}{4.7276} = 60.9500,$$

que excede a $f_{0.05}(1, 7) = 5.59$, la variable x_1 es significativa y se introduce al modelo.

PASO 2. En esta etapa se ajustan tres ecuaciones de regresión y todas incluyen a x_1 . Los resultados importantes para las combinaciones (x_1, x_2) , (x_1, x_3) y (x_1, x_4) son

$$R(\beta_2|\beta_1) = 23.8703, \quad R(\beta_3|\beta_1) = 29.3086, \quad R(\beta_4|\beta_1) = 13.8178.$$

La variable x_3 muestra la mayor suma de cuadrados de regresión en presencia de x_1 . La regresión que implica x_1 y x_3 proporciona un valor nuevo de $s^2 = 0.6307$, y como

$$f = \frac{R(\beta_3|\beta_1)}{s^2} = \frac{29.3086}{0.6307} = 46.47,$$

que excede a $f_{0.05}(1, 6) = 5.99$, la variable x_3 es significativa y se incluye en el modelo junto con x_1 . Ahora debemos someter a x_1 a una prueba de significancia en presencia de x_3 . Encontramos que $R(\beta_1|\beta_3) = 131.349$, en consecuencia,

$$f = \frac{R(\beta_1|\beta_3)}{s^2} = \frac{131.349}{0.6307} = 208.26,$$

que es muy significativa. Por lo tanto, se mantiene x_1 junto con x_3 .

PASO 3. Con x_1 y x_3 incluidas en el modelo, ahora se requiere $R(\beta_2|\beta_1, \beta_3)$ y $R(\beta_4|\beta_1, \beta_3)$ para determinar cuál de las dos variables restantes, si es que acaso se puede incluir alguna, se debe incluir en esta etapa. Del análisis de regresión, usando x_2 junto con x_1 y x_3 , se observa que $R(\beta_2|\beta_1, \beta_3) = 0.7948$, y cuando x_4 se utiliza con x_1 y x_3 , se obtiene $R(\beta_4|\beta_1, \beta_3) = 0.1855$. El valor de s^2 es 0.5979 para la combinación (x_1, x_2, x_3) , y 0.7198 para la combinación (x_1, x_2, x_4) . Como ningún valor f es significativo al nivel $\alpha = 0.05$, el modelo final de regresión sólo incluye las variables x_1 y x_3 . Se encuentra que la ecuación de estimación es

$$\hat{y} = 20.1084 + 0.4136x_1 + 2.0253x_3,$$

y el coeficiente de determinación para este modelo es $R^2 = 0.9882$.

Aunque (x_1, x_3) es la combinación elegida mediante la regresión por etapas, no es necesariamente la combinación de dos variables que proporciona el valor más grande de R^2 . De hecho, ya observamos que la combinación (x_2, x_3) da un valor de $R^2 = 0.9905$. Desde luego, el procedimiento por etapas nunca tomó en cuenta dicha combinación. Se podría plantear un argumento racional de que en realidad hay una diferencia despreciable en el desempeño entre esas dos ecuaciones de estimación, al menos en términos del

porcentaje de variación explicado. Sin embargo, es interesante observar que el procedimiento de eliminación hacia atrás proporciona la combinación (x_2, x_3) en la ecuación final (véase el ejercicio 12.49 en la página 494). ■

Resumen

La función principal de cada uno de los procedimientos explicados en esta sección consiste en exponer las variables a una metodología sistemática, diseñada para garantizar la inclusión final de las mejores combinaciones de las mismas. Es evidente que no es seguro que esto pase en todos los problemas y, por supuesto, es posible que la multicolinealidad sea tan extensa que no haya más alternativa que apoyarse en procedimientos de estimación diferentes de los mínimos cuadrados. Tales procedimientos de estimación se estudian en Myers (1990), listado en la bibliografía.

Los procedimientos secuenciales que se estudian aquí son tres de los muchos métodos de ese tipo que aparecen en la literatura y que están incluidos en diversos paquetes de regresión por computadora. Estos métodos fueron diseñados para ser eficientes en cuanto al cálculo pero, por supuesto, no proporcionan resultados para todos los subconjuntos posibles de variables. Debido a esto los procedimientos son más eficaces para conjuntos de datos que incluyen un **número grande de variables**. En el caso de los problemas de regresión que implican un número relativamente pequeño de variables, los paquetes modernos de cómputo para la regresión permiten el cálculo y resumen la información cuantitativa de todos los modelos para cada subconjunto posible de variables. En la sección 12.11 se proporcionan ilustraciones.

Elección de valores P

Como es de esperarse, la elección del modelo final con estos procedimientos podría depender en gran medida del valor P que se seleccione. Además, un procedimiento es más exitoso cuando es forzado a probar una gran cantidad de variables posibles. Por esta razón, cualquier procedimiento hacia delante es más útil cuando se utiliza un valor P relativamente grande. A esto se debe que algunos programas de cómputo empleen un valor P predeterminado de 0.50.

12.10 Estudio de los residuales y violación de las suposiciones (verificación del modelo)

Anteriormente en este capítulo se sugirió que los residuales, o errores en el ajuste de regresión, con frecuencia proporcionan información que puede ser muy valiosa para el analista de datos. Los $e_i = y_i - \hat{y}_i$, $i = 1, 2, \dots, n$, que son el equivalente numérico de los ϵ_i , los errores del modelo, a menudo revelan la posible violación de las suposiciones o la presencia de datos de puntos “sospechosos”. Suponga que el vector \mathbf{x}_i denota los valores de las variables regresoras que corresponden al i -ésimo punto de los datos, complementado por un 1 en la posición inicial. Es decir,

$$\mathbf{x}_i' = [1, x_{1i}, x_{2i}, \dots, x_{ki}].$$

Considere la cantidad

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i, \quad i = 1, 2, \dots, n.$$

El lector debería notar que en la sección 12.5 se utilizó h_{ii} para calcular los intervalos de confianza de la respuesta media. Además de σ^2 , h_{ii} representa la varianza del valor ajustado \hat{y}_i . Los valores h_{ii} son los elementos de la diagonal de la **matriz “SOMBRERO”**

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

que desempeña un papel importante en cualquier estudio de residuales y en otros aspectos modernos del análisis de regresión (véase Myers, 1990, citado en la bibliografía). El término *matriz SOMBRERO* se deriva del hecho de que \mathbf{H} genera las “y sombrero”, o los valores ajustados cuando se multiplica por el vector \mathbf{y} de respuestas observadas. Es decir, $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$, por lo tanto,

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

donde $\hat{\mathbf{y}}$ es el vector cuyo i -ésimo elemento es \hat{y}_i .

Si se hacen las suposiciones acostumbradas de que los ϵ_i son independientes y están distribuidos normalmente, con media cero y varianza σ^2 , las propiedades estadísticas de los residuales se establecen con facilidad. Entonces,

$$E(e_i) = E(y_i - \hat{y}_i) = 0 \quad \text{y} \quad \sigma_{\epsilon_i}^2 = (1 - h_{ii})\sigma^2,$$

para $i = 1, 2, \dots, n$. (Para mayores detalles véase Myers, 1990). Es posible demostrar que los valores de la diagonal de la matriz SOMBRERO están acotados de acuerdo con la desigualdad

$$\frac{1}{n} \leq h_{ii} \leq 1.$$

Además, $\sum_{i=1}^n h_{ii} = k + 1$, el número de parámetros de la regresión. Como resultado, cualquier punto de los datos cuyo elemento diagonal SOMBRERO sea grande, es decir, esté muy por encima del valor promedio de $(k + 1)/n$, está en una posición dentro del conjunto de datos donde la varianza de \hat{y}_i es relativamente grande y la varianza de un residuo es relativamente pequeña. Como resultado, el analista de datos puede tener una idea de qué tan grande puede ser un residuo antes de que su desviación de cero se pueda atribuir a algo distinto del azar. Muchos de los paquetes comerciales para computadora que permiten calcular la regresión producen el conjunto de **residuales estudentizados**.

Residuo
estudentizado

$$r_i = \frac{e_i}{s\sqrt{1 - h_{ii}}}, \quad i = 1, 2, \dots, n$$

Aquí, cada residuo se **dividió entre una estimación de su desviación estándar** creando un estadístico *tipo t* diseñado para dar al analista una cantidad sin escala que proporcione información sobre el **tamaño** del residual. Además, a menudo los paquetes de cómputo comunes proporcionan valores de otro conjunto de residuales tipo estudentizados denominados **valores R de Student**.

Residual R de
Student

$$t_i = \frac{e_i}{s_{-i}\sqrt{1 - h_{ii}}}, \quad i = 1, 2, \dots, n,$$

donde s_{-i} es un estimador de la desviación estándar del error calculado con el ***i*-ésimo punto de los datos eliminado**.

Hay tres tipos de transgresiones de las suposiciones fáciles de detectar mediante el uso de los residuales o *gráficas de residuales*. Aunque las gráficas de los residuales brutos, los e_i , ayudan a esto, con frecuencia es más informativo graficar los residuales estudentizados. Las tres transgresiones son las siguientes:

1. Presencia de valores extremos
2. Varianza heterogénea del error
3. Especificación inadecuada del modelo

En el caso 1 elegimos definir un **valor extremo** como un punto de los datos que se desvía de la suposición común de que $E(\epsilon_i) = 0$ para un valor específico de i . Si hay una razón para creer que un punto de un dato específico es un valor extremo que ejerce una gran influencia sobre el modelo ajustado, r_i o t_i , esto podría estar informando algo. Es de esperarse que los valores R de Student sean más sensibles a los valores extremos que los valores r_i .

En realidad, en el caso de que $E(\epsilon_i) = 0$, t_i es un valor de una variable aleatoria que sigue una distribución t con $n - 1 - (k + 1) = n - k - 2$ grados de libertad. Por consiguiente, es posible utilizar una prueba t de dos colas para proporcionar información con el fin de detectar si el punto i -ésimo es o no un valor extremo.

Aunque el estadístico R de Student t_i produce una prueba t exacta para detectar un valor extremo en una ubicación específica, la distribución t no se aplicaría para probar simultáneamente varios valores extremos en todas las ubicaciones. Como resultado, los residuales estudentizados o valores R de Student se deberían usar estrictamente como herramientas de diagnóstico *sin* un mecanismo de prueba de hipótesis formal. La implicación es que dichos estadísticos resaltan puntos de los datos en los que el error del ajuste es mayor de lo esperado por el azar. Los valores R de Student de gran magnitud sugieren la necesidad de “verificar” los datos con todos los recursos disponibles. La práctica de eliminar observaciones de conjuntos de datos de la regresión no debería llevarse a cabo de forma indiscriminada. (Para más información sobre el uso de los diagnósticos sobre valores extremos véase Myers, 1990, en la bibliografía).

Ilustración de la detección de valores extremos

Estudio de caso 12.1: Método para capturar saltamontes. En un experimento biológico, que fue efectuado en el Departamento de Entomología de Virginia Tech, se hicieron n ensayos experimentales con dos métodos diferentes para capturar saltamontes. Los métodos consistieron en la captura por caída de la red y la captura por barrido de la red. El número promedio de saltamontes atrapados con cada método se registró en un conjunto de cuadrantes del campo en una fecha determinada. También se registró una variable regresora adicional, la altura promedio de las plantas en los cuadrantes. Los datos experimentales aparecen en la tabla 12.10.

El objetivo consiste en estimar cuántos saltamontes se capturan empleando sólo el método del barrido de la red, que es menos costoso. Hay cierta preocupación por la validez del cuarto punto de los datos. La captura observada utilizando el método de caída de la red que se reportó parece inusualmente alta, dadas las demás condiciones, de hecho se pensó que la cifra podía ser errónea. Ajuste un modelo del tipo

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

para los 17 puntos de los datos y estudie los residuales para determinar si el punto 4 es un valor extremo.

Tabla 12.10: Conjunto de datos para el estudio de caso 12.1

Observación	Captura por caída de la red, y	Captura por barrido de la red, x_1	Altura de las plantas, x_2 (cm)
1	18.0000	4.15476	52.705
2	8.8750	2.02381	42.069
3	2.0000	0.15909	34.766
4	20.0000	2.32812	27.622
5	2.3750	0.25521	45.879
6	2.7500	0.57292	97.472
7	3.3333	0.70139	102.062
8	1.0000	0.13542	97.790
9	1.3333	0.12121	88.265
10	1.7500	0.10937	58.737
11	4.1250	0.56250	42.386
12	12.8750	2.45312	31.274
13	5.3750	0.45312	31.750
14	28.0000	6.68750	35.401
15	4.7500	0.86979	64.516
16	1.7500	0.14583	25.241
17	0.1333	0.01562	36.354

Solución: Un paquete de cómputo generó el modelo de regresión ajustado

$$\hat{y} = 3.6870 + 4.1050x_1 - 0.0367x_2$$

junto con los estadísticos $R^2 = 0.9244$ y $s^2 = 5.580$. También se obtuvieron los residuales y otra información de diagnóstico que fueron registrados en la tabla 12.11.

Como se esperaba, el residual en la cuarta ubicación parece inusualmente grande, a saber, 7.769. La cuestión fundamental aquí es si este residual es más grande de lo que se esperaría debido al azar. El error estándar del residual para el punto 4 es 2.209. El valor R de Student t_4 que se obtuvo es 9.9315. Al considerarlo como el valor de una variable aleatoria que tiene una distribución t con 13 grados de libertad, se concluiría sin duda que el residuo de la cuarta observación se estima algo mayor que 0, y que la medición del presunto error es apoyada por el estudio de los residuales. Observe que ningún otro valor de los residuales proporciona un valor R de Student que sea motivo de alarma. ▀

Gráfica de los residuales para el estudio de caso 12.1

En el capítulo 11 estudiamos con cierto detalle la utilidad de graficar los residuos en el análisis de regresión. Es frecuente que con base en dichas gráficas se detecte la violación de las suposiciones del modelo. En la regresión múltiple en ocasiones es útil graficar la probabilidad normal de los residuales o los residuales en comparación con \hat{y} . Sin embargo, a menudo es preferible graficar los residuales estudentizados.

Recuerde que la preferencia por los residuales estudentizados sobre los residuales ordinarios para propósitos de graficación se debe a que, como la varianza de i -ésimo

Tabla 12.11: Información sobre los residuales para el conjunto de datos del estudio de caso 12.1

Obs.	y_i	\hat{y}_i	$y_i - \hat{y}_i$	h_{ii}	$s \sqrt{1 - h_{ii}}$	r_i	t_i
1	18.000	18.809	-0.809	0.2291	2.074	-0.390	-0.3780
2	8.875	10.452	-1.577	0.0766	2.270	-0.695	-0.6812
3	2.000	3.065	-1.065	0.1364	2.195	-0.485	-0.4715
4	20.000	12.231	7.769	0.1256	2.209	3.517	9.9315
5	2.375	3.052	-0.677	0.0931	2.250	-0.301	-0.2909
6	2.750	2.464	0.286	0.2276	2.076	0.138	0.1329
7	3.333	2.823	0.510	0.2669	2.023	0.252	0.2437
8	1.000	0.656	0.344	0.2318	2.071	0.166	0.1601
9	1.333	0.947	0.386	0.1691	2.153	0.179	0.1729
10	1.750	1.982	-0.232	0.0852	2.260	-0.103	-0.0989
11	4.125	4.442	-0.317	0.0884	2.255	-0.140	-0.1353
12	12.875	12.610	0.265	0.1152	2.222	0.119	0.1149
13	5.375	4.383	0.992	0.1339	2.199	0.451	0.4382
14	28.000	29.841	-1.841	0.6233	1.450	-1.270	-1.3005
15	4.750	4.891	-0.141	0.0699	2.278	-0.062	-0.0598
16	1.750	3.360	-1.610	0.1891	2.127	-0.757	-0.7447
17	0.133	2.418	-2.285	0.1386	2.193	-1.042	-1.0454

residuo depende del i -ésimo elemento en la diagonal SOMBRERO, las varianzas de los residuos diferirán si hay dispersión en las diagonales SOMBRERO. Así, es probable que la apariencia de una gráfica de residuales sugiera heterogeneidad debido a que los propios residuales no se comportan, en general, de manera ideal. El propósito de utilizar residuales estudentizados es proporcionar un tipo de *estandarización*. Es evidente que si se conociera σ , en condiciones ideales, es decir, en las que el modelo fuera correcto y la varianza homogénea, se tendría

$$E\left(\frac{e_i}{\sigma\sqrt{1-h_{ii}}}\right) = 0 \quad \text{y} \quad \text{Var}\left(\frac{e_i}{\sigma\sqrt{1-h_{ii}}}\right) = 1.$$

De manera que los residuales estudentizados producen un conjunto de estadísticos que en condiciones ideales se comportan en forma estándar. La figura 12.5 presenta una gráfica con los valores **R de Student** para los datos de los saltamontes del estudio de caso 12.1. Advierta que el valor para la observación 4 se destaca de los demás. La gráfica *R* de Student se generó con el programa SAS. La gráfica presenta los residuales en comparación con los valores \hat{y} .

Verificación de la normalidad

El lector debe recordar, de acuerdo con lo que se estudió en el capítulo 11, la importancia de verificar la normalidad utilizando una gráfica de probabilidad normal. La misma recomendación es válida para el caso de la regresión lineal múltiple. Las gráficas de probabilidad normal se pueden generar utilizando software estándar para regresión. Sin embargo, como ya se indicó, éstas pueden ser más eficaces si se usan residuales estudentizados o valores *R* de Student en vez de residuales comunes.

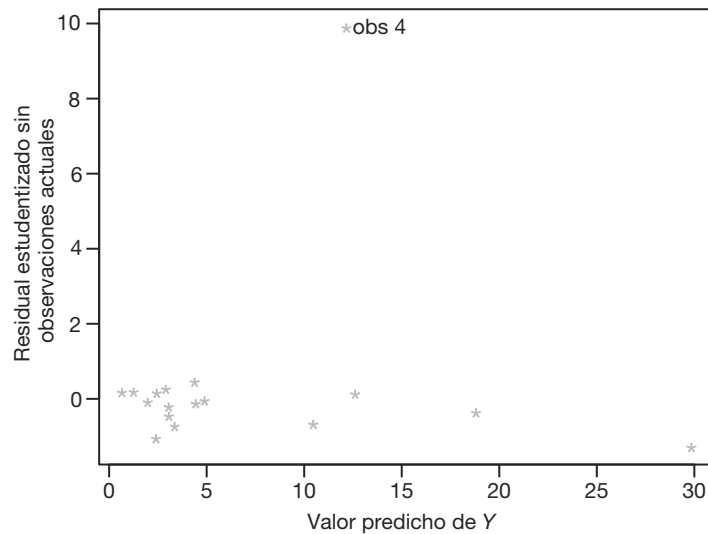


Figura 12.5: Valores R de Student graficados en comparación con los valores predichos para los datos de los saltamontes del estudio de caso 12.1.

12.11 Validación cruzada, C_p , y otros criterios para la selección del modelo

Para muchos problemas de regresión el experimentador debe elegir entre varios modelos alternativos o formas de modelo que se desarrollan a partir del mismo conjunto de datos. Con mucha frecuencia se requiere el modelo que predice o estima mejor la respuesta media. El experimentador debe tomar en cuenta los tamaños relativos de los valores de s^2 para los posibles modelos y, sin duda, la naturaleza general de los intervalos de confianza sobre la respuesta media. También se debe considerar lo bien que el modelo predice los valores de la respuesta que **no se hayan utilizado para construir los posibles modelos**. Los modelos deben estar sujetos a **validación cruzada**. Entonces, lo que se requiere son los errores de la validación cruzada en lugar de los errores del ajuste. Estos errores en la predicción son los **residuales PRESS**.

$$\delta_i = y_i - \hat{y}_{i, -i}, \quad i = 1, 2, \dots, n,$$

donde $\hat{y}_{i, -i}$ es la predicción del i -ésimo punto de los datos por medio de un modelo que no utiliza el i -ésimo punto en el cálculo de los coeficientes. Estos residuales PRESS se calculan mediante la fórmula

$$\delta_i = \frac{e_i}{1 - h_{ii}}, \quad i = 1, 2, \dots, n.$$

(La derivación se encuentra en Myers, 1990).

Uso del estadístico PRESS

La motivación para utilizar PRESS y la utilidad de los residuales PRESS es muy fácil de entender. El propósito de extraer o *separar* puntos de datos, uno a la vez, consiste en

permitir el empleo de metodologías separadas para ajustar y evaluar un modelo específico. Para evaluar un modelo la “-i” indica que el residual PRESS proporciona un error de predicción donde la observación que se predice es *independiente del ajuste del modelo*.

Los criterios que utilizan los residuales PRESS son dados por

$$\sum_{i=1}^n |\delta_i| \quad \text{y} \quad \text{PRESS} = \sum_{i=1}^n \delta_i^2.$$

(El término **PRESS** es un acrónimo que se forma con las iniciales de los términos de la frase en inglés *prediction sum of squares*, que se traduce como **suma de cuadrados de predicción**). Se sugiere que se utilicen ambos criterios. Es posible que PRESS sea dominado por uno o algunos residuales PRESS grandes. Es evidente que el criterio sobre $\sum_{i=1}^n |\delta_i|$ es menos sensible a un número pequeño de valores grandes.

Además del estadístico PRESS en sí, el analista puede simplemente calcular un estadístico similar a R^2 que refleje el desempeño de la predicción. Con frecuencia a este estadístico se le denomina R^2_{pred} y se calcula como sigue:

R^2 de predicción Dado un modelo ajustado con valor específico para PRESS, R^2_{pred} es dado por

$$R^2_{\text{pred}} = 1 - \frac{\text{PRESS}}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Observe que R^2_{pred} es tan sólo el estadístico común R^2 donde la *SCE* fue reemplazada por el estadístico PRESS.

En el siguiente estudio de caso se proporciona un ejemplo en el que se ajustan muchos posibles modelos a un conjunto de datos y se elige el mejor de ellos. No se emplean los procedimientos secuenciales descritos en la sección 12.9. En vez de eso se ilustra el papel que desempeñan los residuales PRESS y otros valores estadísticos cuando se trata de seleccionar la mejor ecuación de regresión.

Estudio de caso 12.2: Patada de fútbol. La fuerza de las piernas es un requisito necesario para que un pateador tenga éxito en el fútbol americano. Una medida de la calidad de una buena patada es el “tiempo de vuelo” del balón, es decir, el tiempo que el balón se mantiene en el aire antes de ser atrapado por el regresador de patadas. Para determinar cuáles factores de la fuerza de las piernas influyen en el tiempo de vuelo y desarrollar un modelo empírico para predecir esta respuesta, el Departamento de Salud, Educación Física y Recreación de Virginia Tech llevó a cabo un estudio sobre *La relación entre variables seleccionadas de desempeño físico y la capacidad de despejes en el fútbol*. Se eligieron 13 pateadores para el experimento y cada uno pateó 10 veces el balón. En la tabla 12.12 aparece el registro del tiempo de vuelo promedio, junto con las medidas de fuerza usadas en el análisis.

Cada variable regresora se define como sigue:

1. **FPD**, fuerza de la pierna derecha (libras)
2. **FPI**, fuerza de la pierna izquierda (libras)
3. **FTD**, flexibilidad muscular del tendón derecho (grados)
4. **FTI**, flexibilidad muscular del tendón izquierdo (grados)

5. Potencia, fuerza general de las piernas (pie-libras)

Determine el modelo más adecuado para predecir el tiempo de vuelo.

Tabla 12.12: Datos para el estudio de caso 12.2.

Pateador	Tiempo de vuelo, y (seg)	FPD, x_1	FPI, x_2	FTD, x_3	FTI, x_4	Potencia, x_5
1	4.75	170	170	106	106	240.57
2	4.07	140	130	92	93	195.49
3	4.04	180	170	93	78	152.99
4	4.18	160	160	103	93	197.09
5	4.35	170	150	104	93	266.56
6	4.16	150	150	101	87	260.56
7	4.43	170	180	108	106	219.25
8	3.20	110	110	86	92	132.68
9	3.02	120	110	90	86	130.24
10	3.64	130	120	85	80	205.88
11	3.68	120	140	89	83	153.92
12	3.60	140	130	92	94	154.64
13	3.85	160	150	95	95	240.57

Solución: Al buscar el mejor modelo posible para predecir el tiempo de vuelo se obtuvo la información de la tabla 12.13 utilizando un paquete de cómputo para regresión. Los modelos están clasificados en orden ascendente con respecto a los valores del estadístico PRESS. Esta presentación brinda información suficiente acerca de todos los modelos posibles con el fin de permitir que el usuario elimine algunos de ellos. Al parecer, el mejor modelo para predecir el tiempo de vuelo para los pateadores es el que contiene a x_2 y x_5 (FPI y potencia), denotadas por x_2, x_5 . Asimismo, observe que todos los modelos con valores bajos de PRESS, de s^2 , de $\sum_{i=1}^n |\delta_i|$, y con valores altos de R^2 , contienen esas dos variables. Para obtener información de los residuales de la regresión ajustada

$$\hat{y}_i = b_0 + b_2 x_{2i} + b_5 x_{5i},$$

se generaron los residuales y los residuales PRESS. El modelo de predicción real (véase el ejercicio 12.47 de la página 494) es dado por

$$\hat{y} = 1.10765 + 0.01370x_2 + 0.00429x_5.$$

En la tabla 12.14 se listan los residuales, los valores de la diagonal testada y los valores PRESS.

Observe el ajuste relativamente bueno de los modelos de regresión con dos variables para los datos. Los residuales PRESS reflejan la capacidad de la ecuación de regresión para predecir el tiempo de vuelo si se hicieran predicciones independientes. Por ejemplo, para el pateador número 4 el tiempo de vuelo de 4.180 tendría un error de predicción de 0.039 si se construyera el modelo usando a los 12 pateadores restantes. Para este modelo el error promedio de la predicción, o error de validación cruzada, es

$$\frac{1}{13} \sum_{i=1}^n |\delta_i| = 0.1489 \text{ segundos},$$

Tabla 12.13: Comparación de diferentes modelos de regresión

Modelo	s^2	$\sum \delta_i $	PRESS	R^2
x_2x_5	0.036907	1.93583	0.54683	0.871300
$x_1x_2x_5$	0.041001	2.06489	0.58998	0.871321
$x_2x_4x_5$	0.037708	2.18797	0.59915	0.881658
$x_2x_3x_5$	0.039636	2.09553	0.66182	0.875606
$x_1x_2x_4x_5$	0.042265	2.42194	0.67840	0.882093
$x_1x_2x_3x_5$	0.044578	2.26283	0.70958	0.875642
$x_2x_3x_4x_5$	0.042421	2.55789	0.86236	0.881658
$x_1x_3x_5$	0.053664	2.65276	0.87325	0.831580
$x_1x_4x_5$	0.056279	2.75390	0.89551	0.823375
x_1x_5	0.059621	2.99434	0.97483	0.792094
x_2x_3	0.056153	2.95310	0.98815	0.804187
x_1x_3	0.059400	3.01436	0.99697	0.792864
$x_1x_2x_3x_4x_5$	0.048302	2.87302	1.00920	0.882096
x_2	0.066894	3.22319	1.04564	0.743404
x_3x_5	0.065678	3.09474	1.05708	0.770971
x_1x_2	0.068402	3.09047	1.09726	0.761474
x_3	0.074518	3.06754	1.13555	0.714161
$x_1x_3x_4$	0.065414	3.36304	1.15043	0.794705
$x_2x_3x_4$	0.062082	3.32392	1.17491	0.805163
x_2x_4	0.063744	3.59101	1.18531	0.777716
$x_1x_2x_3$	0.059670	3.41287	1.26558	0.812730
x_3x_4	0.080605	3.28004	1.28314	0.718921
x_1x_4	0.069965	3.64415	1.30194	0.756023
x_1	0.080208	3.31562	1.30275	0.692334
$x_1x_3x_4x_5$	0.059169	3.37362	1.36867	0.834936
$x_1x_2x_4$	0.064143	3.89402	1.39834	0.798692
$x_3x_4x_5$	0.072505	3.49695	1.42036	0.772450
$x_1x_2x_3x_4$	0.066088	3.95854	1.52344	0.815633
x_5	0.111779	4.17839	1.72511	0.571234
x_4x_5	0.105648	4.12729	1.87734	0.631593
x_4	0.186708	4.88870	2.82207	0.283819

que es pequeño comparado con el tiempo de vuelo promedio para los 13 pateadores. ─

En la sección 12.9 indicamos que a menudo es aconsejable utilizar todos los subconjuntos posibles de regresión cuando se busca el mejor modelo. La mayoría de los programas comerciales de cómputo para estadística contienen una rutina de *todas las regresiones posibles*. Tales algoritmos calculan diversos criterios para todos los subconjuntos de términos del modelo. Es evidente que criterios como R^2 , s^2 y PRESS son razonables para elegir entre subconjuntos de candidatos. Otro estadístico muy popular y útil, en particular para las ciencias físicas e ingeniería, es el estadístico C_p , que se describe a continuación.

Tabla 12.14: Residuales PRESS

Pateador	y_i	\hat{y}_i	$e_i = y_i - \hat{y}_i$	h_{ii}	δ_i
1	4.750	4.470	0.280	0.198	0.349
2	4.070	3.728	0.342	0.118	0.388
3	4.040	4.094	-0.054	0.444	-0.097
4	4.180	4.146	0.034	0.132	0.039
5	4.350	4.307	0.043	0.286	0.060
6	4.160	4.281	-0.121	0.250	-0.161
7	4.430	4.515	-0.085	0.298	-0.121
8	3.200	3.184	0.016	0.294	0.023
9	3.020	3.174	-0.154	0.301	-0.220
10	3.640	3.636	0.004	0.231	0.005
11	3.680	3.687	-0.007	0.152	-0.008
12	3.600	3.553	0.047	0.142	0.055
13	3.850	4.196	-0.346	0.154	-0.409

El estadístico C_p

Muy a menudo la selección del modelo más adecuado implica tomar en cuenta muchas cosas. Evidentemente el número de términos del modelo es importante; el tema de la parsimonia no debe ignorarse. Por otro lado, el analista no debe sentirse satisfecho con un modelo demasiado simple hasta el punto de una simplificación excesiva. En este sentido, un estadístico único que implica un compromiso aceptable es C_p . (Véase Mallows, 1973, en la bibliografía).

El estadístico C_p apela de forma muy adecuada al sentido común y se desarrolla tomando en cuenta el equilibrio apropiado entre el sesgo excesivo en que se incurre cuando se subajusta, es decir, cuando se eligen muy pocos términos para el modelo; y la varianza excesiva de la predicción que se genera cuando se sobreajusta, o sea cuando hay redundancias en el modelo. El estadístico C_p es una función simple del número total de parámetros en el posible modelo y la media cuadrada del error s^2 .

Aquí no presentaremos el desarrollo completo del estadístico C_p . (Para mayores detalles se recomienda consultar a Myers, 1990, listado en la bibliografía). El C_p para un subconjunto particular de modelos es *un estimado* de lo siguiente:

$$\Gamma_{(p)} = \frac{1}{\sigma^2} \sum_{i=1}^n \text{Var}(\hat{y}_i) + \frac{1}{\sigma^2} \sum_{i=1}^n (\text{Sesgo } \hat{y}_i)^2.$$

Se descubre que, bajo las suposiciones estándar de los mínimos cuadrados que se indicaron con anterioridad en este capítulo, y asumiendo que el modelo “verdadero” es aquel que contiene todas las posibles variables,

$$\frac{1}{\sigma^2} \sum_{i=1}^n \text{Var}(\hat{y}_i) = p \quad (\text{número de parámetros en el posible modelo})$$

(véase el ejercicio de repaso 12.63) y un estimado no sesgado de

$$\frac{1}{\sigma^2} \sum_{i=1}^n (\text{Sesgo } \hat{y}_i)^2 \text{ es dado por } \frac{1}{\sigma^2} \sum_{i=1}^n (\widehat{\text{Sesgo}} \hat{y}_i)^2 = \frac{(s^2 - \sigma^2)(n - p)}{\sigma^2}.$$

En las ecuaciones anteriores s^2 es el cuadrado medio del error para el posible modelo y σ^2 es la varianza del error de la población. Así, si asumimos que se dispone de algún estimado $\hat{\sigma}^2$ para σ^2 , entonces C_p es dado por la siguiente ecuación:

Estadístico C_p

$$C_p = p + \frac{(s^2 - \hat{\sigma}^2)(n - p)}{\hat{\sigma}^2},$$

donde p es el número de parámetros en el modelo, s^2 es el cuadrado medio del error para el modelo candidato y $\hat{\sigma}^2$ es un estimador de σ^2 .

Es evidente que el científico debería adoptar modelos con valores pequeños de C_p . El lector observará que, a diferencia del estadístico PRESS, C_p carece de una escala. Además, se puede obtener cierta información acerca de qué tan adecuado es un posible modelo observando su valor de C_p . Por ejemplo, $C_p > p$ indica que un modelo está sesgado debido a que está subajustado, mientras que $C_p \approx p$ indica un modelo razonable.

Con frecuencia hay confusión respecto a la procedencia de $\hat{\sigma}^2$ en la fórmula para C_p . Es evidente que el científico o ingeniero no tienen acceso a la cantidad σ^2 de la población. En aplicaciones donde se dispone de corridas repetidas, digamos en situaciones de diseño experimental, se dispone de un estimado de σ^2 independiente del modelo (véase los capítulos 11 y 15). Sin embargo, la mayoría de paquetes de cómputo utilizan $\hat{\sigma}^2$ como el *cuadrado medio del error del modelo más completo*. Evidentemente, si éste no es un buen estimado, la parte de sesgo del estadístico C_p puede ser negativa. Por consiguiente, C_p puede ser menor que p .

Ejemplo 12.12: Considere el conjunto de datos de la tabla 12.15, los cuales reflejan el interés de un fabricante de grava asfáltica en la relación que existe entre las ventas durante un año específico y los factores que influyen en ellas. (Los datos fueron tomados de Kutner *et al.*, 2004, véase la bibliografía).

En los subconjuntos de modelos posibles, hay tres que revisten interés especial. Estos tres son los de x_2x_3 , $x_1x_2x_3$ y $x_1x_2x_3x_4$. A continuación se presenta la información pertinente para comparar los tres modelos. Para ayudar a la toma de decisiones incluimos los estadísticos PRESS de los tres modelos.

Modelo	R^2	R^2_{pred}	s^2	PRESS	C_p
x_2x_3	0.9940	0.9913	44.5552	782.1896	11.4013
$x_1x_2x_3$	0.9970	0.9928	24.7956	643.3578	3.4075
$x_1x_2x_3x_4$	0.9971	0.9917	26.2073	741.7557	5.0

A partir de la información de la tabla parece claro que el modelo $x_1x_2x_3$ es mejor que los otros dos. Observe que para el modelo completo $C_p = 5.0$. Esto ocurre porque la *parte de sesgo* es igual a cero y $\hat{\sigma}^2 = 26.2073$ es el cuadrado medio del error del modelo completo. ■

La figura 12.6 es una salida de resultados de la función PROC REG del SAS, la cual muestra información sobre todas las regresiones posibles. A partir de ella es posible hacer comparaciones de otros modelos con (x_1, x_2, x_3) . Observe que (x_1, x_2, x_3) parece muy bueno en comparación con todos los demás modelos.

Como verificación final del modelo (x_1, x_2, x_3) , la figura 12.7 presenta una gráfica de probabilidad normal de los residuales del modelo.

Tabla 12.15: Datos para el ejemplo 12.12

Distrito	Cuentas promocionales, x_1	Cuentas activas, x_2	Marcas en competencia, x_3	Potencial, x_4	Ventas, y (miles)
1	5.5	31	10	8	\$ 79.3
2	2.5	55	8	6	200.1
3	8.0	67	12	9	163.2
4	3.0	50	7	16	200.1
5	3.0	38	8	15	146.0
6	2.9	71	12	17	177.7
7	8.0	30	12	8	30.9
8	9.0	56	5	10	291.9
9	4.0	42	8	4	160.0
10	6.5	73	5	16	339.4
11	5.5	60	11	7	159.6
12	5.0	44	12	12	86.3
13	6.0	50	6	6	237.5
14	5.0	39	10	4	107.2
15	3.5	55	10	4	155.0

Dependent Variable: sales							
Number in Model	C(p)	R-Square	Adjusted R-Square	MSE	Variables in Model		
3	3.4075	0.9970	0.9961	24.79560	x1	x2	x3
4	5.0000	0.9971	0.9959	26.20728	x1	x2	x3 x4
2	11.4013	0.9940	0.9930	44.55518	x2	x3	
3	13.3770	0.9940	0.9924	48.54787	x2	x3	x4
3	1053.643	0.6896	0.6049	2526.96144	x1	x3	x4
2	1082.670	0.6805	0.6273	2384.14286	x3	x4	
2	1215.316	0.6417	0.5820	2673.83349	x1	x3	
1	1228.460	0.6373	0.6094	2498.68333	x3		
3	1653.770	0.5140	0.3814	3956.75275	x1	x2	x4
2	1668.699	0.5090	0.4272	3663.99357	x1	x2	
2	1685.024	0.5042	0.4216	3699.64814	x2	x4	
1	1693.971	0.5010	0.4626	3437.12846	x2		
2	3014.641	0.1151	-.0324	6603.45109	x1	x4	
1	3088.650	0.0928	0.0231	6248.72283	x4		
1	3364.884	0.0120	-.0640	6805.59568	x1		

Figura 12.6: Salida de resultados del SAS de todos los subconjuntos posibles sobre los datos de las ventas para el ejemplo 12.12.

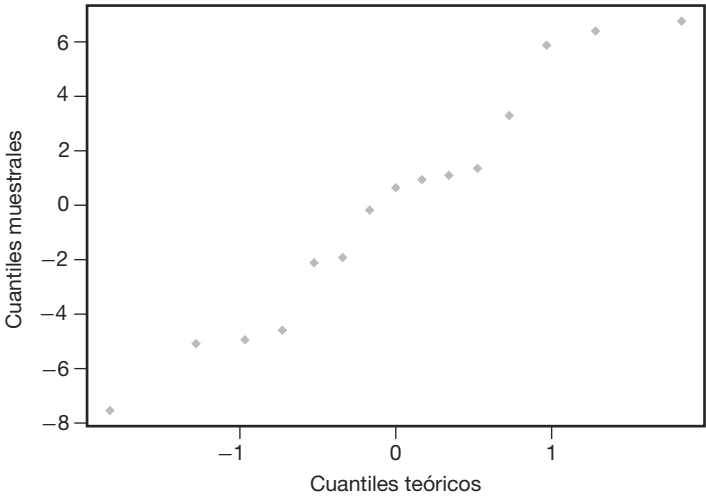


Figura 12.7: Gráfica de la probabilidad normal de los residuales, utilizando el modelo $x_1x_2x_3$ para el ejemplo 12.12.

Ejercicios

12.47 Considere los datos sobre el “tiempo de vuelo” de los pateadores que se presentaron en el estudio de caso 12.2 y, utilizando sólo las variables x_2 y x_3 ,

- a) Verifique la ecuación de regresión que se presenta en la página 489.
- b) Prediga el tiempo de vuelo para un pateador con $FPI = 180$ libras y potencia = 260 pie-libras.
- c) Construya un intervalo de confianza de 95% para el tiempo de vuelo promedio de un pateador con $FPI = 180$ libras y potencia = 260 pies-libras.

12.48 Para los datos del ejercicio 12.15 de la página 452 utilice las técnicas de

- a) *selección hacia delante* a un nivel de significancia de 0.05 para elegir un modelo de regresión lineal;
- b) *eliminación hacia atrás* a un nivel de significancia de 0.05 para seleccionar un modelo de regresión lineal;
- c) *regresión por etapas* a un nivel de significancia de 0.05 para escoger un modelo de regresión lineal.

12.49 Emplee las técnicas de *eliminación hacia atrás* con $\alpha = 0.05$ para elegir una ecuación de predicción para los datos de la tabla 12.8.

12.50 Para los datos de los pateadores del estudio de caso 12.2 también se registró una respuesta adicional, la “distancia de la patada”. Los siguientes son los valores de distancia promedio para cada uno de los 13 pateadores:

- a) Utilice los datos de distancia en lugar de los de tiempo de vuelo para estimar un modelo de regresión lineal múltiple del tipo

$$\begin{aligned} \mu_Y | x_1, x_2, x_3, x_4, x_5 \\ = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \end{aligned}$$

que permita predecir la distancia de la patada.

- b) Utilice la regresión por etapas a un nivel de significancia de 0.10 para seleccionar una combinación de variables.
- c) Genere valores para s^2 , R^2 , PRESS y $\sum_{i=1}^n |\delta_i|$ para el conjunto completo de 31 modelos. Utilice esta información para determinar la mejor combinación de variables para predecir la distancia de la patada.
- d) Para el modelo final que seleccione, grafique los residuales estandarizados en comparación con Y y elabore una gráfica de probabilidad normal de los residuales ordinarios. Haga comentarios al respecto.

Pateador	Distancia, y (pies)
1	162.50
2	144.00
3	147.50
4	163.50
5	192.00
6	171.75
7	162.00
8	104.93
9	105.67
10	117.59
11	140.25
12	150.17
13	165.16

12.51 El siguiente es un conjunto de datos para y , la cantidad de dinero (en miles de dólares) aportado a la asociación de exalumnos del Virginia Tech por la generación de 1960; y para x , el número de años que han transcurrido desde la graduación:

y	x	y	x
812.52	1	2755.00	11
822.50	2	4390.50	12
1211.50	3	5581.50	13
1348.00	4	5548.00	14
1301.00	8	6086.00	15
2567.50	9	5764.00	16
2526.50	10	8903.00	17

a) Ajuste un modelo de regresión del tipo

$$\mu_{Y|x} = \beta_0 + \beta_1 x.$$

b) Ajuste un modelo cuadrático del tipo

$$\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_{11} x^2.$$

c) Determine cuál de los modelos de los incisos a) o b) es preferible. Utilice s^2 , R^2 y los residuales PRESS para sustentar su decisión.

12.52 Para el modelo del ejercicio 12.50a) pruebe la hipótesis

$$H_0: \beta_4 = 0,$$

$$H_1: \beta_4 \neq 0.$$

Utilice un valor P para su conclusión.

12.53 Para el modelo cuadrático del ejercicio 12.51b) proporcione estimados de las varianzas y las covarianzas de los estimados de β_1 y β_{11} .

12.54 Un cliente del Departamento de Ingeniería Mecánica se acercó al Centro de Consulta de Virginia Tech para que lo ayudaran a analizar un experimento sobre motores con turbina de gas. Se midieron varias salidas del voltaje de los motores con distintas combinaciones de velocidad de las aspas y de la extensión de los sensores. Los datos son los siguientes:

y (voltios)	Velocidad, x_1 (pulg/seg)	Extensión, x_2 (pulg)
1.95	6336	0.000
2.50	7099	0.000
2.93	8026	0.000
1.69	6230	0.000
1.23	5369	0.000
3.13	8343	0.000
1.55	6522	0.006
1.94	7310	0.006
2.18	7974	0.006
2.70	8501	0.006
1.32	6646	0.012
1.60	7384	0.012
1.89	8000	0.012
2.15	8545	0.012
1.09	6755	0.018
1.26	7362	0.018
1.57	7934	0.018
1.92	8554	0.018

a) Ajuste una regresión lineal múltiple a los datos.

b) Calcule las pruebas t sobre los coeficientes. Proporcione valores P .

c) Haga comentarios sobre la calidad del modelo ajustado.

12.55 La blancura del rayón es un factor importante para los científicos que estudian la calidad de las telas. La blancura se ve afectada por la calidad de la pulpa y otras variables de procesamiento. Algunas de las variables son la temperatura del baño con ácido, $^{\circ}\text{C}$ (x_1); la concentración del ácido en cascada, % (x_2); la temperatura del agua, $^{\circ}\text{C}$ (x_3); la concentración del sulfuro, % (x_4); la cantidad del blanqueador de cloro, lb/min (x_5) y la temperatura de terminación de la tela, $^{\circ}\text{C}$ (x_6). A continuación se proporciona un conjunto de datos de especímenes de rayón. La respuesta, y , es la medida de la blancura.

y	x_1	x_2	x_3	x_4	x_5	x_6
88.7	43	0.211	85	0.243	0.606	48
89.3	42	0.604	89	0.237	0.600	55
75.5	47	0.450	87	0.198	0.527	61
92.1	46	0.641	90	0.194	0.500	65
83.4	52	0.370	93	0.198	0.485	54
44.8	50	0.526	85	0.221	0.533	60
50.9	43	0.486	83	0.203	0.510	57
78.0	49	0.504	93	0.279	0.489	49
86.8	51	0.609	90	0.220	0.462	64
47.3	51	0.702	86	0.198	0.478	63
53.7	48	0.397	92	0.231	0.411	61
92.0	46	0.488	88	0.211	0.387	88
87.9	43	0.525	85	0.199	0.437	63
90.3	45	0.486	84	0.189	0.499	58
94.2	53	0.527	87	0.245	0.530	65
89.5	47	0.601	95	0.208	0.500	67

a) Utilice los criterios CME , c_p y PRESS para obtener el “mejor” modelo de todos los subconjuntos de los modelos.

b) Grafique los residuales estandarizados en comparación con Y y dibuje una gráfica de probabilidad normal de los residuales para el “mejor” modelo. Comente sus resultados.

12.56 En un esfuerzo para modelar las remuneraciones de los ejecutivos en el año 1979 se seleccionaron 33 empresas y se recabaron datos acerca de las remuneraciones, las ventas, las utilidades y el empleo. Se reunieron los siguientes datos para el año 1979.

Empresa	Remuneraciones, y (miles)	Ventas, x_1 , (millones)	Utilidades, x_2 , (millones)	Empleo, x_3
1	\$450	\$4600.6	\$128.1	48,000
2	387	9255.4	783.9	55,900
3	368	1526.2	136.0	13,783
4	277	1683.2	179.0	27,765
5	676	2752.8	231.5	34,000
6	454	2205.8	329.5	26,500
7	507	2384.6	381.8	30,800
8	496	2746.0	237.9	41,000
9	487	1434.0	222.3	25,900

(cont.)

Empresa	Remuneraciones, y (miles)	Ventas, x_1 (millones)	Utilidades, x_2 (millones)	Empleo, x_3
10	\$383	\$470.6	\$63.7	8600
11	311	1508.0	149.5	21,075
12	271	464.4	30.0	6874
13	524	9329.3	577.3	39,000
14	498	2377.5	250.7	34,300
15	343	1174.3	82.6	19,405
16	354	409.3	61.5	3586
17	324	724.7	90.8	3905
18	225	578.9	63.3	4139
19	254	966.8	42.8	6255
20	208	591.0	48.5	10,605
21	518	4933.1	310.6	65,392
22	406	7613.2	491.6	89,400
23	332	3457.4	228.0	55,200
24	340	545.3	54.6	7800
25	698	22,862.8	3011.3	337,119
26	306	2361.0	203.0	52,000
27	613	2614.1	201.0	50,500
28	302	1013.2	121.3	18,625
29	540	4560.3	194.6	97,937
30	293	855.7	63.4	12,300
31	528	4211.6	352.1	71,800
32	456	5440.4	655.2	87,700
33	417	1229.9	97.5	14,600

Considere el modelo

$$y_i = \beta_0 + \beta_1 \ln x_{1i} + \beta_2 \ln x_{2i} + \beta_3 \ln x_{3i} + \epsilon_i, \quad i = 1, 2, \dots, 33.$$

- Ajuste la regresión con el modelo anterior.
- ¿Un modelo con un subconjunto de variables es preferible al modelo completo?

12.57 La resistencia a la tracción de una unión de alambre es una característica importante. La siguiente tabla brinda información sobre la resistencia a la tracción, y , la altura del molde, x_1 , la altura del perno, x_2 , la altura del lazo, x_3 , la longitud del alambre, x_4 , el ancho de la unión sobre el molde, x_5 y el ancho del molde sobre el perno, x_6 . (Datos tomados de Myers, Montgomery y Anderson-Cook, 2009).

- Ajuste un modelo de regresión usando todas las variables independientes.
- Utilice la regresión por etapas a un nivel de significancia de entrada de 0.25 y un nivel de significancia de eliminación de 0.05. Proporcione el modelo final.
- Utilice todos los modelos de regresión posibles y

calcule R^2 , C_p , s^2 y R^2 ajustada para todos los modelos.

d) Proporcione el modelo final.

e) Para el modelo del inciso d) grafique los residuos estudentizados (o la R de Student) y haga comentarios al respecto.

y	x_1	x_2	x_3	x_4	x_5	x_6
8.0	5.2	19.6	29.6	94.9	2.1	2.3
8.3	5.2	19.8	32.4	89.7	2.1	1.8
8.5	5.8	19.6	31.0	96.2	2.0	2.0
8.8	6.4	19.4	32.4	95.6	2.2	2.1
9.0	5.8	18.6	28.6	86.5	2.0	1.8
9.3	5.2	18.8	30.6	84.5	2.1	2.1
9.3	5.6	20.4	32.4	88.8	2.2	1.9
9.5	6.0	19.0	32.6	85.7	2.1	1.9
9.8	5.2	20.8	32.2	93.6	2.3	2.1
10.0	5.8	19.9	31.8	86.0	2.1	1.8
10.3	6.4	18.0	32.6	87.1	2.0	1.6
10.5	6.0	20.6	33.4	93.1	2.1	2.1
10.8	6.2	20.2	31.8	83.4	2.2	2.1
11.0	6.2	20.2	32.4	94.5	2.1	1.9
11.3	6.2	19.2	31.4	83.4	1.9	1.8
11.5	5.6	17.0	33.2	85.2	2.1	2.1
11.8	6.0	19.8	35.4	84.1	2.0	1.8
12.3	5.8	18.8	34.0	86.9	2.1	1.8
12.5	5.6	18.6	34.2	83.0	1.9	2.0

12.58 Para el ejercicio 12.57 pruebe $H_0: \beta_1 = \beta_6 = 0$. Proporcione valores P y comente al respecto.

12.59 En el ejercicio 12.28 de la página 462 se tienen los siguientes datos sobre el desgaste de un cojinete:

y (desgaste)	x_1 (viscosidad del aceite)	x_2 (carga)
193	1.6	851
230	15.5	816
172	22.0	1058
91	43.0	1201
113	33.0	1357
125	40.0	1115

a) Puede considerar el siguiente modelo para describir los datos:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + \epsilon_i,$$

para $i = 1, 2, \dots, 6$. El término $x_1 x_2$ es una “interacción”. Ajuste este modelo y estime los parámetros.

b) Utilice los modelos (x_1) , (x_1, x_2) , (x_2) , $(x_1, x_2, x_1 x_2)$ y calcule PRESS, C_p , y s^2 para determinar el “mejor” modelo.

12.12 Modelos especiales no lineales para condiciones no ideales

En gran parte del material anterior de este capítulo y en el del capítulo 11 nos hemos beneficiado mucho de la suposición de que los errores del modelo, los ϵ_i , son normales,

con media igual a cero y varianza constante σ^2 . Sin embargo, en la vida real hay muchas situaciones en las cuales es evidente que la respuesta no es normal. Por ejemplo, existe una gran cantidad de aplicaciones en las que la **respuesta es binaria** (0 o 1), por lo que su naturaleza es de Bernoulli. En las ciencias sociales un problema podría ser el de desarrollar un modelo que prediga si un individuo representa riesgos para un crédito (0 o 1), en función de ciertos regresores socioeconómicos, como sus ingresos, edad, género y nivel de escolaridad. En una prueba biomédica para un fármaco a menudo se observa si el paciente responde o no de manera favorable a éste, en tanto que los regresores podrían incluir la dosis y factores biológicos como la edad, el peso y la presión sanguínea. Nuevamente la respuesta es de naturaleza binaria. También abundan las aplicaciones en las áreas de manufactura en que ciertos factores controlables influyen en el hecho de que un artículo fabricado esté **o no defectuoso**.

Un segundo tipo de aplicación que no es normal y del que haremos una mención breve tiene que ver con el **conteo de datos**. Aquí a menudo es conveniente suponer una respuesta de Poisson. En aplicaciones biomédicas la respuesta que se modela en comparación con las dosis de medicamentos podría ser el número de colonias de células cancerosas. En la industria textil una respuesta razonable que se modela en comparación con ciertas variables de los procesos es el número de imperfecciones por yarda de tela.

Varianza no homogénea

El lector debería notar la comparación de la situación ideal, es decir, la respuesta normal, con la de la respuesta de Bernoulli (o binomial) o la de Poisson. Nos hemos acostumbrado al hecho de que el caso normal es muy especial debido a que la varianza es **independiente de la media**. Resulta claro que éste no es el caso para la respuesta de Bernoulli ni la de Poisson. Por ejemplo, si la respuesta es 0 o 1, lo cual sugiere una respuesta de Bernoulli, entonces el modelo adopta la forma

$$p = f(\mathbf{x}, \beta),$$

donde p es la **probabilidad de un éxito** (por ejemplo, la respuesta = 1). El parámetro p desempeña el papel de $\mu_{y|x}$ en el caso normal. Sin embargo, la varianza de Bernoulli es $p(1 - p)$ que, desde luego, también es una función del regresor \mathbf{x} . Como resultado, la varianza no es constante. Esto descarta el uso de los mínimos cuadrados estándar que hemos utilizado en nuestro trabajo de regresión lineal hasta este momento. Lo mismo se aplica para el caso de Poisson, ya que el modelo adopta la forma

$$\lambda = f(\mathbf{x}, \beta),$$

con $\text{Var}(y) = \mu_y = \lambda$, que varía con \mathbf{x} .

Respuesta binaria (regresión logística)

El enfoque más popular para modelar respuestas binarias es la técnica llamada **regresión logística**, la cual se emplea mucho en las ciencias biológicas, en la investigación biomédica y en la ingeniería. De hecho, se observa que incluso en las ciencias sociales abundan las respuestas binarias. La distribución básica para la respuesta es la de Bernoulli o la binomial. La primera se encuentra en estudios observacionales donde no hay corridas repetidas en cada nivel de regresor; mientras que la segunda será el caso en que se utilice un diseño experimental. Por ejemplo, en un ensayo clínico en el cual se evalúa un fármaco nuevo, el objetivo podría ser el de determinar la dosis del medicamento que es

eficaz. Así, en el experimento se utilizarán ciertas dosis y para cada una de ellas se emplearán a varios sujetos, un caso al que se le denomina **caso agrupado**.

¿Cuál es el modelo para la regresión logística?

En el caso de respuestas binarias la respuesta media es una probabilidad. En la ilustración anterior del ensayo clínico podríamos decir que deseamos estimar la probabilidad de que el paciente responda en forma adecuada al fármaco, $P(\text{éxito})$. Entonces, el modelo se escribe en términos de una probabilidad. Dados los regresores \mathbf{x} , la función logística es dada por

$$p = \frac{1}{1 + e^{-\mathbf{x}'\beta}}.$$

La porción $\mathbf{x}'\beta$ se llama **predictor lineal** y, en el caso de un solo regresor x , se puede escribir $\mathbf{x}'\beta = \beta_0 + \beta_1 x$. Desde luego, no descartamos la inclusión de regresores múltiples y de términos polinomiales en el llamado predictor lineal. En el caso agrupado el modelo implica el modelado de la media de una binomial en vez de una de Bernoulli, por lo que la media es dada por

$$np = \frac{n}{1 + e^{-\mathbf{x}'\beta}}.$$

Características de la función logística

Una gráfica de la función logística revela mucho sobre sus características y del porqué se utiliza para este tipo de problema. En primer lugar, la función es no lineal. Además, la gráfica de la figura 12.8 revela la forma de S con la función que tiende a la asíntota en $p = 1.0$. En este caso, $\beta_1 > 0$. Así, nunca se experimentaría una probabilidad estimada mayor que 1.0.

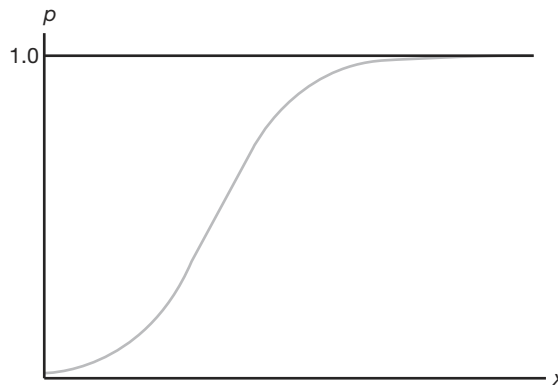


Figura 12.8: La función logística.

Los coeficientes de regresión en el predictor lineal se estiman con el método de probabilidad máxima, tal como se describió en el capítulo 9. La solución de las ecuaciones

de probabilidad requiere una metodología iterativa que no se describe aquí. Sin embargo, presentaremos un ejemplo y analizaremos la salida de resultados por computadora y las conclusiones.

Ejemplo 12.13: El conjunto de datos de la tabla 12.16 se utilizará con el fin de ilustrar el uso de la regresión logística para analizar un ensayo biológico cuantal de agente único en un experimento de toxicidad. Los resultados muestran el efecto de diferentes dosis de nicotina en la mosca común de la fruta.

Tabla 12.16: Conjunto de datos para el ejemplo 12.13

x Concentración (gramos/100 cc)	n_i Número de insectos	y Número de muertes	Porcentaje de muertes
0.10	47	8	17.0
0.15	53	14	26.4
0.20	55	24	43.6
0.30	52	32	61.5
0.50	46	38	82.6
0.70	54	50	92.6
0.95	52	50	96.2

El propósito del experimento era el de obtener un modelo adecuado que relacionara la probabilidad de “muerte” con la concentración. Además, el analista buscaba la denominada **dosis eficaz** (DE), es decir, la concentración de nicotina que da como resultado cierta probabilidad. La DE_{50} tiene interés particular, ya que es la concentración que produce una probabilidad de 0.5 de que el “insecto muera”.

Este ejemplo es agrupado, por lo que el modelo es dado por

$$E(Y_i) = n_i p_i = \frac{n_i}{1 + e^{-(\beta_0 + \beta_1 x_i)}}.$$

Los estimados de β_0 y β_1 , y sus errores estándar, se calculan usando el método de probabilidad máxima. Las pruebas de los coeficientes individuales se calculan utilizando el estadístico χ^2 en lugar del estadístico t , puesto que no hay una varianza común σ^2 . El estadístico χ^2 se obtiene a partir de (coef/error estándar)².

Por consiguiente, obtenemos la siguiente salida de resultados por computadora de la función PROC LOGIST del SAS.

Análisis de los estimados de los parámetros					
	gl	Estimado	Error estándar	Chi cuadrada	Valor P
β_0	1	-1.7361	0.2420	51.4482	< 0.0001
β_1	1	6.2954	0.7422	71.9399	< 0.0001

Ambos coeficientes difieren significativamente de cero. Por consiguiente, el modelo ajustado que se emplea para predecir la probabilidad de “muerte” es dado por

$$\hat{p} = \frac{1}{1 + e^{-(-1.7361 + 6.2954 x)}}.$$



Estimado de la dosis eficaz

El estimado de la DE_{50} para el ejemplo 12.13 se calcula de manera muy sencilla a partir de los estimados b_0 para β_0 y b_1 para β_1 . A partir de la función logística se observa que

$$\log \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 x.$$

Como resultado, para $p = 0.5$ se calcula un estimado de x a partir de

$$b_0 + b_1 x = 0.$$

Así, DE_{50} es dada por

$$x = - \left(\frac{b_0}{b_1} \right) = 0.276 \text{ gramos/100 cc.}$$

Concepto de razón de probabilidad

Otra forma de inferencia que se lleva a cabo de manera adecuada usando la regresión logística se deriva del uso de la razón de probabilidad, la cual está diseñada para determinar cómo se incrementa la **probabilidad de éxitos**, $\frac{p}{1-p}$, a medida que ocurren ciertos cambios en los valores del regresor. Por ejemplo, en el caso del ejemplo 12.13, quizá se deseara saber cómo aumentarían las probabilidades si la dosis se incrementara en, digamos, 0.2 gramos/100 cc.

Definición 12.1: En la regresión logística una **razón de probabilidad** es la razón de la probabilidad de éxito en la condición 2 con respecto a la de la condición 1 en los regresores, es decir,

$$\frac{[p/(1-p)]_2}{[p/(1-p)]_1}.$$

Esto permite que el analista tenga una idea de la utilidad de cambiar el regresor en cierto número de unidades. Ahora, como $\left(\frac{p}{1-p} \right) = e^{\beta_0 + \beta_1 x}$, para el ejemplo 12.13 la razón que refleja el incremento de las probabilidades de éxito cuando aumenta la dosis de nicotina en 0.2 gramos/100 cc es dada por

$$e^{0.2b_1} = e^{(0.2)(6.2954)} = 3.522.$$

La implicación de una razón de probabilidad de 3.522 es que la probabilidad de éxito aumenta en un factor de 3.522 cuando la dosis de nicotina aumenta en 0.2 gramos/100 cc.

Ejercicios

12.60 A partir de un conjunto de datos de respuestas a la dosis de estreptomycin un investigador desea desarrollar una relación entre la proporción de linfoblastos muestreados que contienen aberraciones y la dosis del medicamento. Se aplicaron cinco niveles de dosis a los conejos que se emplearon para el experimento. Los datos son los siguientes (véase Myers, 1990, listado en la bibliografía):

Dosis (mg/kg)	Número de linfoblastos	Número de aberraciones
0	600	15
30	500	96
60	600	187
75	300	100
90	300	145

- a) Ajuste una regresión logística al conjunto de datos, y así estime β_0 y β_1 en el modelo

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}},$$

donde n es el número de linfoblastos, x es la dosis y p la probabilidad de una aberración.

- b) Muestre los resultados de pruebas χ^2 que revelen la significancia de los coeficientes de regresión β_0 y β_1 .
- c) Estime la DE_{50} e interprétela.

12.61 En un experimento para estudiar el efecto de la carga, x , en lb/pulgadas², sobre la probabilidad de falla de especímenes de cierto tipo de tela, varios especímenes se expusieron a cargas de entre 5 lb/pulg² a 90 lb/pulg². Se observaron los números de “fallas”. Los datos son los siguientes:

Carga	Número de especímenes	Número de fallas
5	600	13
35	500	95
70	600	189
80	300	95
90	300	130

- a) Utilice regresión logística para ajustar el modelo

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}},$$

donde p es la probabilidad de falla y x es la carga.

- b) Emplee el concepto de razón de probabilidad para determinar el incremento de la probabilidad de falla que resulta de aumentar la carga en 20 lb/pulg².

Ejercicios de repaso

12.62 En el Departamento de Pesca y Vida Silvestre de Virginia Tech se realizó un experimento para estudiar el efecto de las características de la corriente sobre la biomasa de los peces. Las variables regresoras son las siguientes: profundidad promedio (de 50 células), x_1 ; área de la cubierta en la corriente, es decir, riberas socavadas, troncos, cantos rodados, etc., x_2 ; porcentaje de cubierta de material translúcido (promedio de 12), x_3 ; y un área ≥ 25 centímetros de profundidad, x_4 . La respuesta es y , la biomasa de los peces. Los datos son los siguientes:

Obs.	y	x_1	x_2	x_3	x_4
1	100	14.3	15.0	12.2	48.0
2	388	19.1	29.4	26.0	152.2
3	755	54.6	58.0	24.2	469.7
4	1288	28.8	42.6	26.1	485.9
5	230	16.1	15.9	31.6	87.6
6	0	10.0	56.4	23.3	6.9
7	551	28.5	95.1	13.0	192.9
8	345	13.8	60.6	7.5	105.8
9	0	10.7	35.2	40.3	0.0
10	348	25.9	52.0	40.3	116.6

- a) Ajuste una regresión lineal múltiple que incluya las cuatro variables regresoras.
- b) Utilice C_p , R^2 y s^2 para determinar el mejor subconjunto de variables. Calcule dichos estadísticos para todos los subconjuntos posibles.
- c) Compare lo adecuado de los modelos de los incisos a) y b) para efectos de predecir la biomasa de los peces.

12.63 Demuestre que, en un conjunto de datos de regresión lineal múltiple,

$$\sum_{i=1}^n h_{ii} = p.$$

12.64 Se efectuó un experimento sencillo para ajustar una ecuación de regresión múltiple que relaciona al producto, y , con la temperatura, x_1 , el tiempo de reacción, x_2 , y la concentración de uno de los reactivos, x_3 . Se eligieron dos niveles de cada variable y se hicieron las siguientes mediciones correspondientes a las variables independientes definidas:

y	x_1	x_2	x_3
7.6	-1	-1	-1
5.5	1	-1	-1
9.2	-1	1	-1
10.3	-1	-1	1
11.6	1	1	-1
11.1	1	-1	1
10.2	-1	1	1
14.0	1	1	1

- a) Utilice las variables codificadas y estime la ecuación de regresión lineal múltiple

$$\mu_{y|x_1, x_2, x_3} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

- b) Separe la SCR , es decir, la suma de cuadrados de regresión, en tres componentes con un grado de libertad, atribuibles a x_1 , x_2 y x_3 , respectivamente. Construya una tabla de análisis de varianza donde se indiquen pruebas de significancia sobre cada variable. Comente los resultados.

12.65 En un experimento de ingeniería química relacionado con la transferencia de calor en una capa de fluido superficial se recabaron datos sobre las cuatro variables regresoras siguientes: la tasa de flujo del gas fluido en lb/hr (x_1), la tasa de flujo del gas flotante en lb/hr (x_2), la abertura de la boquilla de entrada del gas flotante en milímetros (x_3) y la temperatura de entrada del gas flotante en °F (x_4). Las respuestas medidas son la eficacia de la transferencia de calor (y_1) y la eficacia térmica (y_2). Los datos son los siguientes:

Obs.	y_1	y_2	x_1	x_2	x_3	x_4
1	41.852	38.75	69.69	170.83	45	219.74
2	155.329	51.87	113.46	230.06	25	181.22
3	99.628	53.79	113.54	228.19	65	179.06
4	49.409	53.84	118.75	117.73	65	281.30
5	72.958	49.17	119.72	117.69	25	282.20
6	107.702	47.61	168.38	173.46	45	216.14
7	97.239	64.19	169.85	169.85	45	223.88
8	105.856	52.73	169.85	170.86	45	222.80
9	99.348	51.00	170.89	173.92	80	218.84
10	111.907	47.37	171.31	173.34	25	218.12
11	100.008	43.18	171.43	171.43	45	219.20
12	175.380	71.23	171.59	263.49	45	168.62
13	117.800	49.30	171.63	171.63	45	217.58
14	217.409	50.87	171.93	170.91	10	219.92
15	41.725	54.44	173.92	71.73	45	296.60
16	151.139	47.93	221.44	217.39	65	189.14
17	220.630	42.91	222.74	221.73	25	186.08
18	131.666	66.60	228.90	114.40	25	285.80
19	80.537	64.94	231.19	113.52	65	286.34
20	152.966	43.18	236.84	167.77	45	221.72

Considere el modelo para predecir la respuesta del coeficiente de transferencia de calor

$$y_{1i} = \beta_0 + \sum_{j=1}^4 \beta_j x_{ji} + \sum_{i=1}^4 \beta_{ji} x_{ji}^2 + \sum_{j \neq l} \beta_{jl} x_{ji} x_{li} + \epsilon_i, \quad i = 1, 2, \dots, 20.$$

- Calcule PRESS y $\sum_{i=1}^n |y_i - \hat{y}_{i,-i}|$ para ajustar el modelo anterior con los mínimos cuadrados de regresión.
- Ajuste un modelo de segundo orden con x_4 eliminada por completo, es decir, elimine todos los términos que impliquen x_4 . Calcule los criterios de predicción para el modelo reducido. Comente qué tan adecuada es x_4 para predecir el coeficiente de transferencia de calor.
- Repita los incisos a) y b) para la eficacia térmica.

12.66 En la fisiología del deporte una medición objetiva de la condición física es el consumo de oxígeno en volumen por unidad de peso corporal por unidad de tiempo. Se estudiaron 31 individuos en un experimento con el fin de modelar el consumo de oxígeno en comparación con la edad en años, x_1 , el peso en kilogramos, x_2 , el tiempo para correr 1 1/2 millas, x_3 , las pulsaciones en reposo, x_4 , las pulsaciones al final de la carrera, x_5 , y las pulsaciones máximas durante la carrera, x_6 .

- Realice una regresión por etapas a un nivel de significancia de 0.25 en la entrada. Proporcione el modelo final.
- Estudie todos los subconjuntos posibles usando s^2 , C_p , R^2 y R^2_{ajus} . Tome una decisión y determine el modelo final.

ID	y	x_1	x_2	x_3	x_4	x_5	x_6
1	44.609	44	89.47	11.37	62	178	182
2	45.313	40	75.07	10.07	62	185	185
3	54.297	44	85.84	8.65	45	156	168
4	59.571	42	68.15	8.17	40	166	172
5	49.874	38	89.02	9.22	55	178	180
6	44.811	47	77.45	11.63	58	176	176
7	45.681	40	75.98	11.95	70	176	180
8	49.091	43	81.19	10.85	64	162	170
9	39.442	44	81.42	13.08	63	174	176
10	60.055	38	81.87	8.63	48	170	186
11	50.541	44	73.03	10.13	45	168	168
12	37.388	45	87.66	14.03	56	186	192
13	44.754	45	66.45	11.12	51	176	176
14	47.273	47	79.15	10.60	47	162	164
15	51.855	54	83.12	10.33	50	166	170
16	49.156	49	81.42	8.95	44	180	185
17	40.836	51	69.63	10.95	57	168	172
18	46.672	51	77.91	10.00	48	162	168
19	46.774	48	91.63	10.25	48	162	164
20	50.388	49	73.37	10.08	76	168	168
21	39.407	57	73.37	12.63	58	174	176
22	46.080	54	79.38	11.17	62	156	165
23	45.441	52	76.32	9.63	48	164	166
24	54.625	50	70.87	8.92	48	146	155
25	45.118	51	67.25	11.08	48	172	172
26	39.203	54	91.63	12.88	44	168	172
27	45.790	51	73.71	10.47	59	186	188
28	50.545	57	59.08	9.93	49	148	155
29	48.673	49	76.32	9.40	56	186	188
30	47.920	48	61.24	11.50	52	170	176
31	47.467	52	82.78	10.50	53	170	172

12.67 Considere los datos del ejercicio de repaso 12.64. Suponga que le interesa agregar algunos términos de “interacción”. En específico, considere el modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_{12} x_{1i} x_{2i} + \beta_{13} x_{1i} x_{3i} + \beta_{23} x_{2i} x_{3i} + \beta_{123} x_{1i} x_{2i} x_{3i} + \epsilon_i.$$

- ¿Aún se tiene ortogonalidad? Comente al respecto.
- Con el modelo ajustado del inciso a), ¿puede usted encontrar intervalos de predicción y de confianza sobre la respuesta media? Explique su respuesta.
- Considere un modelo en el que se eliminó $\beta_{123} x_1 x_2 x_3$. Para determinar si son necesarias las interacciones (como un todo), pruebe

$$H_0: \beta_{12} = \beta_{13} = \beta_{23} = 0.$$

Proporcione el valor P y saque conclusiones.

12.68 Para extraer petróleo crudo se utiliza una técnica de inyección de dióxido de carbono (CO_2). El flujo de CO_2 envuelve el petróleo y lo desplaza. En un experimento se introducen tubos de flujo en muestras de cavidades de petróleo que contienen una cantidad conocida del mismo. Se utilizan tres valores diferentes de

presión de flujo y tres valores diferentes de ángulos de introducción, las cavidades de petróleo se inyectan con CO_2 y se registra el porcentaje de petróleo desplazado. Considere el modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + \epsilon_i.$$

Ajuste el modelo anterior a los datos y sugiera cualquier modificación al modelo que considere necesaria.

Presión lb/pulg ² , x_1	Ángulo de inyección, x_2	Recuperación de petróleo, (%), y
1000	0	60.58
1000	15	72.72
1000	30	79.99
1500	0	66.83
1500	15	80.78
1500	30	89.78
2000	0	69.18
2000	15	80.31
2000	30	91.99

Fuente: Wang, G. C. "Microscopic Investigations of CO_2 Flooding Process", Journal of Petroleum Technology, vol. 34, núm. 8, agosto de 1982.

12.69 Un artículo del *Journal of Pharmaceutical Sciences* (vol. 80, 1991) presenta datos de la solubilidad de una fracción molar de un soluto a temperatura constante. También se midió la dispersión, x_1 , y los parámetros de solubilidad del enlace bipolar y de hidrógeno, x_2 y x_3 . En la tabla siguiente se presenta una parte de los datos. En el modelo, y es el logaritmo negativo de la fracción molar. Ajuste el modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i,$$

para $i = 1, 2, \dots, 20$.

Obs.	y	x_1	x_2	x_3
1	0.2220	7.3	0.0	0.0
2	0.3950	8.7	0.0	0.3
3	0.4220	8.8	0.7	1.0
4	0.4370	8.1	4.0	0.2
5	0.4280	9.0	0.5	1.0
6	0.4670	8.7	1.5	2.8
7	0.4440	9.3	2.1	1.0
8	0.3780	7.6	5.1	3.4
9	0.4940	10.0	0.0	0.3
10	0.4560	8.4	3.7	4.1
11	0.4520	9.3	3.6	2.0
12	0.1120	7.7	2.8	7.1
13	0.4320	9.8	4.2	2.0
14	0.1010	7.3	2.5	6.8
15	0.2320	8.5	2.0	6.6
16	0.3060	9.5	2.5	5.0
17	0.0923	7.4	2.8	7.8
18	0.1160	7.8	2.8	7.7
19	0.0764	7.7	3.0	8.0
20	0.4390	10.3	1.7	4.2

- Pruebe $H_0: \beta_1 = \beta_2 = \beta_3 = 0$.
- Grafique los residuales estudentizados en comparación con x_1 , x_2 y x_3 (tres gráficas). Haga comentarios al respecto.
- Considere dos modelos adicionales que compitan con el modelo anterior:

Modelo 2: Agregue x_1^2, x_2^2, x_3^2 .

Modelo 3: Agregue $x_1^2, x_2^2, x_3^2, x_1 x_2, x_1 x_3, x_2 x_3$.

Utilice PRESS y C_p con estos tres modelos para saber cuál de los tres es el mejor.

12.70 Se realizó un estudio para determinar si los cambios en el estilo de vida podrían sustituir la medicación para reducir la presión sanguínea de los individuos hipertensos. Los factores considerados fueron una dieta saludable con un programa de ejercicios, la dosis común de medicamentos para la hipertensión y ningún tratamiento. También se calculó el índice de masa corporal (IMC) previo al tratamiento, debido a que se sabe que éste afecta la presión sanguínea. La respuesta considerada en este estudio fue el cambio en la presión sanguínea. La variable "grupo" tenía los siguientes niveles.

1 = Dieta saludable y programa de ejercicios

2 = Medicación

3 = Sin tratamiento

- Ajuste un modelo adecuado utilizando los datos anteriores. ¿Parece que el ejercicio y la dieta se pueden utilizar en forma eficaz para disminuir la presión sanguínea? Explique su respuesta a partir de los resultados.
- ¿El ejercicio y la dieta son una alternativa eficaz a la medicación?

(Sugerencia: Para responder a estas preguntas quizás usted desee construir el modelo en más de una forma).

Cambio en la presión sanguínea	Grupo	IMC
-32	1	27.3
-21	1	22.1
-26	1	26.1
-16	1	27.8
-11	2	19.2
-19	2	26.1
-23	2	28.6
-5	2	23.0
-6	3	28.1
5	3	25.3
-11	3	26.7
14	3	22.3

12.71 Demuestre que al elegir el llamado mejor modelo del subconjunto de entre una serie de posibles modelos, elegir el modelo con la menor s^2 equivale a escoger el modelo con el R_{ajus}^2 más pequeño.

12.72 Estudio de caso: Considere el conjunto de datos para el ejercicio 12.12 de la página 452 (datos de un hospital) que se repite a continuación.

Sitio	x_1	x_2	x_3	x_4	x_5	y
1	15.57	2463	472.92	18.0	4.45	566.52
2	44.02	2048	1339.75	9.5	6.92	696.82
3	20.42	3940	620.25	12.8	4.28	1033.15
4	18.74	6505	568.33	36.7	3.90	1003.62
5	49.20	5723	1497.60	35.7	5.50	1611.37
6	44.92	11,520	1365.83	24.0	4.60	1613.27
7	55.48	5779	1687.00	43.3	5.62	1854.17
8	59.28	5969	1639.92	46.7	5.15	2160.55
9	94.39	8461	2872.33	78.7	6.18	2305.58
10	128.02	20,106	3655.08	180.5	6.15	3503.93
11	96.00	13,313	2912.00	60.9	5.88	3571.59
12	131.42	10,771	3921.00	103.7	4.88	3741.40
13	127.21	15,543	3865.67	126.8	5.50	4026.52
14	252.90	36,194	7684.10	157.7	7.00	10,343.81
15	409.20	34,703	12,446.33	169.4	10.75	11,732.17
16	463.70	39,204	14,098.40	331.4	7.05	15,414.94
17	510.22	86,533	15,524.00	371.6	6.35	18,854.45

- a) Los listados de resultados de la función PROC REG del SAS que se presentan en las figuras 12.9 y 12.10 proporcionan una cantidad considerable de información. El propósito es detectar los valores extremos y, a final de cuentas, determinar cuáles términos del modelo deben utilizarse en la versión final.
- b) A menudo ocurre que el papel que desempeña una sola variable regresora no es evidente cuando se estudia en presencia de otras variables; esto se debe a la multicolinealidad. Con esto presente haga comentarios sobre la importancia de x_2 y x_3 en el modelo completo en comparación con su importancia en un modelo en el cual éstas son las únicas variables.
- c) Comente acerca de qué otros análisis se tendrían que hacer.
- d) Elabore análisis apropiados y escriba sus conclusiones respecto al modelo final.

Dependent Variable: y

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	490177488	98035498	237.79	<.0001
Error	11	4535052	412277		
Corrected Total	16	494712540			
	Root MSE	642.08838	R-Square	0.9908	
	Dependent Mean	4978.48000	Adj R-Sq	0.9867	
	Coeff Var	12.89728			

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	
Intercept	Intercept	1	1962.94816	1071.36170	1.83	0.0941	
x1	Average Daily Patient Load	1	-15.85167	97.65299	-0.16	0.8740	
x2	Monthly X-Ray Exposure	1	0.05593	0.02126	2.63	0.0234	
x3	Monthly Occupied Bed Days	1	1.58962	3.09208	0.51	0.6174	
x4	Eligible Population in the Area/100	1	-4.21867	7.17656	-0.59	0.5685	
x5	Average Length of Patients Stay in Days	1	-394.31412	209.63954	-1.88	0.0867	

Figura 12.9: Salida de resultados del SAS para el ejercicio de repaso 12.72; parte I.

	Dependent	Predicted	Std Error		95% CL Mean		95% CL Predict	
Obs	Variable	Value	Mean	Predict				
1	566.5200	775.0251	241.2323	244.0765	1306	-734.6494	2285	
2	696.8200	740.6702	331.1402	11.8355	1470	-849.4275	2331	
3	1033	1104	278.5116	490.9234	1717	-436.5244	2644	
4	1604	1240	268.1298	650.3459	1831	-291.0028	2772	
5	1611	1564	211.2372	1099	2029	76.6816	3052	
6	1613	2151	279.9293	1535	2767	609.5796	3693	
7	1854	1690	218.9976	1208	2172	196.5345	3183	
8	2161	1736	468.9903	703.9948	2768	-13.8306	3486	
9	2306	2737	290.4749	2098	3376	1186	4288	
10	3504	3682	585.2517	2394	4970	1770	5594	
11	3572	3239	189.0989	2823	3655	1766	4713	
12	3741	4353	328.8507	3630	5077	2766	5941	
13	4027	4257	314.0481	3566	4948	2684	5830	
14	10344	8768	252.2617	8213	9323	7249	10286	
15	11732	12237	573.9168	10974	13500	10342	14133	
16	15415	15038	585.7046	13749	16328	13126	16951	
17	18854	19321	599.9780	18000	20641	17387	21255	

		Std Error	Student					
Obs	Residual	Residual	Residual	-2	-1	0	1	2
1	-208.5051	595.0	-0.350					
2	-43.8502	550.1	-0.0797					
3	-70.7734	578.5	-0.122					
4	363.1244	583.4	0.622			*		
5	46.9483	606.3	0.0774					
6	-538.0017	577.9	-0.931		*			
7	164.4696	603.6	0.272					
8	424.3145	438.5	0.968			*		
9	-431.4090	572.6	-0.753		*			
10	-177.9234	264.1	-0.674		*			
11	332.6011	613.6	0.542			*		
12	-611.9330	551.5	-1.110		**			
13	-230.5684	560.0	-0.412					
14	1576	590.5	2.669			*****		
15	-504.8574	287.9	-1.753		***			
16	376.5491	263.1	1.431			**		
17	-466.2470	228.7	-2.039		****			

Figura 12.10: Salida de resultados del SAS para el ejercicio de repaso 12.72; parte II.

12.13 Posibles riesgos y errores conceptuales; relación con el material de otros capítulos

En este capítulo estudiamos varios procedimientos para usarlos en un “intento” por encontrar el mejor modelo. Sin embargo, una de las confusiones más importantes en el trabajo de los científicos e ingenieros novatos es que existe un **modelo lineal verdadero**, y que es posible encontrarlo. En la mayoría de fenómenos de la ciencia las relaciones entre las variables científicas son de naturaleza no lineal y se desconoce el modelo verdadero. Los modelos estadísticos lineales son **aproximaciones empíricas**.

En ocasiones, la decisión sobre cuál modelo adoptar depende de la información que se necesita obtener de éste. ¿Se usará para realizar predicciones? ¿Para explicar el papel que desempeña cada regresor? Esta “decisión” podría ser difícil ante la presencia de colinealidad. Es un hecho que para muchos problemas de regresión hay modelos múltiples con un desempeño muy similar. Para mayores detalles véase la referencia de Myers (1990).

Uno de los abusos más nocivos del material de este capítulo consiste en dar demasiada importancia a R^2 en la selección del llamado mejor modelo. Es importante recordar que para cualquier conjunto de datos se puede obtener una R^2 tan grande como se desee, dentro de la restricción de que $0 \leq R^2 \leq 1$. **Prestar demasiada atención a R^2 con frecuencia conduce a un sobreajuste.**

En este capítulo se dio mucha importancia a la detección de los valores extremos. Un clásico y grave abuso de la estadística radica en la decisión relacionada con la detección de los valores extremos. Esperamos que quede claro que el analista no debería por ningún motivo detectar los valores extremos, eliminarlos del conjunto de datos, ajustar un modelo nuevo, informar sobre los valores extremos, y así sucesivamente. Se trata de un procedimiento tentador y desastroso para llegar a un modelo que se ajuste bien a los datos, el cual conlleva a un ejemplo de **cómo mentir con estadísticos**. Si se detecta un valor extremo, lo correcto es revisar la historia de los datos en busca de posibles errores de captura o de procedimiento antes de eliminarlos del conjunto de datos. Se debe recordar que, por definición, un valor extremo es aquel para el cual el modelo no se ajusta bien. El problema podría no estar en los datos sino en la selección del modelo. Cambiar el modelo quizás haría que el punto no se detecte como un valor extremo.

Existen muchos tipos de respuestas que ocurren de forma natural en la práctica, pero que no se pueden utilizar en un análisis de mínimos cuadrados estándar porque sus supuestos de mínimos cuadrados clásicos no se cumplen. Los supuestos que suelen fallar son los de los errores normales y de la varianza homogénea. Por ejemplo, si la respuesta es una proporción, digamos la proporción de artículos defectuosos, la distribución de las respuestas se relaciona con la distribución binomial. Una segunda respuesta que ocurre con frecuencia en la práctica es la del conteo de Poisson. Evidentemente, la distribución no es normal, y la varianza de la respuesta, que es igual a la media de Poisson, varía de una observación a otra. Para conocer más detalles sobre estas condiciones poco ideales véase Myers y colaboradores (2008), citado en la bibliografía.

Capítulo 13

Experimentos con un solo factor: generales

13.1 Técnica del análisis de varianza

En el material sobre estimación y prueba de hipótesis que se cubrió en los capítulos 9 y 10 en cada caso nos limitamos a considerar sólo dos parámetros de la población. Ése fue el caso, por ejemplo, en la prueba de la igualdad de dos medias de la población, en la cual se usaron muestras independientes de poblaciones normales con varianza común pero desconocida, y en donde se necesitaba obtener un estimado agrupado de σ^2 .

El material que se refiere a las inferencias de dos muestras representa un caso especial de lo que se denomina *problema de un solo factor*. Por ejemplo, en el ejercicio 10.35 de la página 357 se midió el tiempo de supervivencia para dos muestras de ratones, en donde una muestra recibió un tratamiento de suero contra la leucemia y la otra no lo recibió. En este caso decimos que hay *un factor*, es decir, un *tratamiento*, y el factor se halla en *dos niveles*. Si en el proceso de muestreo se utilizaran varios tratamientos en competencia, se necesitarían más muestras de ratones. En ese caso el problema implicaría un factor con más de dos niveles, por lo tanto, con más de dos muestras.

En el problema de $k > 2$ muestras se supone que hay k muestras provenientes de k poblaciones. Un procedimiento muy común que se utiliza cuando se prueban medias de la población se denomina **análisis de varianza**, o **ANOVA**.

Si el lector ha estudiado el material acerca de la teoría de la regresión, el análisis de varianza no será, por supuesto, una técnica nueva para él. Utilizamos el método del análisis de varianza para partir la suma total de cuadrados en dos partes, una parte debida a la regresión y otra debida al error.

Suponga que en un experimento industrial a un ingeniero le interesa la forma en que la absorción media de humedad del concreto varía para 5 agregados de concreto diferentes. Las muestras se exponen a la humedad durante 48 horas y se decide que para cada agregado deben probarse 6 muestras, lo que hace que se requiera probar un total de 30 muestras. En la tabla 13.1 se presentan los datos registrados.

El modelo que se considera para esta situación es el siguiente. Se tomaron 6 observaciones de cada una de las 5 poblaciones, con medias $\mu_1, \mu_2, \dots, \mu_5$, respectivamente. Deseamos probar

$$H_0: \mu_1 = \mu_2 = \dots = \mu_5,$$

$$H_1: \text{Al menos dos de las medias no son iguales.}$$

Tabla 13.1: Absorción de humedad en agregados para concreto

Agregado:	1	2	3	4	5	
	551	595	639	417	563	
	457	580	615	449	631	
	450	508	511	517	522	
	731	583	573	438	613	
	499	633	648	415	656	
	632	517	677	555	679	
Total	3320	3416	3663	2791	3664	16,854
Media	553.33	569.33	610.50	465.17	610.67	561.80

Además, estamos interesados en realizar comparaciones individuales entre estas 5 medias de la población.

Dos fuentes de variabilidad en los datos

En el procedimiento del análisis de varianza se supone que cualquier variación que exista entre los promedios de los agregados se atribuye a 1) la variación en la absorción entre observaciones *dentro* de los tipos de agregados, y 2) la variación *entre* los tipos de agregados, es decir, a las diferencias en la composición química de los agregados. Por supuesto, la **variación dentro de los agregados** se debe a varias causas. Quizá las condiciones de temperatura y humedad no se mantuvieron constantes durante el experimento. Es posible que haya habido cierta cantidad de heterogeneidad en los lotes de materias primas que se usaron. En todo caso debe considerarse la variación dentro de la muestra como una **variación aleatoria o al azar**. Parte del objetivo del análisis de varianza consiste en determinar si las diferencias entre las 5 medias muestrales son lo que se esperaría debido sólo a la variación aleatoria o si, más bien, se trata de una variación más allá de los simples efectos del azar, como las diferencias en la composición química de los agregados.

En esta etapa surgen muchas preguntas acerca del problema anterior. Por ejemplo, ¿cuántas muestras deben probarse para cada agregado? Ésta es una pregunta que desafía continuamente al analista. Además, ¿qué pasaría si la variación dentro de la muestra fuera tan grande que al procedimiento estadístico le resultara difícil detectar las diferencias sistemáticas? ¿Es posible controlar de manera sistemática fuentes externas de variación y así eliminarlas de la parte que llamamos variación aleatoria? En las secciones siguientes intentaremos responder éstas y otras preguntas.

13.2 La estrategia del diseño de experimentos

En los capítulos 9 y 10 se estudiaron los conceptos de la estimación y la prueba de hipótesis para el caso de dos muestras, bajo la importante perspectiva de la manera en que se realiza el experimento. Esto forma parte de la categoría amplia de los diseños experimentales. Por ejemplo, para la **prueba *t* agrupada** que se estudió en el capítulo 10, se supone que los niveles de los factores (los tratamientos en el ejemplo de los ratones) se asignan al azar a las unidades experimentales (los ratones). En los capítulos 9 y 10 analizamos el

concepto de unidades experimentales y lo ilustramos por medio de varios ejemplos. En pocas palabras, las unidades experimentales son las unidades (ratones, pacientes, especímenes de concreto, tiempo) que **proporcionan la heterogeneidad que conduce al error experimental** en una investigación científica. La asignación aleatoria elimina el sesgo que podría originarse con una asignación sistemática. El objetivo consiste en distribuir en forma uniforme entre los niveles de los factores los riesgos que introduce la heterogeneidad de las unidades experimentales. Una asignación al azar simula mejor las condiciones que se asumen en el modelo. En la sección 13.7 analizamos la **formación de bloques** en los experimentos. En los capítulos 9 y 10 se presentó el concepto, cuando se efectuaron comparaciones entre las medias usando el **emparejamiento**, es decir, la división de las unidades experimentales en pares homogéneos denominados **bloques**. Entonces, los niveles de los factores o tratamientos se asignan al azar dentro de los bloques. El propósito de la formación de bloques es reducir el error experimental efectivo. En este capítulo se extiende de manera natural el emparejamiento a bloques de tamaño mayor, con el análisis de varianza como la herramienta analítica principal.

13.3 Análisis de varianza de un factor: diseño completamente aleatorizado (ANOVA de un factor)

De k poblaciones se seleccionan muestras aleatorias de tamaño n . Las k poblaciones diferentes se clasifican con base en un criterio único, como tratamientos o grupos distintos. En la actualidad el término **tratamiento** se utiliza por lo general para designar las diversas clasificaciones, ya sean diferentes agregados, analistas, fertilizadores o regiones del país.

Suposiciones e hipótesis del ANOVA de un solo factor

Se supone que las k poblaciones son independientes y que están distribuidas en forma normal con medias $\mu_1, \mu_2, \dots, \mu_k$, y varianza común σ^2 . Como se indicó en la sección 13.2, estas suposiciones son más aceptables mediante la aleatoriedad. Se desean obtener métodos adecuados para probar las hipótesis

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k,$$

$$H_1: \text{Al menos dos de las medias no son iguales.}$$

Sea que y_{ij} denote la j -ésima observación del i -ésimo tratamiento, y el acomodo de los datos es el que se observa en la tabla 13.2. Aquí, Y_i es el total de todas las observaciones de la muestra, del i -ésimo tratamiento, \bar{y}_i , es la media de todas las observaciones en la muestra del i -ésimo tratamiento, $Y_{..}$ es el total de todas las nk observaciones, y $\bar{y}_{..}$ es la media de todas las nk observaciones.

Modelo de ANOVA para un solo factor

Cada observación puede escribirse en la forma

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

donde ϵ_{ij} mide la desviación que tiene la observación j -ésima de la i -ésima muestra, con respecto de la media del tratamiento correspondiente. El término ϵ_{ij} representa el error aleatorio y desempeña el mismo papel que los términos del error en los modelos de

Tabla 13.2: k muestras aleatorias

Tratamiento:	1	2	...	i	...	k	
	y_{11}	y_{21}	...	y_{i1}	...	y_{k1}	
	y_{12}	y_{22}	...	y_{i2}	...	y_{k2}	
	\vdots	\vdots		\vdots		\vdots	
	y_{1n}	y_{2n}	...	y_{in}	...	y_{kn}	
Total	$Y_{1.}$	$Y_{2.}$...	$Y_{i.}$...	$Y_{k.}$	$Y_{..}$
Media	$\bar{y}_{1.}$	$\bar{y}_{2.}$...	$\bar{y}_{i.}$...	$\bar{y}_{k.}$	$\bar{y}_{..}$

regresión. Una forma alternativa y preferible de esta ecuación se obtiene sustituyendo $\mu_i = \mu + \alpha_i$, sujeta a la restricción $\sum_{i=1}^k \alpha_i = 0$. Por lo tanto, se escribe

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

donde μ tan sólo es la **media general** de todas las μ_i , es decir,

$$\mu = \frac{1}{k} \sum_{i=1}^k \mu_i,$$

y α_i se denomina el **efecto** del i -ésimo tratamiento.

La hipótesis nula de que k medias de la población son iguales, en comparación con la alternativa de que al menos dos de las medias son distintas, ahora se puede reemplazar por las hipótesis equivalentes.

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0,$$

$$H_1: \text{Al menos una de las } \alpha_i \text{ no es igual a cero.}$$

Resolución de la variabilidad total en componentes

Nuestra prueba se basará en una comparación de dos estimados independientes de la varianza poblacional común σ^2 . Dichos estimadores se obtendrán haciendo la partición de la variabilidad total de nuestros datos, denotados mediante la sumatoria doble

$$\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2,$$

en dos componentes.

Teorema 13.1: Identidad de la suma de cuadrados

$$\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

En lo que sigue, será conveniente identificar los términos de la identidad de la suma de cuadrados con la siguiente notación:

Tres medidas
importantes de
variabilidad

$$STC = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \text{suma total de cuadrados,}$$

$$SCT = n \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 = \text{suma de los cuadrados del tratamiento,}$$

$$SCE = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 = \text{suma de los cuadrados del error.}$$

Entonces, la identidad de la suma de los cuadrados se puede representar simbólicamente con la ecuación

$$STC = SCT + SCE$$

La identidad anterior expresa cómo las variaciones entre los tratamientos y dentro de los tratamientos contribuyen a la suma total de cuadrados. Sin embargo, se puede obtener mucha información si se investiga el **valor esperado tanto de SCT como de SCE**. Eventualmente calcularemos estimados de la varianza que determinan la razón que

Teorema 13.2:

$$E(SCT) = (k - 1)\sigma^2 + n \sum_{i=1}^k \alpha_i^2$$

La prueba del teorema se deja como ejercicio para el lector (véase el ejercicio 13.53 de la página 556).

Si H_0 es verdadera, un estimado de σ^2 basado en $k - 1$ grados de libertad es dado por la expresión:

Media cuadrática
del tratamiento

$$s_1^2 = \frac{SCT}{k - 1}$$

Si H_0 es verdadera y por ello cada α_i en el teorema 13.2 es igual a cero, se observa que

$$E\left(\frac{SCT}{k - 1}\right) = \sigma^2,$$

y s_1^2 es un estimado no sesgado de σ^2 . Sin embargo, si H_1 es verdadera, se tiene que

$$E\left(\frac{SCT}{k - 1}\right) = \sigma^2 + \frac{n}{k - 1} \sum_{i=1}^k \alpha_i^2,$$

y s_1^2 estima a σ^2 más un término adicional, que mide la variación debida a los efectos sistemáticos.

Otro estimado independiente de σ^2 , basado en $k(n - 1)$ grados de libertad, es la fórmula familiar:

Cuadrado medio
del error

$$s^2 = \frac{SCE}{k(n - 1)}$$

Resulta aleccionador puntualizar la importancia de los valores esperados de los cuadrados medios a los que recién nos referimos. En la sección siguiente se estudia el empleo de una **razón F** con el cuadrado medio del tratamiento en el numerador. Se observa que cuando H_1 es verdadera, la presencia de la condición $E(s_1^2) > E(s^2)$ sugiere que la razón F se utiliza en el contexto de una **prueba unilateral de cola superior**. Es decir, cuando H_1 es verdadera se esperaría que el numerador s_1^2 fuera mayor que el denominador.

Uso de la prueba F en el ANOVA

El estimado s^2 es no sesgado, independientemente de la veracidad o falsedad de la hipótesis nula (véase el ejercicio de repaso 13.52 de la página 556). Es importante señalar que la identidad de la suma de cuadrados ha hecho la partición no sólo de la variabilidad total de los datos, sino también del número total de grados de libertad. Es decir,

$$nk - 1 = k - 1 + k(n - 1).$$

Razón F para probar la igualdad de las medias

Cuando H_0 es verdadera, la razón $f = s_1^2/s^2$ es un valor de la variable aleatoria F , que tiene la distribución F con $k - 1$ y $k(n - 1)$ grados de libertad (véase el teorema 8.8). Como s_1^2 sobrestima a σ^2 cuando H_0 es falsa, se tiene una prueba de una cola con la región crítica localizada por completo en la cola derecha de la distribución.

A un nivel de significancia de α se rechaza la hipótesis nula H_0 cuando

$$f > f_\alpha[k - 1, k(n - 1)].$$

Otro método, el del valor P , sugiere que la evidencia a favor o en contra de H_0 es

$$P = P\{f[k - 1, k(n - 1)] > f\}.$$

Los cálculos para un problema de análisis de varianza por lo general se resumen en forma tabular, como se observa en la tabla 13.3.

Tabla 13.3: Análisis de varianza para el ANOVA de un solo factor

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	f calculada
Tratamientos	SCT	$k - 1$	$s_1^2 = \frac{SCT}{k - 1}$	$\frac{s_1^2}{s^2}$
Error	SCE	$k(n - 1)$	$s^2 = \frac{SCE}{k(n - 1)}$	
Total	STC	$kn - 1$		

Ejemplo 13.1: Pruebe la hipótesis de que $\mu_1 = \mu_2 = \dots = \mu_5$ a un nivel de significancia de 0.05 para los datos de la tabla 13.1 sobre la absorción de humedad por varios tipos de agregados para cemento.

Solución: Las hipótesis son

$H_0: \mu_1 = \mu_2 = \cdots = \mu_5,$
 $H_1: \text{Al menos dos de las medias no son iguales.}$
 $\alpha = 0.05$

Región crítica: $f > 2.76$ con $v_1 = 4$ y $v_2 = 25$ grados de libertad. Los cálculos de la suma de cuadrados proporcionan

$STC = 209,377, \quad SCT = 85,356,$
 $SCE = 209,377 - 85,356 = 124,021.$

En la figura 13.1 se presentan estos resultados y el resto de los cálculos del procedimiento ANOVA del programa SAS.

The GLM Procedure						
Dependent Variable: moisture						
		Sum of				
Source	DF	Squares	Mean Square	F Value	Pr > F	
Model	4	85356.4667	21339.1167	4.30	0.0088	
Error	25	124020.3333	4960.8133			
Corrected Total	29	209376.8000				
R-Square	Coeff Var	Root MSE	moisture Mean			
0.407669	12.53703	70.43304	561.8000			
Source	DF	Type I SS	Mean Square	F Value	Pr > F	
aggregate	4	85356.46667	21339.11667	4.30	0.0088	

Figura 13.1: Salida de resultados del programa SAS para el procedimiento de análisis de varianza.

Decisión: Rechazar H_0 y concluir que los agregados no tienen la misma media de absorción. El valor P para $f = 4.30$ es 0.0088, que es menor que 0.05. ▀

Además del ANOVA, se construyeron gráficas de caja para cada agregado, las cuales se presentan en la figura 13.2. Al observar las gráficas vemos que es evidente que no todos los agregados tienen la misma absorción. De hecho, parece que el agregado 4 destaca del resto. En el ejercicio 13.21 de la página 531 se incluye un análisis más formal que revela este resultado.

Durante el trabajo experimental es frecuente que se pierdan algunas de las observaciones deseadas. Los animales experimentales mueren, el material experimental se daña o los seres humanos abandonan el estudio. El análisis anterior para el mismo tamaño de la muestra aún es válido si modificamos ligeramente las fórmulas de la suma de cuadrados. Ahora suponemos que las k muestras aleatorias son de tamaño n_1, n_2, \dots, n_k , respectivamente.

Suma de cuadrados;
tamaños desiguales
de las muestras

$$STC = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2, \quad SCT = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y}_{..})^2, \quad SCE = STC - SCT$$

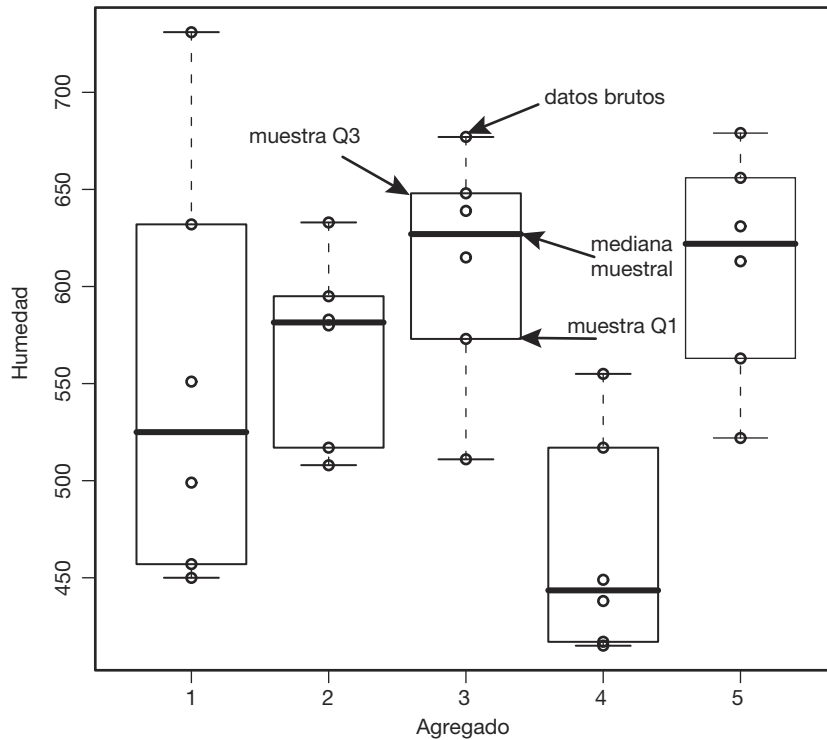


Figura 13.2: Gráficas de caja para la absorción de la humedad en agregados de concreto.

Después se hace la partición de los grados de libertad, como antes: $N - 1$ para STC , $k - 1$ para SCT y $N - 1 - (k - 1) = N - k$ para SCE , donde $N = \sum_{i=1}^k n_i$.

Ejemplo 13.2: Parte de un estudio realizado en Virginia Tech se diseñó para medir los niveles de actividad de la fosfatasa alcalina sérica (en unidades de Bessey-Lowry) en niños con trastornos convulsivos que recibían terapia de anticonvulsivantes bajo el cuidado de un médico privado. Se reclutaron 45 sujetos para el estudio y se clasificaron en cuatro grupos de medicamentos:

- G-1: Control (no recibieron anticonvulsivantes ni tenían historia de trastornos convulsivos)
- G-2: Fenobarbital
- G-3: Carbamazepina
- G-4: Otros anticonvulsivantes

De las muestras de sangre tomadas a cada sujeto se determinó el nivel de actividad de la fosfatasa alcalina sérica y se registró tal como se observa en la tabla 13.4. Pruebe la hipótesis de que, a un nivel de significancia de 0.05, el nivel promedio de actividad de la fosfatasa alcalina sérica es el mismo para los cuatro grupos de medicamentos.

Tabla 13.4: Nivel de actividad de la fosfatasa alcalina sérica

G-1		G-2	G-3	G-4
49.20	97.50	97.07	62.10	110.60
44.54	105.00	73.40	94.95	57.10
45.80	58.05	68.50	142.50	117.60
95.84	86.60	91.85	53.00	77.71
30.10	58.35	106.60	175.00	150.00
36.50	72.80	0.57	79.50	82.90
82.30	116.70	0.79	29.50	111.50
87.85	45.15	0.77	78.40	
105.00	70.35	0.81	127.50	
95.22	77.40			

Solución: A un nivel de significancia de 0.05, las hipótesis son

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4,$$

H_1 : Al menos dos de las medias no son iguales.

Región crítica: $f > 2.836$, al interpolar los valores de la tabla A.6.

Cálculos: $Y_1 = 1460.25$, $Y_2 = 440.36$, $Y_3 = 842.45$, $Y_4 = 707.41$ y $Y_{..} = 3450.47$. El análisis de varianza se incluye en la salida de resultados de *MINITAB* que se presenta en la figura 13.3.

One-way ANOVA: G-1, G-2, G-3, G-4					
Source	DF	SS	MS	F	P
Factor	3	13939	4646	3.57	0.022
Error	41	53376	1302		
Total	44	67315			
S = 36.08 R-Sq = 20.71% R-Sq(adj) = 14.90%					
Individual 95% CIs For Mean Based on Pooled StDev					
Level	N	Mean	StDev	--+-+-----+-----+-----+-----	
G-1	20	73.01	25.75	(----*-----)	
G-2	9	48.93	47.11	(------*-----)	
G-3	9	93.61	46.57	(------*-----)	
G-4	7	101.06	30.76	(------*-----)	
				--+-+-----+-----+-----+-----	
				30 60 90 120	
Pooled StDev = 36.08					

Figura 13.3: Análisis de *MINITAB* para los datos de la tabla 13.4.

Decisión: Rechazar H_0 y concluir que los niveles de actividad promedio de la fosfatasa alcalina sérica para los cuatro grupos de medicamentos no son los mismos. El valor calculado de P es 0.022. \blacksquare

Para concluir nuestro estudio del análisis de varianza para la clasificación de un solo factor mencionaremos las ventajas de elegir muestras del mismo tamaño en vez de otras de tamaños distintos. La primera ventaja es que la razón f no es sensible a pequeñas desviaciones de la suposición de varianzas iguales para las k poblaciones cuando las muestras son del mismo tamaño. La segunda consiste en que muestras del mismo tamaño minimizan la probabilidad de cometer un error tipo II.

13.4 Pruebas de la igualdad de varias varianzas

Aunque la razón f que se obtiene con el procedimiento del análisis de varianza no es sensible a las desviaciones de la suposición de varianzas iguales para las k poblaciones normales cuando las muestras son de igual tamaño, debe tenerse precaución y efectuar una prueba preliminar sobre la homogeneidad de las varianzas. En el caso de muestras de tamaños distintos, salta a la vista que es aconsejable realizar una prueba como ésta, si existe duda razonable acerca de la homogeneidad de las varianzas de la población. Por lo tanto, suponga que se desea probar la hipótesis nula

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

en comparación con la alternativa

$$H_1: \text{No todas las varianzas son iguales.}$$

La prueba que usaremos, denominada **prueba de Bartlett**, se basa en un estadístico cuya distribución muestral proporciona valores críticos exactos cuando los tamaños de las muestras son iguales. Dichos valores críticos para tamaños de las muestras iguales también se pueden utilizar para obtener aproximaciones muy exactas de los valores críticos para tamaños muestrales distintos.

En primer lugar calculamos las k varianzas muestrales $s_1^2, s_2^2, \dots, s_k^2$ a partir de muestras de tamaño n_1, n_2, \dots, n_k , con $\sum_{i=1}^k n_i = N$. En segundo lugar combinamos las varianzas muestrales para obtener la estimación agrupada

$$s_p^2 = \frac{1}{N - k} \sum_{i=1}^k (n_i - 1) s_i^2.$$

Ahora,

$$b = \frac{[(s_1^2)^{n_1-1} (s_2^2)^{n_2-1} \cdots (s_k^2)^{n_k-1}]^{1/(N-k)}}{s_p^2}$$

es un valor de una variable aleatoria B que tiene la **distribución de Bartlett**. Para el caso especial en que $n_1 = n_2 = \cdots = n_k = n$, se rechaza H_0 a un nivel de significancia α si

$$b < b_k(\alpha; n),$$

donde $b_k(\alpha; n)$ es el valor crítico que deja una área de tamaño α en el extremo izquierdo de la distribución de Bartlett. En la tabla A.10 se incluyen los valores críticos, $b_k(\alpha; n)$, para $\alpha = 0.01$ y 0.05 ; $k = 2, 3, \dots, 10$; y valores seleccionados de n , desde 3 hasta 100.

Cuando los tamaños de las muestras son distintos, se rechaza la hipótesis nula al nivel de significancia α si

$$b < b_k(\alpha; n_1, n_2, \dots, n_k),$$

donde

$$b_k(\alpha; n_1, n_2, \dots, n_k) \approx \frac{n_1 b_k(\alpha; n_1) + n_2 b_k(\alpha; n_2) + \dots + n_k b_k(\alpha; n_k)}{N}.$$

Igual que antes, todas las $b_k(\alpha; n_i)$ para los tamaños muestrales n_1, n_2, \dots, n_k se obtienen de la tabla A.10.

Ejemplo 13.3: Utilice la prueba de Bartlett a un nivel de significancia de 0.01 para probar la hipótesis de que las varianzas de la población de los cuatro grupos de medicamentos del ejemplo 13.2 son iguales.

Solución: Tenemos la hipótesis

$$\begin{aligned} H_0: \sigma_1^2 &= \sigma_2^2 = \sigma_3^2 = \sigma_4^2, \\ H_1: &\text{Las varianzas no son iguales,} \\ &\text{con } \alpha = 0.01. \end{aligned}$$

Región crítica: Si nos remitimos al ejemplo 13.2, tenemos que $n_1 = 20$, $n_2 = 9$, $n_3 = 9$, $n_4 = 7$, $N = 45$ y $k = 4$. Por lo tanto, se rechaza cuando

$$\begin{aligned} b &< b_4(0.01; 20, 9, 9, 7) \\ &\approx \frac{(20)(0.8586) + (9)(0.6892) + (9)(0.6892) + (7)(0.6045)}{45} \\ &= 0.7513. \end{aligned}$$

Cálculos: El primero se obtiene

$$s_1^2 = 662.862, s_2^2 = 2219.781, s_3^2 = 2168.434, s_4^2 = 946.032,$$

y después

$$\begin{aligned} s_p^2 &= \frac{(19)(662.862) + (8)(2219.781) + (8)(2168.434) + (6)(946.032)}{41} \\ &= 1301.861. \end{aligned}$$

Ahora,

$$b = \frac{[(662.862)^{19} (2219.781)^8 (2168.434)^8 (946.032)^6]^{1/41}}{1301.861} = 0.8557.$$

Decisión: no rechazar la hipótesis y concluir que las varianzas de la población de los cuatro grupos de medicamentos no son significativamente distintas. ■

Aunque la prueba de Bartlett se utiliza con mayor frecuencia para probar la homogeneidad de varianzas, se dispone de otros métodos. Un método creado por Cochran proporciona un procedimiento de cálculo sencillo, aunque está limitado a situaciones en

que los tamaños muestrales son iguales. La **prueba de Cochran** es especialmente útil para detectar si alguna de las varianzas es mucho mayor que las demás. El estadístico que se emplea es:

$$G = \frac{S_i^2 \text{ más grande}}{\sum_{i=1}^k S_i^2},$$

y se rechaza la hipótesis de igualdad de varianzas si $g > g_\alpha$, donde el valor de g_α se obtiene de la tabla A.11.

Para ilustrar la prueba de Cochran nos remitiremos otra vez a los datos de la tabla 13.1 sobre la absorción de humedad de los agregados para concreto. ¿Se justificó la suposición de varianzas iguales al realizar el análisis de varianza en el ejemplo 13.1? Se encontró que

$$s_1^2 = 12,134, \quad s_2^2 = 2303, \quad s_3^2 = 3594, \quad s_4^2 = 3319, \quad s_5^2 = 3455.$$

Por lo tanto,

$$g = \frac{12,134}{24,805} = 0.4892,$$

que no excede el valor de la tabla $g_{0.05} = 0.5065$. En consecuencia, se concluye que es razonable la suposición de que las varianzas son iguales.

Ejercicios

13.1 Se están considerando seis máquinas diferentes para la fabricación de sellos de goma y se están comparando con respecto a la resistencia a la tensión del producto. Se utiliza una muestra aleatoria de cuatro sellos hechos con cada máquina para determinar si la resistencia media a la tensión varía de una máquina a otra. A continuación se presentan las medidas de la resistencia a la tensión en kilogramos por centímetro cuadrado $\times 10^{-1}$:

Máquina					
1	2	3	4	5	6
17.5	16.4	20.3	14.6	17.5	18.3
16.9	19.2	15.7	16.7	19.2	16.2
15.8	17.7	17.8	20.8	16.5	17.5
18.6	15.4	18.9	18.9	20.5	20.1

Realice el análisis de varianza a un nivel de significancia de 0.05 e indique si la resistencia promedio a la tensión de las seis máquinas difiere o no de manera significativa.

13.2 Los datos que se presentan en la siguiente tabla representan el número de horas de alivio proporcionadas por cinco marcas diferentes de tabletas para el dolor de cabeza administradas a 25 sujetos que tenían fiebre de 38°C o más. Realice el análisis de varianza y, a un nivel de significancia de 0.05, pruebe la hipótesis de que las cinco marcas proporcionan el mismo número medio de horas de alivio. Analice los resultados.

Tabletas				
A	B	C	D	E
5.2	9.1	3.2	2.4	7.1
4.7	7.1	5.8	3.4	6.6
8.1	8.2	2.2	4.1	9.3
6.2	6.0	3.1	1.0	4.2
3.0	9.1	7.2	4.0	7.6

13.3 En el artículo “Shelf-Space Strategy in Retailing”, que se publicó en *Proceedings: Southern Marketing Association*, se investigó el efecto que tenía la altura de los anaqueles en los supermercados sobre las ventas de alimento enlatado para perro. Durante un periodo de 8 días se llevó a cabo un experimento en un supermercado pequeño acerca de las ventas de una marca de alimento para perro conocida como Arf y se utilizaron tres niveles de altura de anaqueles: a las rodillas, a la cintura y a los ojos. Cada día se cambió al azar tres veces la altura del anaqueles en la que estaba dicho alimento. Las secciones restantes de la góndola que contenía la marca dada se llenaban con una mezcla de marcas de comida canina, las cuales resultaban tanto familiares como desconocidas para los consumidores de esa área geográfica específica. Se presentan las ventas diarias, expresadas en cientos de dólares, del alimento Arf para las tres alturas del anaqueles. Con base en los datos, ¿existe una diferencia significativa en el promedio de ventas diarias de dicho alimento, con base en la altura del anaqueles? Utilice un nivel de significancia de 0.01.

Altura de anaquel		
Nivel de las rodillas	Nivel de la cintura	Nivel de los ojos
77	88	85
82	94	85
86	93	87
78	90	81
81	91	80
86	94	79
77	90	87
81	87	93

13.4 La inmovilización de los venados silvestres de cola blanca usando tranquilizantes da a los investigadores la oportunidad de estudiarlos de cerca y obtener información fisiológica valiosa. En el estudio denominado *Influence of Physical Restraint and Restraint Facilitating Drugs on Blood Measurements of White-Tailed Deer and Other Selected Mammals*, realizado en Virginia Tech, los biólogos de la vida silvestre probaron el tiempo del “derribamiento” (el periodo transcurrido entre la inyección y la inmovilización) de tres sustancias inmovilizadoras distintas. En este caso la inmovilización se define como el punto en que el animal ya no tiene control muscular suficiente para permanecer de pie. Se asignaron 30 venados machos de cola blanca al azar a cada uno de tres tratamientos. El grupo A recibió 5 miligramos de cloruro de succinilcolina líquida (SCC); al grupo B se le suministraron 8 miligramos de SCC en polvo; y al grupo C, 200 miligramos de hidrocloreto de fenciclidina. A continuación se presentan los tiempos de derribamiento, en minutos. Haga un análisis de varianza a un nivel de significancia de 0.01 y determine si el tiempo promedio de derribamiento es o no igual para las tres sustancias.

Grupo		
A	B	C
11	10	4
5	7	4
14	16	6
7	7	3
10	7	5
7	5	6
23	10	8
4	10	3
11	6	7
11	12	3

13.5 La enzima mitocondrial transhidrogenasa NADPH:NAD, de la tenia de la rata común (*Hymenolepis diminuta*) cataliza el hidrógeno en la transferencia de NADPH a NAD, lo que produce NADH. Se sabe que esta enzima desempeña un papel vital en el metabolismo anaerobio de la tenia, y recientemente se planteó la hipótesis de que podría servir como una bomba de intercambio de protones, es decir, para transferir protones a través de la membrana mitocondrial. Un estudio sobre el *Effect*

of Various Substrate Concentrations on the Conformational Variation of the NADPH:NAD Transhydrogenase of Hymenolepis diminuta llevado a cabo por la Bowling Green State University, se diseñó para evaluar la capacidad de dicha enzima para sufrir cambios en su conformación o su forma. Podría considerarse que los cambios en la actividad específica de la enzima ocasionados por las variaciones en la concentración de NADP sustentan la teoría del cambio de conformación. La enzima en cuestión se localiza en la membrana interior de las mitocondrias de la tenia. Se homogeneizaron las tenias y se aisló la enzima mediante una serie de centrifugaciones. Después se agregaron diferentes concentraciones de NADP a la solución de enzima aislada y la mezcla se incubó durante tres minutos en un baño de agua a 56°C. Luego, se analizó la enzima con un espectrómetro de rayo dual y se calcularon los resultados que se presentan a continuación, en términos de la actividad específica de la enzima, en nanomoles por minuto por miligramo de proteína. Pruebe la hipótesis de que la actividad específica promedio es la misma para las cuatro concentraciones, a un nivel de significancia de 0.01.

Concentración de NADP (nm)				
0	80	160	360	
11.01	11.38	11.02	6.04	10.31
12.09	10.67	10.67	8.65	8.30
10.55	12.33	11.50	7.76	9.48
11.26	10.08	10.31	10.13	8.89
			9.36	

13.6 Un estudio midió la tasa de sorción (ya sea absorción o adsorción) de tres tipos diferentes de solventes químicos orgánicos. Estos solventes se utilizan para limpiar partes industriales metálicas, y son desechos potencialmente riesgosos. Se probaron muestras independientes de solventes de cada tipo y se registraron sus tasas de sorción como un porcentaje molar. (Véase McClave, Dietrich y Sincich, 1997).

Aromáticos		Cloroalcalinos		Ésteres		
1.06	0.95	1.58	1.12	0.29	0.43	0.06
0.79	0.65	1.45	0.91	0.06	0.51	0.09
0.82	1.15	0.57	0.83	0.44	0.10	0.17
0.89	1.12	1.16	0.43	0.55	0.53	0.17
1.05				0.61	0.34	0.60

¿Existe una diferencia significativa en la tasa promedio de sorción de los tres solventes? Utilice un valor *P* para sus conclusiones. ¿Qué solvente usaría?

13.7 Se ha demostrado que el fertilizante fosfato amoniacal de magnesio, MgNH_4PO_4 , es un proveedor eficaz de los nutrientes necesarios para el crecimiento de las plantas. Los compuestos que suministra son muy solubles en agua, lo cual permite su aplicación directa sobre la superficie del suelo o que se mezcle con el sustrato de crecimiento durante el proceso de encapsu-

lamiento. Se efectuó un estudio denominado *Effect of Magnesium Ammonium Phosphate on Height of Chrysanthemums* en George Mason University para determinar el nivel óptimo posible de la fertilización con base en la mejoría de la respuesta de crecimiento vertical del crisantemo. Se dividieron 40 semillas de crisantemo en 4 grupos de diez plantas cada uno. Se sembró cada una en una maceta similar que contenía un medio uniforme de crecimiento. Se agregó a cada grupo de plantas una concentración cada vez mayor de MgNH_4PO_4 , medido en gramos por bushel. Los cuatro grupos de plantas se cultivaron durante cuatro semanas en condiciones uniformes en un invernadero. A continuación se presentan los tratamientos y los cambios respectivos de sus alturas, medidas en centímetros:

Tratamiento							
50 g/bu		100 g/bu		200 g/bu		400 g/bu	
13.2	12.4	16.0	12.6	7.8	14.4	21.0	14.8
12.8	17.2	14.8	13.0	20.0	15.8	19.1	15.8
13.0	14.0	14.0	23.6	17.0	27.0	18.0	26.0
14.2	21.6	14.0	17.0	19.6	18.0	21.1	22.0
15.0	20.0	22.2	24.4	20.2	23.2	25.0	18.2

A un nivel de significancia de 0.05, ¿podría concluirse que concentraciones diferentes de MgNH_4PO_4 afectan la altura promedio que alcanzan los crisantemos? ¿Qué cantidad del fertilizante parece ser la mejor?

13.8 Para el conjunto de datos del ejercicio 13.7 use la prueba de Bartlett para probar si las varianzas son iguales. Utilice $\alpha = 0.05$.

13.9 Utilice la prueba de Bartlett a un nivel de significancia de 0.01 para probar la homogeneidad de las varianzas en el ejercicio 13.5 de la página 519.

13.10 Utilice la prueba de Cochran a un nivel de significancia de 0.01 para probar la homogeneidad de las varianzas en el ejercicio 13.4 de la página 519.

13.11 Utilice la prueba de Bartlett a un nivel de significancia de 0.05 para probar la homogeneidad de las varianzas en el ejercicio 13.6 de la página 519.

13.5 Comparaciones de un grado de libertad

El análisis de varianza en la clasificación de un solo factor, o experimento de un solo factor, como se le denomina con frecuencia, tan sólo indica si puede rechazarse o no la hipótesis de medias de tratamientos iguales. Por lo general, el experimentador preferiría efectuar un análisis más profundo. Como ilustración, en el ejemplo 13.1, mediante el rechazo de la hipótesis nula, concluimos que las medias no son iguales, pero aún no sabemos en dónde residen las diferencias entre los agregados. Es probable que el ingeniero intuya de antemano que los agregados 1 y 2 deberían poseer propiedades similares de absorción, al igual que los agregados 3 y 5. Sin embargo, sería interesante estudiar las diferencias entre los dos grupos. Así, parece apropiado probar las hipótesis

$$\begin{aligned}H_0: \mu_1 + \mu_2 - \mu_3 - \mu_5 &= 0, \\H_1: \mu_1 + \mu_2 - \mu_3 - \mu_5 &\neq 0.\end{aligned}$$

Se observa que la hipótesis es una función lineal de las medias de la población, en las cuales los coeficientes suman cero.

Definición 13.1: Cualquier función lineal de la forma

$$\omega = \sum_{i=1}^k c_i \mu_i,$$

donde $\sum_{i=1}^k c_i = 0$ se llama **comparación** o **contraste** en las medias de los tratamientos.

Con frecuencia el experimentador puede hacer comparaciones múltiples probando la significancia de los contrastes de las medias de los tratamientos, es decir, probando una hipótesis del siguiente tipo:

Hipótesis para un
contraste

$$H_0: \sum_{i=1}^k c_i \mu_i = 0,$$

$$H_1: \sum_{i=1}^k c_i \mu_i \neq 0,$$

donde $\sum_{i=1}^k c_i = 0$.

La prueba se efectúa calculando primero un contraste similar de las medias de las muestras,

$$w = \sum_{i=1}^k c_i \bar{Y}_i.$$

Como $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_k$ son variables aleatorias independientes que tienen distribuciones normales con medias $\mu_1, \mu_2, \dots, \mu_k$ y varianzas $\sigma_1^2/n_1, \sigma_2^2/n_2, \dots, \sigma_k^2/n_k$, respectivamente, el teorema 7.11 nos garantiza que w es un valor de la variable aleatoria normal W con

$$\text{media } \mu_W = \sum_{i=1}^k c_i \mu_i \text{ y varianza } \sigma_W^2 = \sigma^2 \sum_{i=1}^k \frac{c_i^2}{n_i}.$$

Por lo tanto, cuando H_0 es verdadera, $\mu_W = 0$ y, según el ejemplo 7.5, el estadístico

$$\frac{W^2}{\sigma_W^2} = \frac{\left(\sum_{i=1}^k c_i \bar{Y}_i \right)^2}{\sigma^2 \sum_{i=1}^k (c_i^2/n_i)}$$

se distribuye como una variable aleatoria chi cuadrada con 1 grado de libertad.

Estadístico de
prueba para
probar un
contraste

Nuestra hipótesis se prueba a un nivel de significancia α calculando

$$f = \frac{\left(\sum_{i=1}^k c_i \bar{Y}_i \right)^2}{s^2 \sum_{i=1}^k (c_i^2/n_i)} = \frac{\left[\sum_{i=1}^k (c_i Y_i / n_i) \right]^2}{s^2 \sum_{i=1}^k (c_i^2/n_i)} = \frac{SC_W}{s^2}.$$

Aquí f es un valor de la variable aleatoria F que tiene distribución F con 1 y $N - k$ grados de libertad.

Cuando los tamaños de las muestras son iguales a n ,

$$SC_W = \frac{\left(\sum_{i=1}^k c_i Y_i \right)^2}{n \sum_{i=1}^k c_i^2}.$$

La cantidad SC_W , que se denomina **suma de cuadrados de los contrastes**, indica la parte de la SCT que se explica por el contraste en cuestión.

Esta suma de cuadrados se empleará para probar la hipótesis de que

$$\sum_{i=1}^k c_i \mu_i = 0.$$

Con frecuencia es de interés probar contrastes múltiples, en particular contrastes que son linealmente independientes u ortogonales. Como resultado, se vuelve necesaria la siguiente definición:

Definición 13.2: Se dice que los dos contrastes

$$\omega_1 = \sum_{i=1}^k b_i \mu_i \quad \text{y} \quad \omega_2 = \sum_{i=1}^k c_i \mu_i$$

son **ortogonales**, si $\sum_{i=1}^k b_i c_i / n_i = 0$, o bien, cuando las n_i son iguales a n , si

$$\sum_{i=1}^k b_i c_i = 0.$$

Si ω_1 y ω_2 son ortogonales, entonces las cantidades SC_{ω_1} y SC_{ω_2} son componentes de SCT , cada una con un solo grado de libertad. La suma de cuadrados de los tratamientos con $k - 1$ grados de libertad se puede dividir en, a lo sumo, $k - 1$ sumas de cuadrados de contrastes independientes con un solo grado de libertad que satisfacen la identidad

$$SCT = SC_{\omega_1} + SC_{\omega_2} + \cdots + SC_{\omega_{k-1}},$$

si los contrastes son ortogonales entre sí.

Ejemplo 13.4: Remítase al ejemplo 13.1 y calcule la suma de cuadrados de los contrastes que corresponden a los contrastes ortogonales

$$\omega_1 = \mu_1 + \mu_2 - \mu_3 - \mu_5, \quad \omega_2 = \mu_1 + \mu_2 + \mu_3 - 4\mu_4 + \mu_5,$$

y efectúe las pruebas de significancia adecuadas. En este caso es de interés *a priori* comparar los dos grupos (1, 2) y (3, 5). Un contraste importante e independiente consiste en realizar la comparación entre el conjunto de agregados (1, 2, 3, 5) y el agregado 4.

Solución: Es evidente que los dos contrastes son ortogonales, puesto que

$$(1)(1) + (1)(1) + (-1)(1) + (0)(-4) + (-1)(1) = 0.$$

El segundo contraste indica una comparación entre los agregados (1, 2, 3 y 5) y el agregado 4. Podemos escribir dos contrastes adicionales ortogonales a los dos primeros, es decir:

$$\omega_3 = \mu_1 - \mu_2 \text{ (agregado 1 contra agregado 2),}$$

$$\omega_4 = \mu_3 - \mu_5 \text{ (agregado 3 contra agregado 5).}$$

De los datos de la tabla 13.1, se tiene que

$$SC_{w_1} = \frac{(3320 + 3416 - 3663 - 3664)^2}{6[(1)^2 + (1)^2 + (-1)^2 + (-1)^2]} = 14,553,$$

$$SC_{w_2} = \frac{[3320 + 3416 + 3663 + 3664 - 4(2791)]^2}{6[(1)^2 + (1)^2 + (1)^2 + (1)^2 + (-4)^2]} = 70,035.$$

En la tabla 13.5 se presenta un análisis de varianza más extenso. Se observa que las dos sumas de cuadrados de los contrastes explican casi toda la suma de cuadrados de los agregados. Existe una diferencia significativa entre las propiedades de absorción de los agregados, y el contraste ω_1 es significativo marginalmente. Sin embargo, el valor f de 14.12 para ω_2 es muy significativo, y se rechaza la hipótesis

$$H_0: \mu_1 + \mu_2 + \mu_3 + \mu_5 = 4\mu_4$$

Tabla 13.5: Análisis de varianza usando contrastes ortogonales

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Agregados	85,356	4	21,339	4.30
(1, 2) vs. (3,5)	14,553	1	14,553	2.93
(1, 2, 3, 5) vs. 4	70,035	1	70,035	14.12
Error	124,021	25	4961	
Total	209,377	29		

Los contrastes ortogonales permiten al profesional dividir la variación del tratamiento en componentes independientes. Por lo general el experimentador tiene interés en hacer ciertos contrastes. Eso ocurrió en nuestro ejemplo, donde las consideraciones *a priori* sugerían que los agregados (1, 2) y (3, 5) constituirían grupos distintos con propiedades diferentes de absorción, un planteamiento que no obtuvo mucho respaldo con la prueba de significancia. Sin embargo, la segunda comparación apoyó la conclusión de que el agregado 4 parecía “destacar” de los demás. En este caso no fue necesaria la partición completa de *SCT*, dado que dos de las cuatro comparaciones independientes posibles explicaban la mayor parte de la variación en los tratamientos.

En la figura 13.4 se presenta un procedimiento GLM del programa SAS, que proporciona un conjunto completo de contrastes ortogonales. Observe que la suma de cuadrados de los cuatro contrastes se agrega a la suma de cuadrados de los agregados. Asimismo, note que los últimos dos contrastes (1 contra 2, 3 contra 5) revelan comparaciones insignificantes. ■

13.6 Comparaciones múltiples

El análisis de varianza es un procedimiento poderoso para probar la homogeneidad de un conjunto de medias. No obstante, si se rechazara la hipótesis nula y se aceptara la alternativa que se planteó (que no todas las medias son iguales), aún no se sabría cuáles de las medias de la población son iguales y cuáles son diferentes.

The GLM Procedure						
Dependent Variable: moisture						
		Sum of				
Source	DF	Squares	Mean Square	F Value	Pr > F	
Model	4	85356.4667	21339.1167	4.30	0.0088	
Error	25	124020.3333	4960.8133			
Corrected Total	29	209376.8000				
R-Square		Coeff Var	Root MSE	moisture Mean		
0.407669		12.53703	70.43304	561.8000		
Source	DF	Type I SS	Mean Square	F Value	Pr > F	
aggregate	4	85356.46667	21339.11667	4.30	0.0088	
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
aggregate	4	85356.46667	21339.11667	4.30	0.0088	
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F	
(1,2,3,5) vs. 4	1	70035.00833	70035.00833	14.12	0.0009	
(1,2) vs. (3,5)	1	14553.37500	14553.37500	2.93	0.0991	
1 vs. 2	1	768.00000	768.00000	0.15	0.6973	
3 vs. 5	1	0.08333	0.08333	0.00	0.9968	

Figura 13.4: Un conjunto de contrastes ortogonales.

A menudo es de interés efectuar varias **comparaciones por pares** (quizá todas las que sean posibles) entre los tratamientos. En realidad, una comparación por pares se puede ver como un contraste simple, es decir, una prueba de

$$H_0: \mu_i - \mu_j = 0,$$

$$H_1: \mu_i - \mu_j \neq 0,$$

para toda $i \neq j$. Hacer todas las comparaciones posibles por pares entre las medias puede ser muy benéfico cuando no se conocen *a priori* contrastes complejos particulares. Por ejemplo, suponga que se desea probar las hipótesis siguientes, con los datos de los agregados de la tabla 13.1:

$$H_0: \mu_1 - \mu_5 = 0,$$

$$H_1: \mu_1 - \mu_5 \neq 0.$$

La prueba se desarrolla usando una F , una t , o el método de los intervalos de confianza. Si se usa la t , se tiene que

$$t = \frac{\bar{y}_1 - \bar{y}_5}{s\sqrt{2/n}},$$

donde s es la raíz cuadrada del cuadrado medio del error y $n = 6$ es el tamaño de la muestra por tratamiento. En este caso,

$$t = \frac{553.33 - 610.67}{\sqrt{4961}\sqrt{1/3}} = -1.41.$$

El valor P para la prueba t con 25 grados de libertad es 0.17. Así que no hay evidencia suficiente para rechazar H_0 .

Relación entre T y F

Anteriormente se expuso el uso de una prueba t agrupada, junto con los lineamientos que se estudiaron en el capítulo 10. El estimado agrupado se tomó del cuadrado medio del error con el fin de aprovechar los grados de libertad que están agrupados en las cinco muestras. Además, probamos un contraste. El lector debería observar que si el valor t se eleva al cuadrado, el resultado tiene exactamente la misma forma que el valor de f para una prueba del contraste, analizada en la sección anterior. En efecto,

$$f = \frac{(\bar{y}_1 - \bar{y}_5)^2}{s^2(1/6 + 1/6)} = \frac{(553.33 - 610.67)^2}{4961(1/3)} = 1.988,$$

que es, por supuesto, t^2 .

Método del intervalo de confianza para una comparación por pares

Es fácil resolver el mismo problema de una comparación por pares (o un contraste) usando el método del intervalo de confianza. Es claro que, si se calcula un intervalo de confianza del $100(1 - \alpha)\%$ sobre $\mu_1 - \mu_5$, se tiene que

$$\bar{y}_1 - \bar{y}_5 \pm t_{\alpha/2} s \sqrt{\frac{2}{6}},$$

donde $t_{\alpha/2}$ es el punto superior de $100(1 - \alpha/2)\%$ de una distribución t con 25 grados de libertad (grados de libertad que provienen de s^2). Esta conexión inmediata entre las pruebas de hipótesis y los intervalos de confianza debería ser evidente a partir de los análisis de los capítulos 9 y 10. La prueba de un contraste simple $\mu_1 - \mu_5$ no implica más que observar si el intervalo de confianza anterior cubre o no al cero. Al sustituir los números se tiene que el intervalo de confianza de 95%:

$$(553.33 - 610.67) \pm 2.060 \sqrt{4961} \sqrt{\frac{1}{3}} = -57.34 \pm 83.77.$$

Por consiguiente, como el intervalo de confianza cubre al cero, el contraste no es significativo. En otras palabras, no hay diferencia significativa entre las medias de los agregados 1 y 5.

Tasa de error por experimento

Se presentan muchas dificultades cuando el analista intenta hacer muchas o todas las comparaciones por pares posibles. Para el caso de k medias habrá, desde luego, $r = k(k - 1)/2$ comparaciones por pares posibles. Si se suponen comparaciones independientes, la tasa de error por experimento o tasa de error por familia, es decir, la probabilidad de un falso rechazo de al menos una de las hipótesis, es dada por $1 - (1 - \alpha)^r$, donde α es la probabilidad seleccionada del error tipo I para una comparación específica. Es claro que esta medida del error tipo I por experimento sería bastante grande. Por

ejemplo, aun si sólo hubiera 6 comparaciones, digamos, en el caso de 4 medias, y $\alpha = 0.05$, la tasa de experimento-juicio sería

$$1 - (0.95)^6 \approx 0.26.$$

Cuando se prueban muchas comparaciones por pares, por lo general existe la necesidad de hacer el contraste efectivo sobre una sola comparación más conservadora. Es decir, usando el método del intervalo de confianza, los intervalos de confianza serían mucho más anchos que $\pm t_{\alpha/2} s \sqrt{2/n}$ que se emplea para el caso de una sola comparación.

Prueba de Tukey

Hay varios métodos estándar para realizar comparaciones por pares que den credibilidad a la tasa del error tipo I. Aquí se analizarán e ilustrarán dos de ellos. El primero, denominado **procedimiento de Tukey**, permite la formación de intervalos de confianza del $100(1 - \alpha)\%$ simultáneos para todas las comparaciones por pares. El método se basa en la distribución del rango *estudentizado*. El punto apropiado del percentil es una función de α , k y $v =$ grados de libertad para s^2 . En la tabla A.12 se presenta una lista de puntos porcentuales superiores adecuados para $\alpha = 0.05$. El método de Tukey de comparaciones por pares implica encontrar una diferencia significativa entre las medias i y j ($i \neq j$) si $|\bar{y}_i - \bar{y}_j|$ excede a $q(\alpha, k, v) \sqrt{\frac{s^2}{n}}$.

El procedimiento de Tukey se ilustra con facilidad. Considere un ejemplo hipotético en el que se tienen 6 tratamientos en un diseño completamente aleatorizado de un solo factor, en el que se hacen 5 observaciones por tratamiento. Suponga que el cuadrado medio del error tomado de la tabla del análisis de varianza es $s^2 = 2.45$ (24 grados de libertad). Las medias muestrales están en orden ascendente,

$$\begin{array}{cccccc} \bar{y}_2 & \bar{y}_5 & \bar{y}_1 & \bar{y}_3 & \bar{y}_6 & \bar{y}_4 \\ 14.50 & 16.75 & 19.84 & 21.12 & 22.90 & 23.20. \end{array}$$

Con $\alpha = 0.05$, el valor de $q(0.05, 6, 24)$ es 4.37. Así, todas las diferencias absolutas se comparan con

$$4.37 \sqrt{\frac{2.45}{5}} = 3.059.$$

Como resultado, las siguientes representan medias que, usando el procedimiento de Tukey, se encuentra que son significativamente diferentes:

$$\begin{array}{cccccc} 4 \text{ y } 1, & 4 \text{ y } 5, & 4 \text{ y } 2, & 6 \text{ y } 1, & 6 \text{ y } 5, \\ 6 \text{ y } 2, & 3 \text{ y } 5, & 3 \text{ y } 2, & 1 \text{ y } 5, & 1 \text{ y } 2. \end{array}$$

¿De dónde proviene el nivel α en la prueba de Tukey?

Se mencionó brevemente el concepto de **intervalos de confianza simultáneos** que se emplean para el procedimiento de Tukey. El lector obtendrá una perspectiva útil del concepto de comparaciones múltiples, si comprende el significado de los intervalos de confianza simultáneos.

En el capítulo 9 vimos que, si se calcula un intervalo de confianza de 95% para, digamos, una media μ , entonces la probabilidad de que el intervalo cubra la media verdadera μ es 0.95.

Sin embargo, como vimos antes, para el caso de comparaciones múltiples la probabilidad efectiva de interés está ligada con la tasa de error por experimento, y debe hacerse énfasis en que los intervalos de confianza del tipo $\bar{y}_i - \bar{y}_j \pm q(\alpha, k, v)s\sqrt{1/n}$ no son independientes, ya que todos implican a s y muchos utilizan los mismos promedios, las \bar{y}_i . A pesar de tales dificultades, si se utiliza la $q(0.05, k, v)$, el nivel de confianza simultáneo está controlado en un 95%. Lo mismo es cierto para $q(0.01, k, v)$, es decir, el nivel de confianza está controlado en un 99%. En el caso de $\alpha = 0.05$, hay una probabilidad de 0.05 de que se encuentre falsamente que al menos un par de mediciones son diferentes (falso rechazo de al menos una hipótesis nula). En el caso de $\alpha = 0.01$, la probabilidad correspondiente será 0.01.

Prueba de Duncan

El segundo procedimiento que se estudiará se denomina **procedimiento de Duncan o prueba de Duncan de rango múltiple**. Este procedimiento también se basa en el concepto general del rango estudentizado. El rango de cualquier subconjunto de p medias muestrales debe exceder cierto valor antes de que se encuentre que cualquiera de las p medias es diferente. Este valor recibe el nombre de **rango de menor significancia** para las p medias, y se denota por R_p , donde

$$R_p = r_p \sqrt{\frac{s^2}{n}}.$$

Los valores de la cantidad r_p , llamados **rango estudentizado de menor significancia**, dependen del nivel de significancia deseado y del número de grados de libertad del cuadrado medio del error. Estos valores se obtienen de la tabla A.13 para $p = 2, 3, \dots, 10$ medias.

Para ilustrar el procedimiento de prueba de rango múltiple, consideremos el ejemplo hipotético en el cual se comparan 6 tratamientos con 5 observaciones por tratamiento. Se trata del mismo ejemplo que se empleó para ilustrar la prueba de Tukey. Se obtiene R_p multiplicando cada r_p por 0.70. Los resultados de estos cálculos se resumen como sigue:

p	2	3	4	5	6
r_p	2.919	3.066	3.160	3.226	3.276
R_p	2.043	2.146	2.212	2.258	2.293

Si se comparan estos rangos de menor significancia con las diferencias en medias ordenadas, se llega a las conclusiones siguientes:

1. Como $\bar{y}_4 - \bar{y}_2 = 8.70 > R_6 = 2.293$, se concluye que μ_4 y μ_2 son significativamente distintas.
2. Si se comparan $\bar{y}_4 - \bar{y}_5$ y $\bar{y}_6 - \bar{y}_2$ con R_5 , se concluye que μ_4 es significativamente mayor que μ_5 y que μ_6 es significativamente mayor que μ_2 .
3. Si se comparan $\bar{y}_4 - \bar{y}_1$, $\bar{y}_6 - \bar{y}_5$ y $\bar{y}_3 - \bar{y}_2$ con R_4 , se concluye que cada diferencia es significativa.
4. Si se comparan $\bar{y}_4 - \bar{y}_3$, $\bar{y}_6 - \bar{y}_1$, $\bar{y}_3 - \bar{y}_5$ y $\bar{y}_1 - \bar{y}_2$ con R_3 , se encuentra que todas las diferencias son significativas excepto para $\mu_4 - \mu_3$. Por lo tanto, μ_3 , μ_4 y μ_6 constituyen un subconjunto de medias homogéneas.
5. Si se comparan $\bar{y}_3 - \bar{y}_1$, $\bar{y}_1 - \bar{y}_5$ y $\bar{y}_5 - \bar{y}_2$ con R_2 , se concluye que sólo μ_3 y μ_1 no son significativamente distintas.

Se acostumbra resumir las conclusiones anteriores con el dibujo de una línea debajo de cualquier subconjunto de medias adyacentes que no sean significativamente diferentes. Así, tenemos

$$\begin{array}{cccccc} \bar{y}_2. & \bar{y}_5. & \bar{y}_1. & \bar{y}_3. & \bar{y}_6. & \bar{y}_4. \\ \hline 14.50 & 16.75 & 19.84 & 21.12 & 22.90 & 23.20 \end{array}$$

Es evidente que en este caso los resultados con los procedimientos de Tukey y Duncan son muy similares. El procedimiento de Tukey no detectó ninguna diferencia entre 2 y 5; mientras que el de Duncan sí lo hizo.

Prueba de Dunnett: comparación de tratamientos con un control

En muchos problemas científicos y de ingeniería no nos interesa hacer inferencias acerca de todas las comparaciones posibles entre las medias de los tratamientos del tipo $\mu_i - \mu_j$. En vez de ello es frecuente que el experimento dicte la necesidad de comparar simultáneamente cada *tratamiento* con un *control*. Un procedimiento de prueba desarrollado por C. W. Dunnett determina diferencias significativas entre cada media de tratamiento y el control, con un solo nivel conjunto de significancia α . Para ilustrar el procedimiento de Dunnett, se considerarán los datos experimentales de la tabla 13.6 para la clasificación de un solo factor, donde se estudió el efecto de tres catalizadores sobre el producto de una reacción. Como control se emplea un cuarto tratamiento en el que no se aplica un catalizador.

Tabla 13.6: Producto de una reacción

Control	Catalizador 1	Catalizador 2	Catalizador 3
50.7	54.1	52.7	51.2
51.5	53.8	53.9	50.8
49.2	53.1	57.0	49.7
53.1	52.5	54.1	48.0
52.7	54.0	52.5	47.2
$\bar{y}_0. = 51.44$	$\bar{y}_1. = 53.50$	$\bar{y}_2. = 54.04$	$\bar{y}_3. = 49.38$

En general, se desea probar las k hipótesis

$$\left. \begin{array}{l} H_0: \mu_0 = \mu_i \\ H_1: \mu_0 \neq \mu_i \end{array} \right\} \quad i = 1, 2, \dots, k,$$

donde μ_0 representa el producto medio para la población de medidas en que se utiliza el control. Como se mencionó en la sección 13.3, se espera que las suposiciones habituales del análisis de varianza sigan siendo válidas. Para probar la hipótesis nula especificada con H_0 en comparación con alternativas bilaterales para una situación experimental donde existen k tratamientos, sin incluir el control, y n observaciones por tratamiento, primero calculamos los valores

$$d_i = \frac{\bar{y}_i. - \bar{y}_0.}{\sqrt{2s^2/n}}, \quad i = 1, 2, \dots, k.$$

Como antes, la varianza muestral s^2 se obtiene a partir del cuadrado medio del error en el análisis de varianza. Ahora bien, la región crítica para rechazar H_0 a un nivel de significancia α , se establece con la desigualdad

$$|d_i| > d_{\alpha/2}(k, \nu),$$

donde ν es el número de grados de libertad para el cuadrado medio del error. Los valores de la cantidad $d_{\alpha/2}(k, \nu)$ para una prueba de dos colas se incluyen en la tabla A.14 para $\alpha = 0.05$ y $\alpha = 0.01$, para diversos valores de k y ν .

Ejemplo 13.5: Para los datos de la tabla 13.6, pruebe la hipótesis que compara cada catalizador con el control, usando alternativas bilaterales. Como nivel de significancia conjunto elija $\alpha = 0.05$.

Solución: El cuadrado medio del error con 16 grados de libertad se obtiene de la tabla de análisis de varianza, usando todos los tratamientos $k + 1$. El cuadrado medio del error es dado por

$$s^2 = \frac{36.812}{16} = 2.30075 \text{ y } \sqrt{\frac{2s^2}{n}} = \sqrt{\frac{(2)(2.30075)}{5}} = 0.9593.$$

Entonces,

$$d_1 = \frac{53.50 - 51.44}{0.9593} = 2.147, \quad d_2 = \frac{54.04 - 51.44}{0.9593} = 2.710,$$

$$d_3 = \frac{49.38 - 51.44}{0.9593} = -2.147.$$

De la tabla A.14 el valor crítico para $\alpha = 0.05$ resulta ser $d_{0.025}(3, 16) = 2.59$. Como $|d_1| < 2.59$ y $|d_3| < 2.59$, se concluye que tan sólo la producción media para el catalizador 2 es significativamente diferente de la respuesta media de la reacción utilizando el control. ■

Muchas aplicaciones prácticas imponen la necesidad de una prueba de una cola para comparar los tratamientos con un control. En efecto, cuando un farmacólogo está interesado en el efecto de varias dosis de un medicamento sobre el nivel del colesterol, y su control consiste en una dosis de cero, sería interesante determinar si cada dosis produce una reducción significativamente mayor que la del control. En la tabla A.15 se presentan los valores críticos de $d_{\alpha}(k, \nu)$ para alternativas unilaterales.

Ejercicios

13.12 Considere los datos del ejercicio de repaso 13.45 de la página 555. Efectúe pruebas de significancia sobre los siguientes contrastes:

- B contra A , C y D ;
- C contra A y D ;
- A contra D .

13.13 El propósito del estudio *The Incorporation of a Chelating Agent into a Flame Retardant Finish of a Cotton Flannelette and the Evaluation of Selected Fabric Properties*, llevado a cabo en Virginia Tech, fue evaluar el uso de un agente quelante como parte del acabado retardante del fuego de la franela de algodón, determinando sus efectos en la inflamabilidad después de lavar la tela en condiciones específicas. Se pre-

pararon dos baños, uno con celulosa de carboximetilo y otro sin ella. Se lavaron 12 piezas de tela 5 veces en el baño I, y otras 12 piezas se lavaron 10 veces en el baño I. Esto se repitió con 24 piezas adicionales de tela en el baño II. Después de los lavados se midieron las longitudes quemadas de la tela, así como los tiempos de combustión. Por conveniencia, definamos los siguientes tratamientos:

Tratamiento 1: 5 lavados en el baño I,

Tratamiento 2: 5 lavados en el baño II,

Tratamiento 3: 10 lavados en el baño I,

Tratamiento 4: 10 lavados en el baño II.

Los registros del tiempo de combustión, en segundos, son los siguientes:

Tratamiento			
1	2	3	4
13.7	6.2	27.2	18.2
23.0	5.4	16.8	8.8
15.7	5.0	12.9	14.5
25.5	4.4	14.9	14.7
15.8	5.0	17.1	17.1
14.8	3.3	13.0	13.9
14.0	16.0	10.8	10.6
29.4	2.5	13.5	5.8
9.7	1.6	25.5	7.3
14.0	3.9	14.2	17.7
12.3	2.5	27.4	18.3
12.3	7.1	11.5	9.9

- a) Efectúe un análisis de varianza a un nivel de significancia de 0.01, y determine si hay diferencias significativas entre las medias de los tratamientos.
- b) Use contrastes de un solo grado de libertad con $\alpha = 0.01$ para comparar el tiempo medio de combustión del tratamiento 1 en comparación con el tratamiento 2, y también del tratamiento 3 en comparación con el 4.

13.14 El Departamento de Alimentación y Nutrición Humana de Virginia Tech realizó el estudio *Loss of Nitrogen Through Sweat by Preadolescent Boys Consuming Three Levels of Dietary Protein* para determinar la pérdida de nitrógeno por transpiración con varios niveles dietéticos de proteínas. En el experimento participaron 12 hombres preadolescentes cuyas edades iban de 7 años 8 meses a 9 años 8 meses, y a quienes de les calificó de clínicamente saludables. Cada muchacho estuvo sujeto a una de tres dietas controladas en las cuales consumía 29, 54 u 84 gramos de proteínas por día. Los siguientes datos representan la pérdida de nitrógeno corporal por transpiración, en miligramos, recabados durante los dos días últimos del periodo de experimentación:

Nivel de proteínas		
29 gramos	54 gramos	84 gramos
190	318	390
266	295	321
270	271	396
	438	399
	402	

- a) Realice un análisis de varianza a un nivel de significancia de 0.05, para demostrar que las pérdidas medias de nitrógeno por transpiración son diferentes con los tres niveles de proteínas.
- b) Utilice una prueba de Tukey para determinar cuáles niveles de proteínas difieren significativamente entre sí en la pérdida media de nitrógeno.

13.15 Utilice la prueba de Tukey a un nivel de significancia de 0.05, para analizar las medias de las 5 mar-

cas diferentes de tabletas para el dolor de cabeza del ejercicio 13.2 de la página 518.

13.16 Se realizó una investigación para determinar la fuente de reducción en el rendimiento de cierto producto químico. Se sabía que la pérdida en el rendimiento ocurría en el licor madre, es decir, el material eliminado en la etapa de filtración. Se pensaba que mezclas distintas del material original podrían ocasionar reducciones diferentes del rendimiento en la etapa de licor madre. A continuación se presentan los resultados de la reducción porcentual para tres lotes de cada una de cuatro mezclas seleccionadas con anterioridad.

Mezcla			
1	2	3	4
25.6	25.2	20.8	31.6
24.3	28.6	26.7	29.8
27.9	24.7	22.2	34.3

- a) Haga el análisis de varianza al nivel de significancia $\alpha = 0.05$.
- b) Utilice la prueba de Duncan de rango múltiple para determinar cuáles mezclas difieren.
- c) Resuelva el inciso b usando la prueba de Tukey.

13.17 En el estudio, denominado *An Evaluation of the Removal Method for Estimating Benthic Populations and Diversity*, realizado por Virginia Tech en el río Jackson, se emplearon 5 procedimientos distintos de muestreo para determinar los conteos de especies. Se seleccionaron 20 muestras al azar y los 5 procedimientos de muestreo se repitieron 4 veces. Se registraron los siguientes conteos de especies:

Procedimiento de muestreo				
Disminución	De Hess modificado	Remoción del sustrato de		
		Surber	Kicknet	Kicknet
85	75	31	43	17
55	45	20	21	10
40	35	9	15	8
77	67	37	27	15

- a) ¿Hay alguna diferencia significativa en el conteo promedio de especies para los distintos procedimientos de muestreo? Use un valor P en su conclusión.
- b) Emplee una prueba de Tukey con $\alpha = 0.05$ para determinar cuáles procedimientos de muestreo difieren.

13.18 Los siguientes datos son valores de presión (psi) en un resorte de torsión para valores distintos del ángulo entre las vueltas del resorte en posición libre.

Ángulo					
67	71	75		79	83
83	84	86	87	89	90
85	85	87	87	90	92
	85	88	88	90	
	86	88	88	91	
	86	88	89		
	87	90			

Calcule un análisis de varianza de un solo factor para este experimento y plantee sus conclusiones acerca del efecto que tiene el ángulo sobre la presión en el resorte. (Tomado de C. R. Hicks, *Fundamental Concepts in the Design of Experiments*, Holt, Rinehart y Winston, Nueva York, 1973).

13.19 Se sospecha que la temperatura del ambiente en que se activan las baterías afecta su vida. Se probaron 30 baterías homogéneas, seis a cada una de cinco temperaturas, y los datos se presentan a continuación (vida activada en segundos). Analice e interprete los datos. (Tomado de C. R. Hicks, *Fundamental Concepts in Design of Experiments*, Holt, Rinehart y Winston, Nueva York, 1973.)

Temperatura(°C)					
0	25	50	75	100	
55	60	70	72	65	
55	61	72	72	66	
57	60	72	72	60	
54	60	68	70	64	
54	60	77	68	65	
56	60	77	69	65	

13.20 La tabla siguiente (tomada de A. Hald, *Statistical Theory with Engineering Applications*, John Wiley & Sons, Nueva York, 1952) proporciona las resistencias a la tensión (en desviaciones desde 340) para conductores extraídos de nueve cables que deben usarse para una red de alto voltaje. Cada cable está constituido por 12 conductores. Se desea saber si las resistencias medias de los conductores en los nueve cables son las mismas. Si los cables son diferentes, ¿cuáles son los que difieren? Utilice un valor P en su análisis de varianza.

Cable	Resistencia a la tensión											
1	5	-13	-5	-2	-10	-6	-5	0	-3	2	-7	-5
2	-11	-13	-8	8	-3	-12	-12	-10	5	-6	-12	-10
3	0	-10	-15	-12	-2	-8	-5	0	-4	-1	-5	-11
4	-12	4	2	10	-5	-8	-12	0	-5	-3	-3	0
5	7	1	5	0	10	6	5	2	0	-1	-10	-2
6	1	0	-5	-4	-1	0	2	5	1	-2	6	7
7	-1	0	2	1	-4	2	7	5	1	0	-4	2
8	-1	0	7	5	10	8	1	2	-3	6	0	5
9	2	6	7	8	15	11	-7	7	10	7	8	1

13.21 La salida de resultados que se presenta en la figura 13.5 de la página 532 proporciona información sobre la prueba de Duncan para los datos de los agregados del ejemplo 13.1 obtenidos con la función PROC GLM del programa SAS. Saque conclusiones sobre las comparaciones por pares usando los resultados de la prueba de Duncan.

13.22 Realice la prueba de Duncan para comparaciones por pares con los datos del ejercicio 13.6 de la página 519. Comente los resultados.

13.23 En un experimento biológico se emplearon 4 concentraciones de cierto producto químico para mejorar el crecimiento de cierto tipo de planta con el paso del tiempo. Se utilizaron cinco plantas con cada concentración y se midió su crecimiento, en centímetros. Se obtuvieron los siguientes datos y también se aplicó un control (ausencia de producto químico)

Control	Concentración				
	1	2	3	4	
6.8	8.2	7.7	6.9	5.9	
7.3	8.7	8.4	5.8	6.1	
6.3	9.4	8.6	7.2	6.9	
6.9	9.2	8.1	6.8	5.7	
7.1	8.6	8.0	7.4	6.1	

Utilice una prueba bilateral de Dunnett a un nivel de significancia de 0.05 para comparar de manera simultánea las concentraciones con el control.

13.24 La estructura financiera de una empresa consiste en la forma en que sus activos se dividen en capital y deuda, y el apalancamiento financiero se refiere al porcentaje de activos financiados con endeudamiento. En el artículo *The Effect of Financial Leverage on Return*, Tai Ma, de Virginia Tech, afirma que es posible utilizar el apalancamiento financiero para incrementar la tasa de rendimiento sobre el capital. Dicho de otra manera, los accionistas pueden recibir rendimientos más elevados sobre el capital propio con la misma cantidad de inversión si usan apalancamiento financiero. Los siguientes datos muestran las tasas de rendimiento sobre el capital utilizando 3 niveles distintos de apalancamiento financiero, así como un nivel de control (deuda igual a cero) para 24 empresas seleccionadas al azar.

Apalancamiento financiero				
Control	Bajo	Medio	Alto	
2.1	6.2	9.6	10.3	
5.6	4.0	8.0	6.9	
3.0	8.4	5.5	7.8	
7.8	2.8	12.6	5.8	
5.2	4.2	7.0	7.2	
2.6	5.0	7.8	12.0	

Fuente: Standard & Poor's *Machinery Industry Survey*, 1975.

- Haga el análisis de varianza a un nivel de significancia de 0.05.
- Use una prueba de Dunnett a un nivel de significancia de 0.01, para determinar si las tasas medias de rendimiento sobre el capital son más elevadas con los niveles bajo, medio y alto de apalancamiento financiero que con el nivel de control.

The GLM Procedure				
Duncan's Multiple Range Test for moisture				
NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.				
Alpha		0.05		
Error Degrees of Freedom		25		
Error Mean Square		4960.813		
Number of Means	2	3	4	5
Critical Range	83.75	87.97	90.69	92.61
Means with the same letter are not significantly different.				
Duncan Grouping		Mean	N	aggregate
A		610.67	6	5
A				
A		610.50	6	3
A				
A		569.33	6	2
A				
A		553.33	6	1
B		465.17	6	4

Figura 13.5: Salida de resultados del SAS para el ejercicio 13.21.

13.7 Comparación de un conjunto de tratamientos en bloques

En la sección 13.2 estudiamos la idea de la formación de bloques, es decir, de aislar conjuntos de unidades experimentales que son razonablemente homogéneas y asignarles tratamientos de forma aleatoria. Ésta es una extensión del concepto de “formar pares” que se analizó en los capítulos 9 y 10, y se hace para reducir el error experimental, ya que las unidades en un bloque tienen más características comunes que las unidades localizadas en diferentes bloques.

El lector no debería considerar los bloques como un segundo factor, aunque ésa sea una forma tentadora de visualizar el diseño. De hecho, el factor principal (los tratamientos) aún lleva el peso mayor del experimento. Las unidades experimentales siguen siendo la fuente del error, igual que en el diseño completamente aleatorizado. Con la formación de bloques simplemente tratamos a dichas unidades de manera más sistemática. De ese modo, se dice que la aleatoriedad tiene restricciones. Antes de iniciar el estudio de la formación de bloques revisaremos dos ejemplos de un **diseño completamente aleatorizado**. El primer ejemplo es un experimento químico diseñado para determinar si hay una diferencia en la reacción media producida por cuatro catalizadores. Las muestras de los materiales que tienen que probarse se extraen de los mismos lotes de materias primas, a la vez que se mantienen constantes otras condiciones como la temperatura y concentración de los reactivos. En este caso, la hora del día en que se efectúan las corridas experimentales podría representar las unidades experimentales, y si el experimentador considera que es posible que haya un ligero efecto del tiempo, aleatorizaría la asignación

de los catalizadores a las corridas para contrarrestar la posible tendencia. Como un segundo ejemplo de dicho diseño, considere un experimento para comparar cuatro métodos para medir una propiedad física en particular de un fluido. Suponga que el proceso de muestreo es destructivo, es decir, que una vez que se ha medido una muestra de la sustancia usando un método, ya no puede medirse nuevamente con ningún otro. Si se decide hacer cinco mediciones con cada método, entonces se seleccionan al azar 20 muestras del material de un lote grande y se utilizan en el experimento para comparar los cuatro métodos de medición. Las unidades experimentales son las muestras seleccionadas al azar. Cualquier variación de una muestra a otra aparecerá en la variación del error, según se mida con s^2 en el análisis.

¿Cuál es el propósito de formar bloques?

Si la variación debida a la heterogeneidad en las unidades experimentales es tan grande que la sensibilidad para detectar diferencias de tratamiento se reduce debido a un valor aumentado de s^2 , un plan mejor sería “bloquear” la variación debida a esas unidades y, por consiguiente, reducir la variación ajena a la que es explicada por bloques más pequeños o más homogéneos. Por ejemplo, suponga que en el ejemplo anterior de los catalizadores se supiera *a priori* que existe en definitiva un efecto significativo diario sobre el producto, y que es posible medir el producto para cuatro catalizadores en un día específico. En lugar de asignar los 4 catalizadores a las 20 corridas de prueba completamente al azar, se eligen, por ejemplo, 5 días y se prueba cada uno de los cuatro catalizadores cada día, asignándolos al azar a las corridas dentro de los días. De esta manera se elimina la variación diaria del análisis y, en consecuencia, el error experimental, que aún incluye cualquier tendencia temporal *dentro de los días*, representa con más precisión la variación aleatoria. A cada día se le denomina **bloque**.

El más directo de los diseños aleatorizados de bloques es aquel en el cual se asigna al azar un tratamiento a la vez a cada bloque. A un plan experimental como éste se le denomina **diseño de bloques completos aleatorizados (BCA)** y cada bloque constituye una sola réplica de los tratamientos.

13.8 Diseños de bloques completos aleatorizados

Un plan clásico para el diseño de bloques completos aleatorizados (BCA) usando tres mediciones en cuatro bloques es el siguiente:

Bloque 1	Bloque 2	Bloque 3	Bloque 4
t_2	t_1	t_3	t_2
t_1	t_3	t_2	t_1
t_3	t_2	t_1	t_3

Las t denotan la asignación de cada uno de 3 tratamientos a los bloques. Desde luego, la asignación verdadera de los tratamientos a las unidades dentro de los bloques se hace al azar. Una vez que ha finalizado el experimento, los datos se pueden registrar como en el siguiente arreglo de 3×4 :

Tratamiento	Bloque: 1	2	3	4
1	y_{11}	y_{12}	y_{13}	y_{14}
2	y_{21}	y_{22}	y_{23}	y_{24}
3	y_{31}	y_{32}	y_{33}	y_{34}

donde y_{11} representa la respuesta que se obtiene al utilizar el tratamiento 1 en el bloque 1, y_{12} es la respuesta que se obtiene al utilizar el tratamiento 1 en el bloque 2,..., y y_{34} es la respuesta que se obtiene al utilizar el tratamiento 3 en el bloque 4.

Ahora vamos a generalizar y a considerar el caso de k tratamientos asignados a b bloques. Los datos se pueden resumir tal como se observa en el arreglo rectangular de $k \times b$ de la tabla 13.7. Se supondrá que las y_{ij} , $i = 1, 2, \dots, k$ y $j = 1, 2, \dots, b$, son valores de variables aleatorias independientes que tienen distribuciones normales con media μ_{ij} y varianza común σ^2 .

Tabla 13.7: Arreglo de $k \times b$ para el diseño de BCA

Tratamiento	Bloque						Total	Media
	1	2	...	j	...	b		
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1b}	$T_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2b}	$T_{2.}$	$\bar{y}_{2.}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots	\vdots
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ib}	$T_{i.}$	$\bar{y}_{i.}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots	\vdots
k	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kb}	$T_{k.}$	$\bar{y}_{k.}$
Total	$T_{.1}$	$T_{.2}$...	$T_{.j}$...	$T_{.b}$	$T_{..}$	
Media	$\bar{y}_{.1}$	$\bar{y}_{.2}$...	$\bar{y}_{.j}$...	$\bar{y}_{.b}$		$\bar{y}_{..}$

Sea $\mu_{i.}$ el promedio (en lugar del total) de las b medias de la población para el i -ésimo tratamiento. Es decir,

$$\mu_{i.} = \frac{1}{b} \sum_{j=1}^b \mu_{ij}, \text{ para } i = 1, \dots, k.$$

De manera similar, el promedio de las medias de la población para el j -ésimo bloque, $\mu_{.j}$, es definido por

$$\mu_{.j} = \frac{1}{k} \sum_{i=1}^k \mu_{ij}, \text{ para } j = 1, \dots, b$$

y el promedio de las bk medias de la población, μ , es definido por

$$\mu = \frac{1}{bk} \sum_{i=1}^k \sum_{j=1}^b \mu_{ij}.$$

Para determinar si parte de la variación de nuestras observaciones se debe a diferencias entre los tratamientos, se considera la siguiente prueba:

Hipótesis de
medias iguales
de los tratamientos

$$H_0: \mu_1 = \mu_2 = \cdots \mu_k = \mu,$$

$$H_1: \text{No todas las } \mu_i \text{ son iguales.}$$

Modelo para el diseño BCA

Cada observación se puede escribir en la forma siguiente:

$$y_{ij} = \mu_{ij} + \epsilon_{ij},$$

donde ϵ_{ij} mide la desviación del valor observado y_{ij} de la media de la población μ_{ij} . La forma preferida de esta ecuación se obtiene sustituyendo

$$\mu_{ij} = \mu + \alpha_i + \beta_j,$$

donde α_i es, como antes, el efecto del i -ésimo tratamiento, y β_j es el efecto del j -ésimo bloque. Se supone que el tratamiento y los efectos de los bloques son aditivos. Por lo tanto, se puede escribir

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

Observe que el modelo se parece al de clasificación de un solo factor; la diferencia esencial es la introducción del efecto de bloque β_j . El concepto básico se parece mucho al de la clasificación de un solo factor, excepto que en el análisis debe tomarse en cuenta el efecto adicional debido a los bloques, ya que ahora la variación se controla de manera sistemática *en dos direcciones*. Si ahora imponemos las restricciones de que

$$\sum_{i=1}^k \alpha_i = 0 \quad \text{y} \quad \sum_{j=1}^b \beta_j = 0,$$

entonces

$$\mu_i = \frac{1}{b} \sum_{j=1}^b (\mu + \alpha_i + \beta_j) = \mu + \alpha_i, \text{ para } i = 1, \dots, k,$$

y

$$\mu_j = \frac{1}{k} \sum_{i=1}^k (\mu + \alpha_i + \beta_j) = \mu + \beta_j, \text{ para } j = 1, \dots, b.$$

La hipótesis nula de que las k medias de los tratamientos μ_i son iguales y, por lo tanto, iguales a μ , ahora es **equivalente a probar la hipótesis**:

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0,$$

$$H_1: \text{Al menos una de las } \alpha_i \text{ no es igual a cero.}$$

Cada una de las pruebas de tratamientos se basará en una comparación de los estimados independientes de la varianza común poblacional σ^2 . Esos estimados se

obtendrán separando la suma total de cuadrados de los datos en tres componentes mediante la siguiente identidad:

Teorema 13.3: Identidad de la suma de cuadrados

$$\begin{aligned} \sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 &= b \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 + k \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 \\ &\quad + \sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \end{aligned}$$

La demostración se deja como ejercicio para el lector.

La identidad de la suma de cuadrados podría presentarse simbólicamente mediante la ecuación

$$STC = SCT + SCB + SCE,$$

donde

$$STC = \sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 \quad = \text{suma total de cuadrados,}$$

$$SCT = b \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 \quad = \text{suma de los cuadrados de los tratamientos,}$$

$$SCB = k \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 \quad = \text{suma de los cuadrados de los bloques,}$$

$$SCE = \sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = \text{suma de los cuadrados del error.}$$

Si se sigue el procedimiento descrito en el teorema 13.2, donde se interpretó a las sumas de cuadrados como funciones de las variables aleatorias independientes, $Y_{11}, Y_{12}, \dots, Y_{kb}$, se puede demostrar que los valores esperados de las sumas de los cuadrados de los tratamientos, los bloques y los errores son dados por

$$E(SCT) = (k-1)\sigma^2 + b \sum_{i=1}^k \alpha_i^2, \quad E(SCB) = (b-1)\sigma^2 + k \sum_{j=1}^b \beta_j^2,$$

$$E(SCE) = (b-1)(k-1)\sigma^2.$$

Como en el caso del problema de un solo factor, tenemos que el cuadrado medio del tratamiento es

$$s_1^2 = \frac{SCT}{k-1}.$$

Si los efectos del tratamiento $\alpha_1 = \alpha_2 = \dots = \alpha_k = 0$, entonces s_1^2 es un estimado insesgado de σ^2 . Sin embargo, si los efectos de los tratamientos no son todos iguales a cero, se tiene que:

Media cuadrada
esperada
del tratamiento

$$E\left(\frac{SCT}{k-1}\right) = \sigma^2 + \frac{b}{k-1} \sum_{i=1}^k \alpha_i^2$$

En este caso s_1^2 sobrestima σ^2 . Un segundo estimado de σ^2 , basado en $b-1$ grados de libertad, es

$$s_2^2 = \frac{SCB}{b-1}.$$

El estimado s_2^2 es un estimado no sesgado de σ^2 si los efectos de los bloques $\beta_1 = \beta_2 = \dots = \beta_b = 0$. Si los efectos de los bloques no son iguales a cero, entonces,

$$E\left(\frac{SCB}{b-1}\right) = \sigma^2 + \frac{k}{b-1} \sum_{j=1}^b \beta_j^2,$$

y s_2^2 sobrestimaré a σ^2 . Un tercer estimado de σ^2 , basado en $(k-1)(b-1)$ grados de libertad e independiente de s_1^2 y s_2^2 , es

$$s^2 = \frac{SCE}{(k-1)(b-1)},$$

que es no sesgado independientemente de la veracidad o falsedad de cualquier hipótesis nula.

Para probar la hipótesis nula de que los efectos de los tratamientos son iguales a cero, se calcula la razón $f_1 = s_1^2/s^2$, que es un valor de la variable aleatoria F_1 , que tiene una distribución F con $k-1$ y $(k-1)(b-1)$ grados de libertad, cuando la hipótesis nula es verdadera. La hipótesis nula se rechaza al nivel de significancia α cuando

$$f_1 > f_\alpha[k-1, (k-1)(b-1)].$$

En la práctica, primero calculamos STC , SCT y SCB , y después, utilizando la identidad de la suma de cuadrados, obtenemos SCE mediante una resta. Los grados de libertad asociados con SCE por lo general también se obtienen por sustracción; es decir,

$$(k-1)(b-1) = kb - 1 - (k-1) - (b-1).$$

Los cálculos necesarios para un problema de análisis de varianza para un diseño de bloques completos aleatorizados se puede resumir como se observa en la tabla 13.8.

Ejemplo 13.6: Se consideran cuatro máquinas diferentes, M_1 , M_2 , M_3 y M_4 , para ensamblar un producto específico. Se decidió que para comparar las máquinas se usarían 6 operadores distintos en un experimento de bloques aleatorizados. Las máquinas se asignaron al azar a cada operador. La operación de las máquinas requiere destreza física, y se anticipó que habría una diferencia en la velocidad con que los operadores trabajaban con las máquinas. En la tabla 13.9 se observan los tiempos (en segundos) requeridos para ensamblar el producto.

A un nivel de significancia de 0.05, pruebe la hipótesis H_0 de que las máquinas se desempeñan con el mismo índice de velocidad promedio.

Tabla 13.8: Análisis de varianza para el diseño de bloques completos aleatorizados

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Tratamientos	SCT	$k - 1$	$s_1^2 = \frac{SCT}{k - 1}$	$f_1 = \frac{s_1^2}{s^2}$
Bloques	SCB	$b - 1$	$s_2^2 = \frac{SCB}{b - 1}$	
Error	SCE	$(k - 1)(b - 1)$	$s^2 = \frac{SCE}{(k - 1)(b - 1)}$	
Total	STC	$kb - 1$		

Tabla 13.9: Tiempo para ensamblar el producto, en segundos

Máquina	Operador						Total
	1	2	3	4	5	6	
1	42.5	39.3	39.6	39.9	42.9	43.6	247.8
2	39.8	40.1	40.5	42.3	42.5	43.1	248.3
3	40.2	40.5	41.3	43.4	44.9	45.1	255.4
4	41.3	42.2	43.5	44.2	45.9	42.3	259.4
Total	163.8	162.1	164.9	169.8	176.2	174.1	1010.9

Solución: Las hipótesis son

H_0 : $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ (los efectos de las máquinas son iguales a cero),

H_1 : Al menos una de las α_i no es igual a cero.

Para producir el análisis de varianza que aparece en la tabla 13.10 se emplean las fórmulas de la suma de cuadrados que se presentan en la página 536 y los grados de libertad. El valor $f = 3.34$ es significativo con $P = 0.048$. Si se emplea $\alpha = 0.05$ como al menos una aproximación burda, se concluye que las máquinas no se desempeñan con el mismo índice de velocidad media. ■

Tabla 13.10: Análisis de varianza para los datos de la tabla 13.9

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Máquinas	15.93	3	5.31	3.34
Operadores	42.09	5	8.42	
Error	23.84	15	1.59	
Total	81.86	23		

Comentarios adicionales acerca de la formación de bloques

En el capítulo 10 presentamos un procedimiento para comparar medias cuando las observaciones estaban *ordenadas por pares*. El procedimiento implicaba “restar” el efecto

debido a la paridad homogénea para así trabajar con las diferencias. Este es un caso especial de diseño de bloques completos aleatorizados con $k = 2$ tratamientos. Las n unidades homogéneas a las que se asignaron los tratamientos adoptan el papel de bloques.

Si hay heterogeneidad en las unidades experimentales, el experimentador no debería confundirse y pensar que siempre es ventajoso reducir el error experimental mediante el uso de pequeños bloques homogéneos. De hecho podría haber casos en los que no es deseable formar bloques. El propósito de reducir la varianza del error es incrementar la *sensibilidad* de la prueba para detectar diferencias en las medias de los tratamientos. Esto se refleja en la potencia del procedimiento de prueba. (En la sección 13.11 se analiza con mayor detalle la potencia del procedimiento de prueba del análisis de varianza). La potencia para detectar ciertas diferencias entre las medias de los tratamientos se incrementa con una disminución de la varianza del error. Sin embargo, la potencia también se ve afectada por los grados de libertad con los que se estima la varianza, y la formación de bloques reduce los grados de libertad que están disponibles desde $k(b - 1)$ para la clasificación de un solo factor, hasta $(k - 1)(b - 1)$. De modo que se podría perder potencia con la formación de bloques si no hay una reducción significativa de la varianza del error.

Interacción entre bloques y tratamientos

Otra suposición importante que está implícita en la escritura del modelo para un diseño de bloques completos aleatorizados es que los efectos de los bloques y del tratamiento son aditivos. Esto equivale a decir que

$$\mu_{ij} - \mu_{ij'} = \mu_{i'j} - \mu_{i'j'} \quad \text{o bien} \quad \mu_{ij} - \mu_{i'j} = \mu_{ij'} - \mu_{i'j'},$$

para cada valor de i, i', j y j' . Es decir, la diferencia entre las medias de la población para los bloques j y j' es la misma para cada tratamiento, y la diferencia entre las medias de la población para los tratamientos i e i' es la misma para cada bloque. Las líneas paralelas de la figura 13.6a ilustran un conjunto de respuestas medias para las cuales los efectos de los tratamientos y los bloques son aditivos, mientras que las líneas que se intersecan en la figura 13.6b exhiben una situación en la que se dice que los efectos de los tratamientos y de los bloques **interactúan**. Con respecto al ejemplo 13.6, si el operador 3 es en promedio 0.5 segundos más rápido que el operador 2 cuando utiliza la máquina 1, entonces el operador 3 será 0.5 segundos más rápido, en promedio, que el operador 2 cuando se empleen las máquinas 2, 3 o 4. En muchos experimentos no se cumple la suposición de aditividad y el análisis descrito en esta sección llevaría a conclusiones erróneas. Por ejemplo, suponga que el operador 3 es 0.5 segundos más rápido, en promedio, que el operador 2 si emplea la máquina 1, pero que es 0.2 segundos más lento, en promedio, que el operador 2 si utiliza la máquina 2. En ese caso los operadores y las máquinas estarían interactuando.

Una inspección de la tabla 13.9 sugiere la posible presencia de interacción. Esta aparente interacción podría ser real o podría deberse al error experimental. El análisis del ejemplo 13.6 se basó en la suposición de que la aparente interacción se debe por completo al error experimental. Si la variabilidad total de nuestros datos se debiera en parte al efecto de la interacción, esa fuente de variación seguiría formando parte de la suma de cuadrados del error, **provocando que el cuadrado medio del error sobrestime a σ^2** , incrementando así la probabilidad de cometer un error tipo II. De hecho, hemos supuesto un modelo incorrecto. Si permitimos que $(\alpha\beta)_{ij}$ denote el efecto de la interacción del

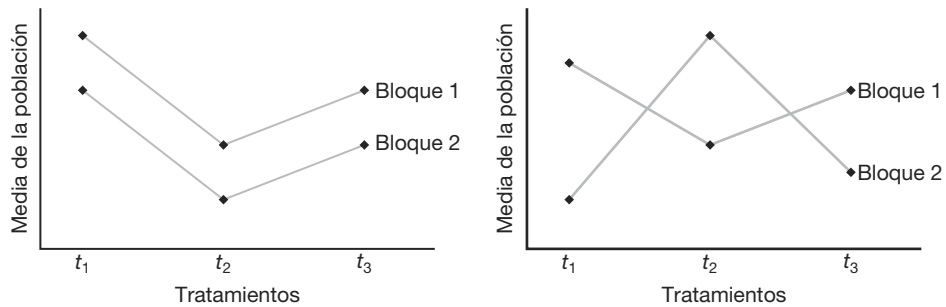


Figura 13.6: Medias de la población para a) resultados aditivos y b) efectos de interacción.

i -ésimo tratamiento y el j -ésimo bloque, podríamos escribir un modelo más adecuado con la forma siguiente:

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij},$$

al que se impondrían las restricciones adicionales

$$\sum_{i=1}^k (\alpha\beta)_{ij} = \sum_{j=1}^b (\alpha\beta)_{ij} = 0, \text{ para } i = 1, \dots, k \text{ y } j = 1, \dots, b.$$

Ahora es fácil comprobar que

$$E \left[\frac{SCE}{(b-1)(k-1)} \right] = \sigma^2 + \frac{1}{(b-1)(k-1)} \sum_{i=1}^k \sum_{j=1}^b (\alpha\beta)_{ij}^2.$$

Así, el cuadrado medio del error es considerado un **estimado sesgado de σ^2 cuando se ha ignorado la interacción existente**. En este momento parecería necesario utilizar un procedimiento para detectar la interacción en aquellos casos en que se sospecha que exista. Tal procedimiento requiere que se disponga de un estimado no sesgado e independiente de σ^2 . Por desgracia, el diseño de bloques aleatorizados no se presta a una prueba de este tipo, a menos que se modifique el diseño inicial del experimento. En el capítulo 14 se estudia este tema de manera detallada.

13.9 Métodos gráficos y verificación del modelo

En varios capítulos de este libro se hace referencia a procedimientos gráficos para mostrar datos y resultados analíticos. En los primeros capítulos se usaron gráficas de tallo y hojas y de caja y extensión como auxiliares visuales para resumir muestras. En el capítulo 10 se emplearon diagnósticos similares para entender mejor los datos de dos problemas de muestreo. En el capítulo 11 se introdujo el concepto de gráfica de residuales para detectar violaciones de las suposiciones estándar. En los últimos años gran parte de la atención dedicada al análisis de datos se ha centrado en los **métodos gráficos**. Al igual que en la regresión, el análisis de varianza se presta a la elaboración de gráficas que

ayudan a resumir los datos y a detectar violaciones. Por ejemplo, una gráfica sencilla de las observaciones brutas alrededor de la media de cada tratamiento proporciona al analista una noción de la variabilidad entre las medias muestrales y dentro de las muestras. La figura 13.7 ilustra una de tales gráficas para los datos de agregados que se presentan en la tabla 13.1. A partir de la apariencia de la gráfica se obtiene incluso una idea de cuáles agregados (si los hubiera) destacan de los demás. Es evidente que el agregado 4 resalta del resto, y que los agregados 3 y 5 forman un grupo homogéneo, así como los agregados 1 y 2.

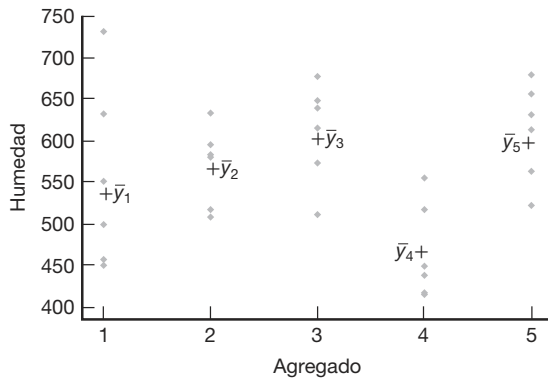


Figura 13.7: Gráfica de los datos alrededor de la media para los datos de los agregados de la tabla 13.1.

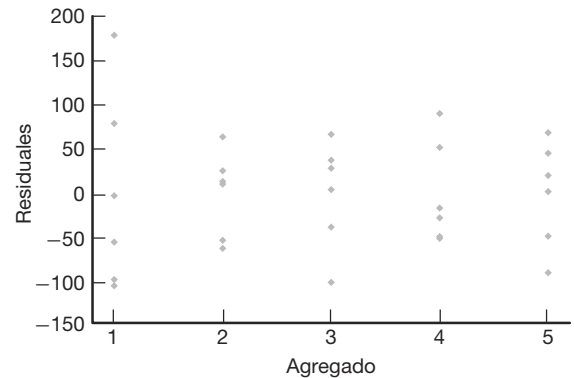


Figura 13.8: Gráfica de los residuales para cinco agregados con los datos de la tabla 13.1.

Como ocurre en el caso de la regresión, los residuales son útiles en el análisis de varianza para dar un diagnóstico sobre la detección de violaciones de los supuestos. Para formar los residuales sólo necesitamos considerar el modelo del problema de un solo factor, que es

$$y_{ij} = \mu_i + \epsilon_{ij}.$$

Es fácil determinar que el estimado de μ_i es \bar{y}_i . Por lo tanto, el ij -ésimo residual es $\bar{y}_{ij} - \bar{y}_j$, lo cual se extiende fácilmente al modelo de bloques completos aleatorizados. Sería aleccionador graficar los residuos para cada agregado con el fin de obtener cierta información sobre la suposición de varianza homogénea. Esta gráfica se muestra en la figura 13.8.

Las tendencias en gráficas como éstas podrían revelar dificultades en ciertas situaciones, especialmente cuando la violación de una suposición en particular se manifiesta en la gráfica. En el caso de la figura 13.8, los residuales parecen indicar que las varianzas *dentro de los tratamientos* son razonablemente homogéneas, excepto la del agregado 1. Hay cierta evidencia gráfica de que la varianza del agregado 1 es más grande que la del resto.

¿Qué es un residual para un diseño de BCA?

El diseño de bloques completos aleatorizados es otra situación experimental en la cual una gráfica permite que el analista se sienta cómodo con una “imagen ideal” o que tal

vez detecte dificultades. Recuerde que el modelo para el diseño de bloques completos aleatorizados es

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, b,$$

con las restricciones impuestas

$$\sum_{i=1}^k \alpha_i = 0, \quad \sum_{j=1}^b \beta_j = 0.$$

Para determinar qué es lo que en realidad constituye un residual considere que

$$\alpha_i = \mu_i - \mu, \quad \beta_j = \mu_j - \mu$$

y que μ se estima por medio de $\bar{y}_{..}$, μ_i se estima por medio de $\bar{y}_{i.}$ y μ_j se estima por medio de $\bar{y}_{.j}$. Como resultado, el *valor ajustado* o pronosticado \bar{y}_{ij} es dado por

$$\hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..},$$

y, entonces, el residual en la observación (i, j) es dado por

$$y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}.$$

Observe que \bar{y}_{ij} , el valor ajustado, es un estimado de la media μ_{ij} . Esto es congruente con la partición de la variabilidad dada en el teorema 13.3, en la que la suma de los cuadrados del error es

$$SCE = \sum_i^k \sum_j^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2.$$

Las técnicas visuales para el diseño de bloques completos aleatorizados requieren que se grafiquen los residuos por separado para cada tratamiento y cada bloque. Si la suposición de varianza homogénea se cumple, el analista debería esperar una variabilidad aproximadamente igual. El lector seguramente recordará que en el capítulo 12 se estudiaron gráficas de los residuales con el objetivo de detectar si el modelo era inadecuado. En el caso del diseño de bloques completos aleatorizados, una grave falla del modelo podría estar relacionada con la suposición de aditividad (lo cual significa que no hay interacción). Si no hay interacción debe surgir un patrón aleatorio.

Considere los datos del ejemplo 13.6, donde los tratamientos son cuatro máquinas y los bloques son seis operadores. Las figuras 13.9 y 13.10 incluyen las gráficas de los residuales para tratamientos separados y bloques separados. La figura 13.11 presenta una gráfica de los residuales contra los valores ajustados. La figura 13.9 revela que quizá la varianza del error no sea la misma para todas las máquinas, y lo mismo podría ocurrir con la varianza del error para cada uno de los seis operadores. Sin embargo, al parecer dos residuales inusualmente grandes son los que provocan la aparente dificultad. La figura 13.11 es una gráfica de residuales que revela evidencia razonable de un comportamiento aleatorio. Sin embargo, sobresalen los dos residuales grandes ya detectados.

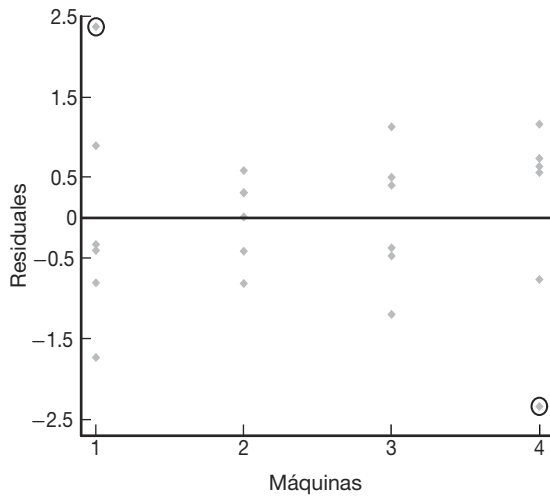


Figura 13.9: Gráfica de residuos para las cuatro máquinas de los datos del ejemplo 13.6.

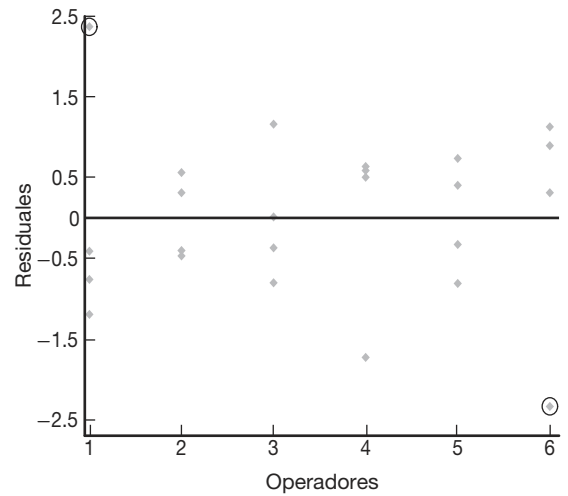


Figura 13.10: Gráfica de residuos para los seis operadores de los datos del ejemplo 13.6.

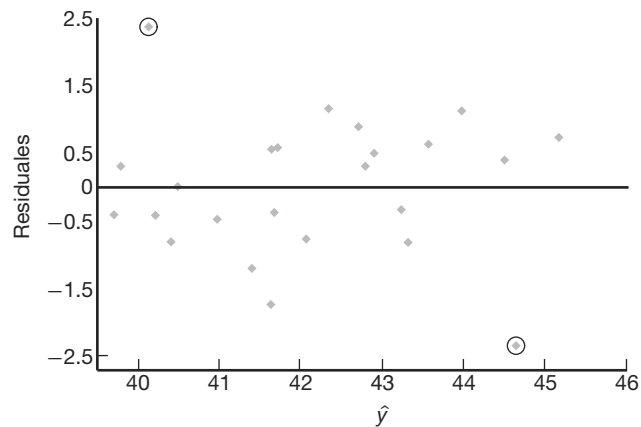


Figura 13.11: Residuales graficados contra los valores ajustados para los datos del ejemplo 13.6.

13.10 Transformaciones de datos en el análisis de varianza

En el capítulo 11 se puso mucha atención a la transformación de la respuesta y en situaciones para las que se ajustaba un modelo de regresión lineal a un conjunto de datos. Es evidente que se aplican los mismos conceptos a la regresión lineal múltiple, aunque esto no se analizó en el capítulo 12. En el estudio de los modelos de regresión se hizo énfasis en las transformaciones de y que producirían un modelo que se ajustara mejor a los datos que uno en el que la y ingresara de forma lineal. Por ejemplo, si la estructura del “tiempo” es de naturaleza exponencial, entonces una transformación logarítmica de

y linealiza la estructura y, por lo tanto, se anticipa más éxito cuando se utiliza la respuesta transformada.

Aunque el propósito fundamental de la transformación de datos que se ha analizado hasta este momento ha sido mejorar el ajuste del modelo, hay otras razones para transformar o reexpresar la respuesta y , y muchas de ellas se relacionan con las suposiciones que se hacen, por ejemplo, las suposiciones de las cuales depende la validez del análisis. Una suposición muy importante en el análisis de varianza es la de la varianza homogénea que se estudió antes en la sección 13.4. Se supone una **varianza común** σ^2 . Si la varianza difiere mucho de un tratamiento a otro, y se realiza el ANOVA estándar que se estudia en este capítulo (y en otros posteriores), los resultados serían muy deficientes. En otras palabras, el análisis de varianza no es **robusto** respecto a la suposición de varianza homogénea. Como se ha dicho hasta el momento, se trata del motivo principal para la graficación de los residuales que estudiamos en la sección anterior y que ilustramos en las figuras 13.9, 13.10 y 13.11. Esas gráficas permiten detectar problemas debidos a una varianza no homogénea. Sin embargo, ¿qué hay que hacer al respecto? ¿Cómo se corrigen?

¿De dónde proviene la varianza no homogénea?

Con frecuencia, aunque no siempre, la varianza no homogénea en el ANOVA existe debido a la distribución de las respuestas. Ahora, por supuesto, se supone la normalidad de la respuesta, pero hay ciertas situaciones en las que se necesitan pruebas de las medias aunque la distribución de la respuesta sea una de las distribuciones no normales que se estudiaron en los capítulos 5 y 6, es decir, la distribución de Poisson, la logarítmica normal, la exponencial y la gamma. Realmente existen problemas del tipo del ANOVA con datos de conteo, duración antes de la falla, etcétera.

En los capítulos 5 y 6 se demostró que, además del caso de la normal, la varianza de una distribución con frecuencia será función de la media, es decir, $\sigma_i^2 = g(\mu_i)$. Por ejemplo, en el caso de la distribución de Poisson, $\text{Var}(Y_i) = \mu_i = \sigma_i^2$, lo que significa que la *varianza es igual a la media*. En el caso de la distribución exponencial, $\text{Var}(Y_i) = \sigma_i^2 = \mu_i^2$, o sea que la *varianza es igual al cuadrado de la media*. Para el caso de la logarítmica normal, una transformación logarítmica produce una distribución normal con varianza constante σ^2 .

Los mismos conceptos que usamos en el capítulo 4 para determinar la varianza de una función no lineal pueden ayudarnos a determinar la naturaleza de la *transformación estabilizadora de la varianza* $g(y_i)$. Recuerde la expansión de las series de Taylor de primer orden de $g(y_i)$ alrededor de $y_i = \mu_i$, donde $g'(\mu_i) = \left[\frac{\partial g(y_i)}{\partial y_i} \right]_{y_i = \mu_i}$. La función de transformación $g(y)$ debe ser independiente de μ para que baste como la transformación estabilizadora de la varianza. De lo anterior

$$\text{Var}[g(y_i)] \approx [g'(\mu_i)]^2 \sigma_i^2.$$

Como resultado, $g(y_i)$ debe ser tal que $g'(\mu_i) \propto \frac{1}{\sigma}$. Así, si se sospecha que la respuesta tiene una distribución de Poisson, $\sigma_i = \mu_i^{1/2}$, de modo que $g'(\mu_i) \propto \frac{1}{\mu_i^{1/2}}$. Entonces, la transformación estabilizadora de la varianza es $g(y_i) = y_i^{1/2}$. A partir de esta ilustración y manipulaciones similares para las distribuciones exponencial y gamma, se obtiene lo siguiente.

Distribución	Transformaciones estabilizadoras de la varianza
Poisson	$g(y) = y^{1/2}$
Exponencial	$g(y) = \ln y$
Gamma	$g(y) = \ln y$

Ejercicios

13.25 Se utilizaron cuatro clases de fertilizante f_1, f_2, f_3 y f_4 para estudiar la cosecha de frijol. El suelo se dividió en 3 bloques, cada uno de los cuales contiene 4 parcelas homogéneas. A continuación se presentan las cosechas en kilogramos por parcela, así como los tratamientos correspondientes:

Bloque 1	Bloque 2	Bloque 3
$f_1 = 42.7$	$f_3 = 50.9$	$f_4 = 51.1$
$f_3 = 48.5$	$f_1 = 50.0$	$f_2 = 46.3$
$f_4 = 32.8$	$f_2 = 38.0$	$f_1 = 51.9$
$f_2 = 39.3$	$f_4 = 40.2$	$f_3 = 53.5$

Realice un análisis de varianza a un nivel de significancia de 0.05 utilizando el modelo de bloques completos aleatorizados.

13.26 Se compararon las cosechas de tres variedades de papas. El experimento se efectuó asignando cada variedad de manera aleatoria a 3 parcelas del mismo tamaño, en 4 lugares diferentes. Se registraron las siguientes cosechas para las variedades A, B y C, en 100 kilogramos por parcela:

Lugar 1	Lugar 2	Lugar 3	Lugar 4
B : 13	C : 21	C : 9	A : 11
A : 18	A : 20	B : 12	C : 10
C : 12	B : 23	A : 14	B : 17

Realice un análisis de varianza de bloques completos aleatorizados con el objetivo de probar la hipótesis de que no hay diferencia en la capacidad de rendimiento de las 3 variedades de papas. Utilice un nivel de significancia de 0.05 y saque conclusiones.

13.27 Los siguientes datos son los porcentajes de aditivos extranjeros, medidos por 5 analistas, de 3 marcas distintas de mermelada de fresa, A, B y C.

Analista 1	Analista 2	Analista 3	Analista 4	Analista 5
B : 2.7	C : 7.5	B : 2.8	A : 1.7	C : 8.1
C : 3.6	A : 1.6	A : 2.7	B : 1.9	A : 2.0
A : 3.8	B : 5.2	C : 6.4	C : 2.6	B : 4.8

A un nivel de significancia de 0.05, realice un análisis de varianza de bloques completos aleatorizados para probar la hipótesis de que el porcentaje de aditivos extranjeros es el mismo para las tres marcas de mermelada. ¿Cuál de ellas parece tener menos aditivos?

13.28 Los siguientes datos representan las calificaciones finales obtenidas por 5 estudiantes en matemáticas, inglés, francés y biología:

Estudiante	Materia			
	Matemáticas	Inglés	Francés	Biología
1	68	57	73	61
2	83	94	91	86
3	72	81	63	59
4	55	73	77	66
5	92	68	75	87

Pruebe la hipótesis de que los cursos tienen la misma dificultad. Use un valor P en sus conclusiones y analice sus hallazgos.

13.29 En el estudio *The Periphyton of the South River, Virginia: Mercury Concentration, Productivity, and Autotrophic Index Studies*, efectuado por el Departamento de Ciencias e Ingeniería Ambientales de Virginia Tech, se midió la concentración total de mercurio en sólidos totales de perifiton en seis estaciones distintas durante seis días diferentes. Determine si el contenido medio de mercurio difiere significativamente entre las estaciones utilizando los siguientes datos. Use un valor P y analice sus hallazgos.

Fecha	Estación					
	CA	CB	E1	E2	E3	E4
8 de abril	0.45	3.24	1.33	2.04	3.93	5.93
23 de junio	0.10	0.10	0.99	4.31	9.92	6.49
1 de julio	0.25	0.25	1.65	3.13	7.39	4.43
8 de julio	0.09	0.06	0.92	3.66	7.88	6.24
15 de julio	0.15	0.16	2.17	3.50	8.82	5.39
23 de julio	0.17	0.39	4.30	2.91	5.50	4.29

13.30 Una planta de energía nuclear produce una gran cantidad de calor que generalmente se descarga en los sistemas de agua. Ese calor eleva la temperatura del líquido, lo cual da como resultado una mayor concentración de clorofila a que, a su vez, alarga la temporada de crecimiento. Para estudiar este efecto se tomaron muestras de agua mensualmente en 3 estaciones, durante un periodo de 12 meses. La estación A es la que se ubica más cerca de una descarga potencial de agua caliente, la estación C es la más lejana de la descarga y la estación B se encuentra entre las estaciones A y C. Se registraron las siguientes concentraciones de clorofila a .

Mes	Estación		
	A	B	C
Enero	9.867	3.723	4.410
Febrero	14.035	8.416	11.100
Marzo	10.700	20.723	4.470
Abril	13.853	9.168	8.010
Mayo	7.067	4.778	34.080
Junio	11.670	9.145	8.990
Julio	7.357	8.463	3.350
Agosto	3.358	4.086	4.500
Septiembre	4.210	4.233	6.830
Octubre	3.630	2.320	5.800
Noviembre	2.953	3.843	3.480
Diciembre	2.640	3.610	3.020

Realice un análisis de varianza y, a un nivel de significancia de 0.05, pruebe la hipótesis de que no hay diferencia en las concentraciones medias de clorofila *a* en las 3 estaciones.

13.31 En un estudio realizado por el Departamento de Salud y Educación Física de Virginia Tech, se asignaron 3 dietas durante 3 días a 6 sujetos utilizando un diseño de bloques completos aleatorizados. Los sujetos, que desempeñan el papel de bloques, recibieron las siguientes 3 dietas en orden aleatorio:

- Dieta 1: grasas mixtas y carbohidratos,
- Dieta 2: alta en grasas,
- Dieta 3: alta en carbohidratos.

Al terminar el periodo de tres días se puso a cada sujeto en una banda caminadora y se midió el tiempo, en segundos, que transcurría hasta que se sentían exhaustos. Efectúe un análisis de varianza separando la dieta, los sujetos y la suma de cuadrados del error. Utilice un valor *P* para determinar si existen diferencias significativas entre las dietas. Los datos registrados son los siguientes:

Dieta	Sujeto					
	1	2	3	4	5	6
1	84	35	91	57	56	45
2	91	48	71	45	61	61
3	122	53	110	71	91	122

13.32 El personal forestal utiliza arsénico orgánico como arboricida. La cantidad de arsénico que absorbe el cuerpo cuando se expone a este producto constituye un grave problema de salud. Es importante que la cantidad de exposición se determine rápido, de manera que pueda retirarse del trabajo a los empleados con niveles elevados de arsénico. En un experimento descrito en el artículo “A Rapid Method for the Determination of Arsenic Concentrations in Urine at Field Locations”, publicado en el *American Industrial Hygiene Association Journal* (Vol. 37, 1976), especímenes de orina de 4 personas del servicio forestal fueron divididos por igual en

tres muestras para que pudiera analizarse el contenido de arsénico en la orina de cada individuo en un laboratorio universitario: las muestras eran analizadas por un químico con un sistema portátil, así como también por un empleado forestal que había recibido una capacitación breve. Se registraron los siguientes niveles de arsénico, en partes por millón:

Individuo	Analista		
	Empleado	Químico	Laboratorio
1	0.05	0.05	0.04
2	0.05	0.05	0.04
3	0.04	0.04	0.03
4	0.15	0.17	0.10

Realice un análisis de varianza y, a un nivel de significancia de 0.05, pruebe la hipótesis de que no hay diferencia en los niveles de arsénico con los tres métodos de análisis.

13.33 Los científicos del Departamento de Patología Vegetal de Virginia Tech realizaron un experimento en el que se aplicaron 5 tratamientos diferentes en 6 lugares distintos de un huerto de manzanas para determinar si había diferencias significativas en el crecimiento entre los tratamientos. Los tratamientos 1 a 4 representan distintos herbicidas y el 5 es un control. El periodo de crecimiento fue de mayo a noviembre de 1982, y los datos de crecimiento nuevo, medido en centímetros, para muestras seleccionadas de los 6 lugares en el huerto, son los siguientes:

Tratamiento	Ubicaciones					
	1	2	3	4	5	6
1	455	72	61	215	695	501
2	622	82	444	170	437	134
3	695	56	50	443	701	373
4	607	650	493	257	490	262
5	388	263	185	103	518	622

Lleve a cabo un análisis de varianza, separando el tratamiento, el lugar y la suma de cuadrados del error. Determine si hay diferencias significativas entre las medias de los tratamientos. Mencione un valor *P*.

13.34 En el artículo “Self-Control and Therapist Control in the Behavioral Treatment of Overweight Women”, publicado en *Behavioral Research and Therapy* (Vol. 10, 1972), se estudiaron dos tratamientos de reducción y otro de control para observar sus efectos en el cambio del peso en mujeres obesas. Los dos tratamientos reductores involucrados fueron un programa autodirigido de reducción de peso y otro controlado por un terapeuta. Se asignó a cada uno de 10 sujetos a uno de los 3 programas de tratamiento en orden aleatorio y se midió la pérdida de peso. Se registraron los siguientes cambios en el peso:

Sujeto	Tratamiento		
	Control	Autodirigido	Con terapeuta
1	1.00	-2.25	-10.50
2	3.75	-6.00	-13.50
3	0.00	-2.00	0.75
4	-0.25	-1.50	-4.50
5	-2.25	-3.25	-6.00
6	-1.00	-1.50	4.00
7	-1.00	-10.75	-12.25
8	3.75	-0.75	-2.75
9	1.50	0.00	-6.75
10	0.50	-3.75	-7.00

Realice un análisis de varianza y, a un nivel de significancia de 0.01, pruebe la hipótesis de que no hay diferencia en las pérdidas de peso promedio para los 3 tratamientos. ¿Cuál tratamiento fue el mejor?

13.35 En el libro *Design of Experiments for the Quality Improvement*, publicado por la Japanese Standards Association (1989) se reportó un estudio sobre la cantidad de tinta que se requiere para obtener el mejor color para cierto tipo de tela. En dos plantas diferentes se administraron tres cantidades de tinta: $\frac{1}{3}$ del porcentaje de wof, es decir, $\frac{1}{3}$ del porcentaje del peso de la tela, 1% de wof y 3% de wof. Después se observó cuatro veces la densidad del color de la tela para cada nivel de tinta aplicada en cada planta.

	Cantidad de tinta					
	1/3%		1%		3%	
Planta 1	5.2	6.0	12.3	10.5	22.4	17.8
	5.9	5.9	12.4	10.9	22.5	18.4
Planta 2	6.5	5.5	14.5	11.8	29.0	23.2
	6.4	5.9	16.0	13.6	29.7	24.0

A un nivel de significancia de 0.05, realice un análisis de varianza para probar la hipótesis de que no hay dife-

rencia en la densidad de color de la tela con los tres niveles de tinta. Considere a las plantas como bloques.

13.36 Se realizó un experimento con el fin de comparar tres tipos de materiales para recubrir alambres de cobre. El propósito del recubrimiento consiste en eliminar los “defectos” del alambre. A cada recubrimiento se le asignaron al azar 10 especímenes distintos, de 5 milímetros de longitud, para que les fuera aplicado. Después se sometió a los 30 especímenes a cierto tipo de desgaste abrasivo. Al final se midió el número de defectos en cada uno y se obtuvieron los siguientes resultados:

	Material											
	1				2				3			
6	8	4	5		3	3	5	4	12	8	7	14
7	7	9	6		2	4	4	5	18	6	7	18
7	8				4	3			8	5		

Suponga que se acepta que se puede aplicar un proceso de Poisson, por lo que el modelo es $Y_{ij} = \mu_i + \epsilon_{ij}$, donde μ_i es la media de la distribución de Poisson y $\sigma_{Y_{ij}}^2 = \mu_i$.

- Haga una transformación apropiada de los datos y un análisis de varianza.
- Determine si hay evidencia suficiente para preferir un material de recubrimiento sobre los demás. Muestre cualesquiera hallazgos que sugieran una conclusión.
- Haga una gráfica de residuales y coméntela.
- Mencione el propósito de la transformación de los datos.
- ¿Qué otra suposición se hace en este caso, que quizá la transformación no cumpla por completo?
- Comente el inciso *e* después de elaborar una gráfica de probabilidad normal sobre los residuales.

13.11 Modelos de efectos aleatorios

A lo largo de este capítulo estudiamos los procedimientos del análisis de varianza en los que el objetivo principal es estudiar el efecto sobre ciertas respuestas de ciertos tratamientos fijos o predeterminados. Los experimentos en los que los tratamientos o los niveles de tratamiento son preseleccionados por el experimentador, y no elegidos al azar, se denominan **experimentos de efectos fijos**. Para el modelo de efectos fijos sólo se hacen inferencias acerca de los tratamientos específicos que se utilizaron en el experimento.

Con frecuencia es importante que el experimentador sea capaz de hacer inferencias acerca de una población de tratamientos a través de un experimento en el que los tratamientos empleados se elijan al azar de entre la población. Por ejemplo, un biólogo podría estar interesado en saber si hay o no una varianza significativa en alguna característica fisiológica debida a un tipo de animal. Los tipos de animales que en realidad se usan en el experimento se eligen al azar y representan los efectos del tratamiento. Un químico podría estar interesado en estudiar el efecto de los laboratorios sobre el análisis químico de

una sustancia; no le interesa un laboratorio en particular, sino una población grande de laboratorios. Así, podría seleccionar al azar un grupo de laboratorios y asignar muestras a cada uno para su análisis. Entonces, la inferencia estadística implicaría 1) probar si los laboratorios contribuyen o no a una varianza diferente de cero en los resultados de los análisis, y 2) estimar la varianza debida a los laboratorios y a la varianza dentro de los mismos.

Modelo y suposiciones para el modelo de efectos aleatorios

El **modelo de efectos aleatorios** de un solo factor se escribe como el modelo de efectos fijos, pero sus términos tienen significados diferentes. La respuesta $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ es ahora un valor de la variable aleatoria

$$Y_{ij} = \mu + A_i + \epsilon_{ij}, \text{ con } i = 1, 2, \dots, k \text{ y } j = 1, 2, \dots, n,$$

donde las A_i tienen distribución normal e independiente con media igual a cero y varianza σ_α^2 , y son independientes de las ϵ_{ij} . Al igual que para el modelo de efectos fijos, las ϵ_{ij} también tienen distribución normal e independiente con media igual a cero y varianza σ^2 . Observe que para un experimento de efectos aleatorios, ya no se aplica la restricción de que $\sum_{i=1}^k \alpha_i = 0$.

Teorema 13.4:

Para el modelo del análisis de varianza de efectos aleatorios de un solo factor,

$$E(SCT) = (k - 1)\sigma^2 + n(k - 1)\sigma_\alpha^2 \quad \text{y} \quad E(SCE) = k(n - 1)\sigma^2.$$

La tabla 13.11 presenta los cuadrados medios esperados tanto para un experimento de efectos fijos como para uno de efectos aleatorios. Los cálculos para un experimento de efectos aleatorios se realizan exactamente de la misma forma que los del experimento de efectos fijos. Es decir, la suma de cuadrados, los grados de libertad y las columnas de los cuadrados medios en la tabla del análisis de varianza son iguales para ambos modelos.

Tabla 13.11: Cuadrados medios esperados para el experimento de un solo factor

Fuente de variación	Grados de libertad	Cuadrados medios	Cuadrados medios esperados	
			Efectos fijos	Efectos aleatorios
Tratamientos	$k - 1$	s_1^2	$\sigma^2 + \frac{n}{k - 1} \sum_i \alpha_i^2$	$\sigma^2 + n\sigma_\alpha^2$
Error	$k(n - 1)$	s^2	σ^2	σ^2
Total	$nk - 1$			

Para el modelo de efectos aleatorios la hipótesis de que todos los efectos del tratamiento son iguales a cero se escribe como sigue:

Hipótesis para un experimento de efectos aleatorios	$H_0: \sigma_\alpha^2 = 0,$ $H_1: \sigma_\alpha^2 \neq 0.$
---	---

Esta hipótesis afirma que los diferentes tratamientos no contribuyen en absoluto a la variabilidad de la respuesta. De la tabla 13.11 es evidente que tanto s_1^2 como s^2 son estimados de σ^2 cuando H_0 es verdadera, y que la razón

$$f = \frac{s_1^2}{s^2}$$

es un valor de la variable aleatoria F que tiene la distribución F con $k - 1$ y $k(n - 1)$ grados de libertad. La hipótesis nula se rechaza a un nivel de significancia α cuando

$$f > f_\alpha[k - 1, k(n - 1)].$$

En muchos estudios científicos y de ingeniería el interés no se centra en la prueba F . El científico sabe que el efecto aleatorio, en efecto, es significativo. Lo más importante es la estimación de los diversos componentes de la varianza. Esto produce un sentido de *jerarquía* en términos de cuáles factores producen la mayor variabilidad y en qué cantidad. En este contexto podría ser interesante cuantificar cuánto más grande es el *componente de la varianza de un solo factor* que el producido por el azar (variación aleatoria).

Estimación de los componentes de la varianza

La tabla 13.11 también se utiliza para estimar los **componentes de la varianza** σ^2 y σ_α^2 . Como s_1^2 estima $\sigma^2 + n\sigma_\alpha^2$, y s^2 estima σ^2 ,

$$\hat{\sigma}^2 = s^2, \quad \hat{\sigma}_\alpha^2 = \frac{s_1^2 - s^2}{n}.$$

Ejemplo 13.7: Los datos de la tabla 13.12 representan observaciones codificadas sobre el producto de un proceso químico en el que se utilizan 5 lotes de materia prima seleccionados al azar. Demuestre que el componente de la varianza del lote es significativamente mayor que cero y obtenga su estimado.

Tabla 13.12: Datos para el ejemplo 13.7

Lote	1	2	3	4	5	
	9.7	10.4	15.9	8.6	9.7	
	5.6	9.6	14.4	11.1	12.8	
	8.4	7.3	8.3	10.7	8.7	
	7.9	6.8	12.8	7.6	13.4	
	8.2	8.8	7.9	6.4	8.3	
	7.7	9.2	11.6	5.9	11.7	
	8.1	7.6	9.8	8.1	10.7	
Total	55.6	59.7	80.7	58.4	75.3	329.7

Solución: La suma total de cuadrados, la del lote y la suma de cuadrados del error son, respectivamente,

$$STC = 194.64, SCT = 72.60 \text{ y } SCE = 194.64 - 72.60 = 122.04.$$

En la tabla 13.13 se presentan estos resultados con el resto de los cálculos.

Tabla 13.13: Análisis de la varianza para el ejemplo 13.7

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Lotes	72.60	4	18.15	4.46
Error	122.04	30	4.07	
Total	194.64	34		

La razón f es significativa al nivel $\alpha = 0.05$, lo que indica que se rechaza la hipótesis de un componente del lote igual a cero. Una estimación del componente de la varianza del lote es

$$\hat{\sigma}_\alpha^2 = \frac{18.15 - 4.07}{7} = 2.01.$$

Observe que mientras que el **componente de la varianza del lote** es significativamente diferente de cero, cuando se compara contra el estimado de σ^2 , es decir, $\hat{\sigma}^2 = CME = 4.07$, pareciera que el componente de varianza del lote no es considerablemente grande. ─

Si el resultado que se obtiene con la fórmula para σ_α^2 es negativo, o sea, que s_1^2 es menor que s^2 , entonces a $\hat{\sigma}_\alpha^2$ se le asigna un valor de cero. Éste es un estimador sesgado. Para tener un mejor estimador de σ_α^2 , suele emplearse un método llamado **verosimilitud restringida** (o **residual**) **máxima (REML** por sus siglas en inglés) (véase Harville, 1977, en la bibliografía). Este tipo de estimador se puede encontrar en muchos paquetes estadísticos para computadora. Los detalles de dicho procedimiento rebasan el alcance de este libro.

Diseño de bloques aleatorizados con bloques aleatorios

En un experimento de bloques completos aleatorizados, donde los bloques representan días, es concebible que el experimentador quiera que los resultados se apliquen no sólo a los días reales utilizados en el análisis, sino a cada día del año. Entonces, seleccionaría al azar los días en que se haría el experimento, así como los tratamientos y el modelo de efectos aleatorios a utilizar.

$$Y_{ij} = \mu + A_i + B_j + \epsilon_{ij}, \text{ para } i = 1, 2, \dots, k \text{ y } j = 1, 2, \dots, b,$$

donde las A_i , B_j y ϵ_{ij} son variables aleatorias independientes con medias igual a cero y varianzas σ_α^2 , σ_β^2 y σ^2 , respectivamente. Se obtienen los cuadrados medios esperados para un diseño de bloques completos aleatorizados de efectos aleatorios usando el mismo procedimiento que se usó en el problema de un solo factor; en la tabla 13.14 se presentan junto con los de un experimento de efectos fijos.

Nuevamente, los cálculos para las sumas de cuadrados y grados de libertad individuales son idénticos a los del modelo de efectos fijos. Las hipótesis

$$H_0: \sigma_\alpha^2 = 0,$$

$$H_1: \sigma_\alpha^2 \neq 0$$

se obtienen calculando

$$f = \frac{s_1^2}{s^2}$$

Tabla 13.14: Cuadrados medios esperados para un diseño de bloques completos aleatorizados

Fuente de variación	Grados de libertad	Cuadrados medios	Cuadrados medios esperados	
			Efectos fijos	Efectos aleatorios
Tratamientos	$k - 1$	s_1^2	$\sigma^2 + \frac{b}{k-1} \sum_i \alpha_i^2$	$\sigma^2 + b\sigma_\alpha^2$
Bloques	$b - 1$	s_2^2	$\sigma^2 + \frac{k}{b-1} \sum_j \beta_j^2$	$\sigma^2 + k\sigma_\beta^2$
Error	$(k-1)(b-1)$	s^2	σ^2	σ^2
Total	$kb - 1$			

y H_0 se rechaza cuando $f > f_\alpha[k-1, (b-1)(k-1)]$.

Los estimados no sesgados de los componentes de la varianza son

$$\hat{\sigma}^2 = s^2, \quad \hat{\sigma}_\alpha^2 = \frac{s_1^2 - s^2}{b}, \quad \hat{\sigma}_\beta^2 = \frac{s_2^2 - s^2}{k}.$$

Las pruebas de las hipótesis referentes a los diversos componentes de la varianza se realizan calculando las razones de los cuadrados medios adecuados, tal como se indica en la tabla 13.14, y comparándolos con los valores f correspondientes de la tabla A.6.

13.12 Estudio de caso

Estudio de caso 13.1: Análisis químico. Se pidió al personal del Departamento de Química de Virginia Tech que analizara un conjunto de datos que se obtuvo para comparar 4 métodos distintos de análisis del aluminio en cierta mezcla deflagradora sólida. Para considerar una amplia gama de laboratorios de análisis se utilizaron 5 de ellos en el experimento. Se seleccionaron esos laboratorios porque suelen realizar esa clase de análisis. Se asignaron al azar 20 muestras de material deflagrador que contenían 2.70% de aluminio, cuatro a cada laboratorio, y se dieron instrucciones acerca de cómo efectuar los análisis químicos utilizando los cuatro métodos. Los datos que se obtuvieron son los siguientes:

Método	Laboratorio					Media
	1	2	3	4	5	
A	2.67	2.69	2.62	2.66	2.70	2.668
B	2.71	2.74	2.69	2.70	2.77	2.722
C	2.76	2.76	2.70	2.76	2.81	2.758
D	2.65	2.69	2.60	2.64	2.73	2.662

Los laboratorios no se consideran efectos aleatorios, ya que no fueron seleccionados al azar de entre una población más grande de ellos. Los datos se analizaron como un diseño de bloques completos aleatorizados. Se dibujaron gráficas de los datos para determinar si era apropiado un modelo aditivo del tipo:

$$y_{ij} = \mu + m_i + l_j + \epsilon_{ij}$$

en otras palabras, un modelo con efectos aditivos. El bloque aleatorizado no es adecuado cuando existe interacción entre los laboratorios y los métodos. Considere la gráfica de la figura 13.12. Aunque es un poco difícil de interpretar porque cada punto representa una sola observación, parece que no hay interacción evidente entre los métodos y los laboratorios.

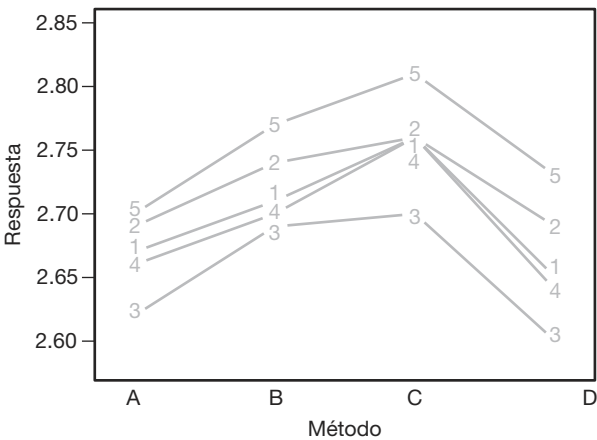


Figura 13.12: Gráfica de interacción para los datos del estudio de caso 13.1.

Gráficas de residuales

Las gráficas de residuales se usaron como indicaciones de diagnóstico con respecto a la suposición de una varianza homogénea. La figura 13.13 presenta una gráfica de residuales contra los métodos de análisis. La variabilidad descrita en los residuales parece ser bastante homogénea. Para completar, en la figura 13.14 se presenta una gráfica de probabilidad normal de los residuales.

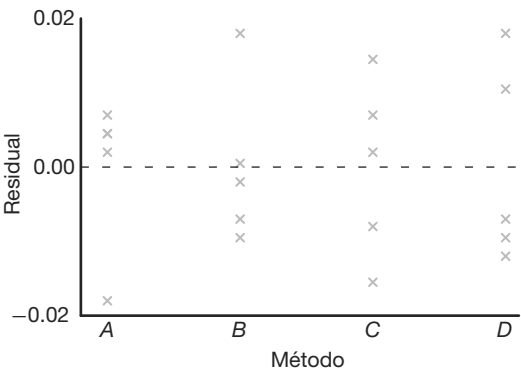


Figura 13.13: Gráfica de residuales en comparación con el método para los datos del estudio de caso 13.1.

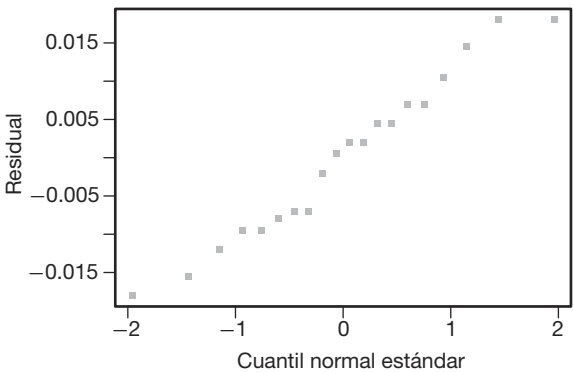


Figura 13.14: Gráfica de probabilidad normal de residuales para los datos del estudio de caso 13.1.

Las gráficas de residuales no muestran problemas con la suposición de errores normales ni con la de varianza homogénea. Para hacer el análisis de varianza se utilizó la

función PROC GLM del programa SAS. En la figura 13.15 se incluye una salida de resultados por computadora con comentarios.

Los valores f y P calculados sí indican una diferencia significativa entre los métodos de análisis. A este análisis le puede seguir un análisis de comparación múltiple para determinar en dónde están las diferencias entre los métodos.

Ejercicios

13.37 Al probar muestras de sangre de un paciente para detectar anticuerpos del VIH un espectrómetro determina la densidad óptica de cada muestra. La densidad óptica se mide como la absorbencia de la luz de cierta longitud de onda. La muestra de sangre es positiva si excede a cierto valor límite que se determina con muestras de control para esa corrida. A los investigadores les interesa comparar la variabilidad del laboratorio para los valores de control positivo. Los datos representan valores de control positivo para 10 corridas distintas en cuatro laboratorios seleccionados al azar.

Corrida	Laboratorio			
	1	2	3	4
1	0.888	1.065	1.325	1.232
2	0.983	1.226	1.069	1.127
3	1.047	1.332	1.219	1.051
4	1.087	0.958	0.958	0.897
5	1.125	0.816	0.819	1.222
6	0.997	1.015	1.140	1.125
7	1.025	1.071	1.222	0.990
8	0.969	0.905	0.995	0.875
9	0.898	1.140	0.928	0.930
10	1.018	1.051	1.322	0.775

- Escriba un modelo adecuado para este experimento.
- Estime el componente de varianza del laboratorio y la varianza dentro de los laboratorios.

13.38 Se efectúa un experimento en el que se compararán 4 tratamientos en 5 bloques. Los datos son los siguientes:

Tratamiento	Bloque				
	1	2	3	4	5
1	12.8	10.6	11.7	10.7	11.0
2	11.7	14.2	11.8	9.9	13.8
3	11.5	14.7	13.6	10.7	15.9
4	12.6	16.5	15.4	9.6	17.1

- Suponga que se trata de un modelo de efectos aleatorios y pruebe la hipótesis de que no hay diferencia entre las medias de los tratamientos, a un nivel de significancia de 0.05.
- Calcule estimados de los componentes de la varianza del tratamiento y del bloque.

13.39 Los siguientes datos muestran el efecto de cuatro operadores, elegidos al azar, sobre la producción de una máquina específica:

	Operador			
	1	2	3	4
	175.4	168.5	170.1	175.2
	171.7	162.7	173.4	175.7
	173.0	165.0	175.7	180.1
	170.5	164.1	170.7	183.7

- Realice un análisis de varianza de efectos aleatorios a un nivel de significancia de 0.05.
- Calcule un estimado del componente de la varianza del operador y del componente de la varianza del error experimental.

13.40 De cinco “vaciados” de metales se tomaron cinco muestras del núcleo y en cada una se analizó la cantidad de un elemento traza. Los siguientes son los datos de los 5 vaciados seleccionados al azar:

Núcleo	Vaciado				
	1	2	3	4	5
1	0.98	0.85	1.12	1.21	1.00
2	1.02	0.92	1.68	1.19	1.21
3	1.57	1.16	0.99	1.32	0.93
4	1.25	1.43	1.26	1.08	0.86
5	1.16	0.99	1.05	0.94	1.41

- La intención es que los vaciados sean idénticos. Por lo tanto, pruebe que el componente de la varianza del “vaciado” es igual a cero. Saque conclusiones.
- Realice un ANOVA completo y obtenga un estimado de la varianza dentro del vaciado.

13.41 Una empresa textil produce cierta tela en un número grande de telares. Los gerentes quieren que los telares sean homogéneos para que la tela que producen tenga una resistencia uniforme. Se sospecha que puede haber una variación significativa entre la resistencia de los telares. Considere los siguientes datos para 4 telares seleccionados al azar. Cada observación es una determinación de la resistencia de la tela expresada en libras por pulgada cuadrada.

	Telar			
	1	2	3	4
	99	97	94	93
	97	96	95	94
	97	92	90	90
	96	98	92	92

- Escriba un modelo para el experimento.
- ¿El componente de la varianza del telar difiere significativamente de cero?
- Haga comentarios sobre la sospecha de los gerentes.

The GLM Procedure						
Class Level Information						
Class		Levels		Values		
Method		4		A B C D		
Lab		5		1 2 3 4 5		
Number of Observations		Read		20		
Number of Observations		Used		20		
Dependent Variable:	Response					
Sum of						
Source	DF	Squares	Mean Square	F Value	Pr > F	
Model	7	0.05340500	0.00762929	42.19	<.0001	
Error	12	0.00217000	0.00018083			
Corrected Total	19	0.05557500				
R-Square	Coeff Var	Root MSE	Response Mean			
0.960954	0.497592	0.013447	2.702500			
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Method	3	0.03145500	0.01048500	57.98	<.0001	
Lab	4	0.02195000	0.00548750	30.35	<.0001	
Observation	Observed		Predicted		Residual	
1	2.67000000		2.66300000		0.00700000	
2	2.71000000		2.71700000		-0.00700000	
3	2.76000000		2.75300000		0.00700000	
4	2.65000000		2.65700000		-0.00700000	
5	2.69000000		2.68550000		0.00450000	
6	2.74000000		2.73950000		0.00050000	
7	2.76000000		2.77550000		-0.01550000	
8	2.69000000		2.67950000		0.01050000	
9	2.62000000		2.61800000		0.00200000	
10	2.69000000		2.67200000		0.01800000	
11	2.70000000		2.70800000		-0.00800000	
12	2.60000000		2.61200000		-0.01200000	
13	2.66000000		2.65550000		0.00450000	
14	2.70000000		2.70950000		-0.00950000	
15	2.76000000		2.74550000		0.01450000	
16	2.64000000		2.64950000		-0.00950000	
17	2.70000000		2.71800000		-0.01800000	
18	2.77000000		2.77200000		-0.00200000	
19	2.81000000		2.80800000		0.00200000	
20	2.73000000		2.71200000		0.01800000	

Figura 13.15: Salida de resultados por computadora del SAS para los datos del estudio de caso 13.1.

Ejercicios de repaso

13.42 El Centro de Consultoría en Estadística de Virginia Tech, junto con el Departamento de Silvicultura, llevaron a cabo un análisis. Se aplicó cierto tratamiento a tres cepas de árbol. Se empleó el producto químico Garlon con el fin de regenerar las raíces de las cepas. Se usó un aerosol con cuatro niveles de concentración de Garlon. Después de cierto tiempo, se observó la altura de los retoños. Realice un análisis de varianza de un solo factor con los siguientes datos. Haga pruebas para saber si la concentración de Garlon tiene un efecto significativo sobre la altura de los retoños. Emplee $\alpha = 0.05$.

Nivel de Garlon							
1		2		3		4	
2.87	2.31	3.27	2.66	2.39	1.91	3.05	0.91
3.91	2.04	3.15	2.00	2.89	1.89	2.43	0.01

13.43 Considere los datos de los agregados del ejemplo 13.1. Efectúe una prueba de Bartlett a un nivel de significancia $\alpha = 0.1$ para determinar si hay heterogeneidad en la varianza entre los agregados.

13.44 En un proceso químico se utilizaron 3 catalizadores y también se incluyó un control (no catalizador). Se tienen los datos siguientes de la producción del proceso:

Catalizador				
Control	1	2	3	
74.5	77.5	81.5	78.1	
76.1	82.0	82.3	80.2	
75.9	80.6	81.4	81.5	
78.1	84.9	79.5	83.0	
76.2	81.0	83.0	82.1	

Use una prueba de Dunnett a un nivel de significancia $\alpha = 0.01$ para determinar si se obtuvo una producción significativamente más alta con los catalizadores que sin ellos.

13.45 Se emplean cuatro laboratorios para efectuar análisis químicos. Se envían muestras del mismo material a los laboratorios para que, como parte del estudio, las analicen para determinar si dan o no, en promedio, los mismos resultados. Los resultados analíticos de los cuatro laboratorios son los siguientes:

Laboratorio			
A	B	C	D
58.7	62.7	55.9	60.7
61.4	64.5	56.1	60.3
60.9	63.1	57.3	60.9
59.1	59.2	55.2	61.4
58.2	60.3	58.1	62.3

- a) Utilice una prueba de Bartlett para demostrar que las varianzas dentro de los laboratorios no difieren de manera significativa a un nivel de significancia $\alpha = 0.05$.

- b) Realice el análisis de varianza y saque conclusiones acerca de los laboratorios.
c) Dibuje una gráfica de probabilidad normal de residuales.

13.46 Se diseñó un experimento para el personal del Departamento de Ciencia Animal de Virginia Tech, con el propósito de estudiar el tratamiento con urea y amoníaco acuoso de la espiga del trigo. El propósito era mejorar el valor nutricional para las ovejas macho. Los tratamientos dietéticos son: control, urea en la alimentación, espiga tratada con amoníaco, espiga tratada con urea. En el experimento se emplearon 24 ovejas y se separaron de acuerdo con su peso relativo. En cada grupo homogéneo había cuatro ovejas (según el peso) y cada una recibió una de las cuatro dietas en orden aleatorio. Se midió el porcentaje de materia seca digerida de las 24 ovejas. Los siguientes son los datos:

Grupo por peso (bloque)						
Dieta	1	2	3	4	5	6
Control	32.68	36.22	36.36	40.95	34.99	33.89
Urea en la alimentación	35.90	38.73	37.55	34.64	37.36	34.35
Tratada con amoníaco	49.43	53.50	52.86	45.00	47.20	49.76
Tratada con urea	46.58	42.82	45.41	45.08	43.81	47.40

- a) Use un análisis de bloques completos aleatorizados para probar las diferencias entre las dietas. Emplee $\alpha = 0.05$.
b) Utilice la prueba de Dunnett para comparar las tres dietas con el control. Utilice $\alpha = 0.05$.
c) Dibuje una gráfica de probabilidad normal de los residuales.

13.47 En un estudio realizado por el personal del Departamento de Bioquímica de Virginia Tech se dieron tres dietas a un grupo de ratas con el objetivo de estudiar el efecto de cada una sobre el zinc dietético residual en el torrente sanguíneo. Se asignaron al azar cinco ratas preñadas a cada grupo dietético, y cada una recibió la dieta en el día 22 del embarazo. Se midió la cantidad de zinc en partes por millón. Los datos son los que siguen:

Dieta	1	0.50	0.42	0.65	0.47	0.44
	2	0.42	0.40	0.73	0.47	0.69
	3	1.06	0.82	0.72	0.72	0.82

Determine si hay una diferencia significativa en el zinc dietético residual entre las tres dietas. Use $\alpha = 0.05$. Lleve a cabo un ANOVA de un solo factor.

13.48 Se realizó un experimento para comparar tres tipos de pintura para buscar evidencia de diferencias en su calidad de desgaste. Las pinturas se expusieron a acciones abrasivas y se registró el tiempo, en horas, que tardaba en observarse la abrasión. Se usaron seis especímenes para cada tipo de pintura. Los datos son los siguientes:

Tipo de pintura								
1			2			3		
158	97	282	515	264	544	317	662	213
315	220	115	525	330	525	536	175	614

- a) Realice un análisis de varianza para determinar si la evidencia sugiere que la calidad del desgaste de las tres pinturas es diferente. Utilice un valor P en sus conclusiones.
- b) Si se encuentran diferencias significativas, diga cuáles son. ¿Hay alguna pintura que destaque? Analice sus hallazgos.
- c) Haga todos los análisis gráficos que necesite para determinar si son válidas las suposiciones que se hicieron en el inciso a . Analice sus hallazgos.
- d) Suponga que se determina que los datos para cada tratamiento tienen una distribución exponencial. ¿Sugiere esto un análisis alternativo? Si fuera así, hágalo y presente sus hallazgos.

13.49 Una empresa que troquila juntas de hojas de caucho, plástico y corcho desea comparar el número medio de juntas producidas por hora para los tres tipos de material. Se eligieron al azar dos máquinas troqueladoras como bloques. Los datos representan el número de juntas (en miles) producidas por hora. En la figura 13.16 de la página 557 se observa la salida de resultados del análisis.

Máquina	Corcho			Material Caucho			Plástico		
A	4.31	4.27	4.40	3.36	3.42	3.48	4.01	3.94	3.89
B	3.94	3.81	3.99	3.91	3.80	3.85	3.48	3.53	3.42

- a) ¿Por qué se eligieron las máquinas troqueladoras como bloques?
- b) Grafique las seis medias para las combinaciones de máquinas y materiales.
- c) ¿Hay un material que sea mejor?
- d) ¿Existe interacción entre los tratamientos y los bloques? Si es así, diga si la interacción ocasiona alguna dificultad seria para llegar a una conclusión adecuada. Explique su respuesta.

13.50 Se hizo un estudio para comparar el rendimiento de tres marcas de gasolina competidoras. Se seleccionaron al azar cuatro modelos de automóvil de tamaño variable. A continuación se presentan los datos, en millas por galón. El orden de prueba es aleatorio para cada modelo.

Modelo	Marca de gasolina		
	A	B	C
A	32.4	35.6	38.7
B	28.8	28.6	29.9
C	36.5	37.6	39.1
D	34.4	36.2	37.9

- a) Analice la necesidad de utilizar más de un solo modelo de automóvil.
- b) Considere el ANOVA de la salida de resultados del SAS en la figura 13.17. ¿Es importante la marca de la gasolina?
- c) ¿Qué marca de gasolina seleccionaría usted? Consulte el resultado de la prueba de Duncan.

13.51 Se utilizaron cuatro localidades diferentes del noreste para hacer mediciones de ozono, en partes por millón. Se recolectaron las cantidades de ozono en cinco muestras de cada localidad.

Localidad				
1	2	3	4	
0.09	0.15	0.10	0.10	
0.10	0.12	0.13	0.07	
0.08	0.17	0.08	0.05	
0.08	0.18	0.08	0.08	
0.11	0.14	0.09	0.09	

- a) ¿Hay información suficiente que sugiera que existen diferencias en los niveles medios de ozono entre las diferentes localidades? Guíese usando un valor P .
- b) Si se encuentran diferencias significativas en el inciso a , determine su naturaleza. Emplee cualesquiera métodos que haya aprendido.

13.52 Demuestre que el cuadrado medio del error

$$s^2 = \frac{SCE}{k(n-1)}$$

para el análisis de varianza en la clasificación de un factor es un estimado no sesgado de σ^2 .

13.53 Demuestre el teorema 13.2.

13.54 Demuestre que la fórmula para calcular la SCB, en el análisis de varianza del diseño de bloques completos aleatorizados, es equivalente al término correspondiente en la identidad del teorema 13.3.

13.55 Para el diseño de bloques aleatorizados con k tratamientos y b bloques, demuestre que

$$E(SCB) = (b-1)\sigma^2 + k \sum_{j=1}^b \beta_j^2.$$

The GLM Procedure					
Dependent Variable: gasket					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.68122778	0.33624556	76.52	<.0001
Error	12	0.05273333	0.00439444		
Corrected Total	17	1.73396111			
R-Square	Coeff Var	Root MSE	gasket Mean		
0.969588	1.734095	0.066291	3.822778		
Source	DF	Type III SS	Mean Square	F Value	Pr > F
material	2	0.81194444	0.40597222	92.38	<.0001
machine	1	0.10125000	0.10125000	23.04	0.0004
material*machine	2	0.76803333	0.38401667	87.39	<.0001
Level of material	Level of machine	-----gasket-----			
		N	Mean	Std Dev	
cork	A	3	4.32666667	0.06658328	
cork	B	3	3.91333333	0.09291573	
plastic	A	3	3.94666667	0.06027714	
plastic	B	3	3.47666667	0.05507571	
rubber	A	3	3.42000000	0.06000000	
rubber	B	3	3.85333333	0.05507571	
Level of material	N	Mean	Std Dev	-----gasket-----	
cork	6	4.12000000	0.23765521		
plastic	6	3.71166667	0.26255793		
rubber	6	3.63666667	0.24287171		
Level of machine	N	Mean	Std Dev	-----gasket-----	
A	9	3.89777778	0.39798800		
B	9	3.74777778	0.21376259		

Figura 13.16: Salida de resultados por computadora del SAS para el ejercicio de repaso 13.49.

The GLM Procedure						
Dependent Variable: MPG						
		Sum of				
Source	DF	Squares	Mean Square	F Value	Pr > F	
Model	5	153.2508333	30.6501667	24.66	0.0006	
Error	6	7.4583333	1.2430556			
Corrected Total	11	160.7091667				
R-Square	Coeff Var	Root MSE	MPG Mean			
0.953591	3.218448	1.114924	34.64167			
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Model	3	130.3491667	43.4497222	34.95	0.0003	
Brand	2	22.9016667	11.4508333	9.21	0.0148	
Duncan's Multiple Range Test for MPG						
NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.						
Alpha				0.05		
Error Degrees of Freedom				6		
Error Mean Square				1.243056		
Number of Means				2	3	
Critical Range				1.929	1.999	
Means with the same letter are not significantly different.						
Duncan Grouping				Mean	N	Brand
A				36.4000	4	C
A						
B A				34.5000	4	B
B						
B				33.0250	4	A

Figura 13.17 Salida de resultados por computadora del SAS para el ejercicio de repaso 13.50.

13.56 Proyecto de grupo: Resulta de interés determinar qué tipo de pelota deportiva se puede lanzar a la mayor distancia. La competencia incluye una pelota de tenis, una de beisbol y una de softbol. Divida el grupo en equipos de cinco estudiantes. Cada equipo debe diseñar y realizar un experimento separado, también debe analizar los datos de su propio experimento. Los cinco miembros del equipo lanzarán cada pelota (después de calentar el brazo el tiempo adecuado). La respuesta experimental será la distancia (en pies) que se lanza la pelota. Los datos de cada equipo incluirán 15 observaciones. Aspectos importantes:

a) No se trata de una competencia entre equipos. La competencia es entre los tres tipos de pelotas. Se esperaría que las conclusiones de cada equipo sean similares.

- b) En cada equipo debe haber hombres y mujeres.
- c) El diseño experimental de cada equipo deberá ser un diseño de bloques completos aleatorizados. Los cinco individuos que lanzan la pelota son los bloques.
- d) Asegúrese de incorporar la aleatorización adecuada para realizar el experimento.
- e) Los resultados deberán contener una descripción del experimento con una tabla de ANOVA que incluya un valor *P* y las conclusiones apropiadas. Se usarán técnicas gráficas y comparaciones múltiples en caso de ser necesarias. Saquen conclusiones prácticas con respecto a las diferencias entre los tipos de pelotas. Sean meticulosos.

13.13 Posibles riesgos y errores conceptuales; relación con el material de otros capítulos

Al igual que otros procedimientos estudiados en capítulos anteriores, el análisis de varianza es razonablemente robusto con respecto a la suposición de normalidad, pero no lo es tanto en cuanto a la suposición de varianza homogénea. También observamos que la prueba de Bartlett para varianzas iguales es sumamente débil en relación con la normalidad.

Este capítulo es sumamente importante, ya que se trata de un punto “de inicio” para temas importantes, como el diseño de experimentos y el análisis de varianza. En el capítulo 14 se tratan los mismos temas, pero en los casos de extensiones a más de un factor y el análisis más complicado por la interpretación de la interacción entre factores. Hay ocasiones en que el papel de la interacción en un experimento científico es más importante que el papel de los factores principales (efectos principales). Ante la presencia de interacciones se hace un énfasis aún mayor en las técnicas gráficas. En los capítulos 14 y 15 será necesario proporcionar más detalles acerca del proceso de aleatorización, ya que el número de combinaciones de factores puede ser muy grande.

Capítulo 14

Experimentos factoriales (dos o más factores)

14.1 Introducción

Considere una situación en la que haya interés por estudiar el efecto de **dos factores**, A y B , sobre alguna respuesta. Por ejemplo, en un experimento químico nos gustaría variar en forma simultánea la presión de reacción y el tiempo de reacción, y estudiar el efecto que cada uno tiene sobre el producto. En un experimento biológico resulta de interés estudiar el efecto que tienen el tiempo de secado y la temperatura sobre la cantidad de sólidos (porcentaje por peso) que queda en muestras de levadura. Igual que en el capítulo 13, el término **factor** se utiliza en un sentido general para denotar cualquier característica del experimento que pueda variar de un ensayo a otro, como la temperatura, el tiempo o la presión. Los **niveles** de un factor se definen como los valores reales que se utilizan en el experimento.

Para cada uno de estos casos es importante determinar no sólo si cada uno de los 2 factores influye en la respuesta, sino también si hay una interacción significativa entre ellos. En lo que se refiere a la terminología, el experimento descrito aquí es de 2 factores, y el diseño experimental podría ser uno completamente aleatorizado, en el que las distintas combinaciones de tratamiento se asignan al azar a todas las unidades experimentales, o bien, un diseño de bloques completos aleatorizados, donde las combinaciones de factores se asignan al azar dentro de los bloques. En el ejemplo de la levadura, si se empleara un diseño completamente aleatorizado, las distintas combinaciones de tratamientos de temperatura y tiempo de secado se asignarían al azar a las muestras de levadura.

Muchos de los conceptos que se estudiaron en el capítulo 13 se extienden a 2 y 3 factores en este capítulo. El objetivo principal de este material es el uso del diseño completamente aleatorizado con un *experimento factorial*. Un experimento factorial con 2 factores implica ensayos experimentales (o uno solo) con todas las combinaciones de factores. Por ejemplo, en el caso de la temperatura y tiempo de secado con, digamos, 3 niveles de cada uno y $n = 2$ corridas por cada una de las 9 combinaciones, tendríamos un *experimento factorial de 2 factores en un diseño completamente aleatorizado*. Ninguno de ellos es un factor de bloqueo; nos interesa la manera en que cada uno influye en el porcentaje de sólidos en las muestras, y si interactúan o no. El biólogo dispondría de 18 muestras físicas de material que constituyen unidades experimentales. Luego, éstas se asignarían al azar a las 18 combinaciones (9 combinaciones de tratamiento, cada una de ellas por duplicado).

Antes de entrar en detalles analíticos, sumas de cuadrados y demás, sería interesante que el lector observe la clara conexión que existe entre lo que hemos descrito y la situación con el problema de un solo factor. Considere el experimento de la levadura. La explicación de los grados de libertad ayuda a que el lector o el analista visualicen la extensión. En un inicio, las 9 combinaciones de tratamientos deberían considerarse como si representaran un factor con 9 niveles (8 grados de libertad). Así, un vistazo inicial a los grados de libertad arroja lo siguiente:

Combinaciones de tratamiento	8
Error	9
Total	17

Efectos principales e interacción

En realidad el experimento se podría analizar como se describe en la tabla anterior. Sin embargo, es probable que la prueba *F* para las combinaciones no dé al analista la información que desea, es decir, el papel de la temperatura y del tiempo de secado. Tres tiempos de secado tienen asociados 2 grados de libertad, y a 3 temperaturas se asocian también 2 grados de libertad. Los factores principales, la temperatura y el tiempo de secado reciben el nombre de **efectos principales**, los cuales representan 4 de los 8 grados de libertad para las *combinaciones de factores*. Los 4 grados de libertad adicionales se asocian con la *interacción* entre los 2 factores. Como resultado, el análisis incluye

Combinaciones	8
Temperatura	2
Tiempo de secado	2
Interacción	4
Error	9
Total	17

En el capítulo 13 vimos que en un análisis de varianza los factores pueden considerarse fijos o aleatorios, dependiendo del tipo de inferencia deseada y de la manera en que se eligieron los niveles. Aquí debemos considerar los efectos fijos, los efectos aleatorios e incluso los casos en que los efectos son mixtos. Conforme avancemos en estos temas pondremos mayor atención a los cuadrados medios esperados. En la siguiente sección nos centraremos en el concepto de interacción.

14.2 Interacción en el experimento de dos factores

En el modelo de bloques aleatorizados que se estudió previamente se supuso que en cada bloque se toma una observación de cada tratamiento. Si la suposición del modelo es correcta, es decir, si los bloques y los tratamientos son los únicos efectos reales y no hay interacción, el valor esperado del cuadrado medio del error es la varianza del error experimental σ^2 . Sin embargo, suponga que existe interacción entre los tratamientos y los bloques, como lo indica el modelo

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

de la sección 13.8. El valor esperado del cuadrado medio del error entonces es dado por

$$E \left[\frac{SCE}{(b-1)(k-1)} \right] = \sigma^2 + \frac{1}{(b-1)(k-1)} \sum_{i=1}^k \sum_{j=1}^b (\alpha\beta)_{ij}^2.$$

Los efectos del tratamiento y los bloques no aparecen en el cuadrado medio del error esperado, pero los efectos de la interacción sí. Entonces, si en el modelo hay interacción, el cuadrado medio del error refleja variación debida al error experimental más una contribución de la interacción y, para este plan experimental, no hay forma de separarlos.

La interacción y la interpretación de los efectos principales

Desde el punto de vista del experimentador, parecería necesario llegar a una prueba significativa sobre la existencia de una interacción, al separar la variación del error verdadero de aquel que se debe a la interacción. Los efectos principales, A y B , adoptan un significado distinto en presencia de la interacción. En el ejemplo biológico anterior el efecto que tiene el tiempo de secado sobre la cantidad de sólidos que quedan en la levadura muy bien podría depender de la temperatura a la que se expusieron las muestras. En general, podrían existir situaciones experimentales en las que el factor A tuviera un efecto positivo sobre la respuesta en un nivel del factor B ; en tanto que con un nivel distinto de B el efecto de A sería negativo. Aquí se usa el término **efecto positivo** para indicar que el producto o la respuesta se incrementan conforme los niveles de un factor dado aumentan de acuerdo con cierto orden definido. En el mismo sentido, un **efecto negativo** corresponde a una disminución de la respuesta al aumentar los niveles del factor.

Considere, por ejemplo, los siguientes datos de temperatura (factor A con niveles t_1 , t_2 y t_3 (en orden creciente) y tiempo de secado d_1 , d_2 y d_3 (también en orden creciente). La respuesta es el porcentaje de sólidos. Estos datos son completamente hipotéticos y se dan para ilustrar un aspecto.

A	B			Total
	d_1	d_2	d_3	
t_1	4.4	8.8	5.2	18.4
t_2	7.5	8.5	2.4	18.4
t_3	9.7	7.9	0.8	18.4
Total	21.6	25.2	8.4	55.2

Es evidente que el efecto de la temperatura sobre el porcentaje de sólidos es positivo para el tiempo breve de secado d_1 , pero negativo para el tiempo prolongado d_3 . Esta **interacción clara** entre la temperatura y el tiempo de secado es evidentemente interesante para el biólogo; sin embargo, con base en los totales de las respuestas para las temperaturas t_1 , t_2 y t_3 , la suma de cuadrados de la temperatura, SCT , producirá un valor de 0. Entonces, se dice que la presencia de la interacción **enmascara** el efecto de la temperatura. Por ello, si se considera el efecto medio de la temperatura, promediado para el tiempo de secado, **no existe efecto alguno**. Entonces, esto define el efecto principal. Pero, por supuesto, es probable que esto no sea pertinente para el biólogo.

Antes de sacar cualquier conclusión final a partir de las pruebas de significancia sobre los efectos principales y los efectos de la interacción, el **experimentador debería observar primero si la prueba para la interacción es o no significativa**. Si la

interacción no es significativa, entonces los resultados de las pruebas sobre los efectos principales son importantes. No obstante, si la interacción debe ser significativa, entonces solamente son importantes aquellas pruebas sobre los efectos principales que resultan significativas. En presencia de una interacción, los efectos principales no significativos bien podrían ser resultado de enmascaramiento e indicar la necesidad de observar la influencia de cada factor a niveles fijos del otro.

Representación gráfica de la interacción

La presencia de interacción, así como su impacto científico, se puede interpretar adecuadamente usando **gráficas de interacción**. Las gráficas proporcionan una clara imagen de la tendencia de los datos para mostrar el efecto que tiene el cambio de un factor conforme se pasa de un nivel a otro del segundo factor. La figura 14.1 ilustra la fuerte interacción entre la temperatura y el tiempo de secado. La interacción se revela en las líneas no paralelas.

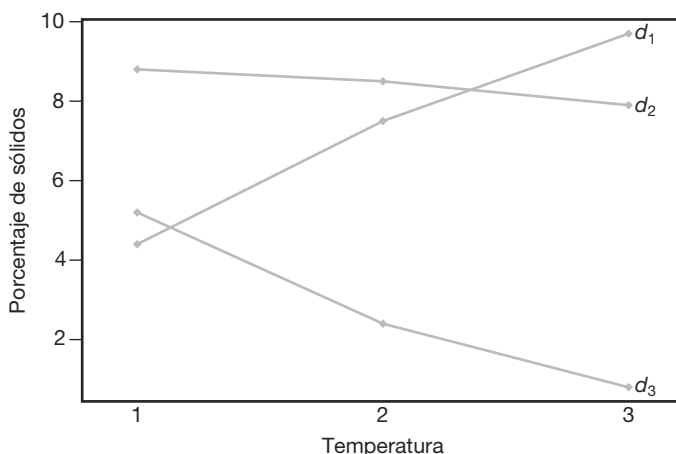


Figura 14.1: Gráfica de la interacción para los datos de temperatura y de tiempo de secado.

El *efecto relativamente fuerte de la temperatura* sobre el porcentaje de sólidos en el tiempo de secado más breve se refleja en la marcada pendiente de d_1 . En el tiempo de secado medio, d_2 , la temperatura tiene muy poco efecto, mientras que en el tiempo de secado prolongado d_3 la pendiente negativa indica un efecto negativo de la temperatura. Las gráficas de interacción como ésta le permiten al científico hacer una interpretación rápida y significativa de la interacción que existe. Debe ser evidente que el **paralelismo** en las gráficas indica la **ausencia de interacción**.

Necesidad de observaciones múltiples

En el experimento de 2 factores, la interacción y el error experimental sólo se separan si se hacen observaciones múltiples con las distintas combinaciones de tratamiento. Para máxima eficiencia debe haber el mismo número n de observaciones para cada combinación. Éstas deben ser verdaderas réplicas, no sólo medidas repetidas. Por ejemplo, en

el caso de la levadura, si para cada combinación de temperatura y tiempo de secado se toman $n = 2$ observaciones, debería haber dos muestras separadas y no sólo mediciones repetidas en la misma muestra. Esto permite que la variabilidad debida a las unidades experimentales aparezca en el “error”, de manera que la variación no es un simple error de medición.

14.3 Análisis de varianza de dos factores

Al presentar las fórmulas generales para el análisis de varianza de un experimento de 2 factores utilizando observaciones repetidas en un diseño completamente aleatorizado, debe considerarse el caso de n réplicas de las combinaciones del tratamiento, determinadas por a niveles del factor A y b niveles del factor B . Las observaciones se podrían clasificar usando un arreglo rectangular, donde los renglones representan los niveles del factor A y las columnas representan los niveles del factor B . Cada combinación de tratamiento define una celda del arreglo. Así, se tienen ab celdas, cada una de las cuales contiene n observaciones. Se denota con y_{ijk} la k -ésima observación tomada en el i -ésimo nivel del factor A y el j -ésimo nivel del factor B . En la tabla 14.1 se muestran las abn observaciones.

Tabla 14.1: Experimento de dos factores con n réplicas

A	B				Total	Media
	1	2	...	b		
1	y_{111}	y_{121}	\cdots	y_{1b1}	$Y_{1..}$	$\bar{y}_{1..}$
	y_{112}	y_{122}	\cdots	y_{1b2}		
	\vdots	\vdots		\vdots		
	y_{11n}	y_{12n}	\cdots	y_{1bn}		
2	y_{211}	y_{221}	\cdots	y_{2b1}	$Y_{2..}$	$\bar{y}_{2..}$
	y_{212}	y_{222}	\cdots	y_{2b2}		
	\vdots	\vdots		\vdots		
	y_{21n}	y_{22n}	\cdots	y_{2bn}		
\vdots	\vdots	\vdots		\vdots	\vdots	\vdots
a	y_{a11}	y_{a21}	\cdots	y_{ab1}	$Y_{a..}$	$\bar{y}_{a..}$
	y_{a12}	y_{a22}	\cdots	y_{ab2}		
	\vdots	\vdots		\vdots		
	y_{a1n}	y_{a2n}	\cdots	y_{abn}		
Total	$Y_{.1.}$	$Y_{.2.}$	\cdots	$Y_{.b.}$	$Y_{...}$	
Media	$\bar{y}_{.1.}$	$\bar{y}_{.2.}$	\cdots	$\bar{y}_{.b.}$		$\bar{y}_{...}$

Las observaciones en la celda (ij) -ésima constituyen una muestra aleatoria de tamaño n de una población que se supone tiene distribución normal con media μ_{ij} y varianza σ^2 . Se supone que todas las ab poblaciones tienen la misma varianza σ^2 . Se

definen los siguientes símbolos útiles, algunos de los cuales se utilizan en la tabla 14.1:

- Y_{ij} = suma de las observaciones en la (ij) -ésima celda,
- $Y_{i.}$ = suma de las observaciones para el i -ésimo nivel del factor A ,
- $Y_{.j}$ = suma de las observaciones para el j -ésimo nivel del factor B ,
- $Y_{...}$ = suma de todas las abn observaciones,
- \bar{y}_{ij} = media de las observaciones en la (ij) -ésima celda,
- $\bar{y}_{i.}$ = media de las observaciones para el i -ésimo nivel del factor A ,
- $\bar{y}_{.j}$ = media de las observaciones para el j -ésimo nivel del factor B ,
- $\bar{y}_{...}$ = media de todas las abn observaciones.

A diferencia de la situación para un solo factor, que se cubrió con amplitud en el capítulo 13, en éste supondremos que las **poblaciones**, de las que se toman n observaciones independientes con distribución idéntica, son **combinaciones** de los factores. Asimismo, se supondrá siempre que de cada combinación de factores se toma un número igual (n) de observaciones. En los casos en que los tamaños de las muestras por combinación son desiguales, los cálculos son más complicados, aunque los conceptos son transferibles.

Modelo e hipótesis para el problema de dos factores

Cada observación de la tabla 14.1 se puede escribir en la siguiente forma:

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk},$$

donde ϵ_{ijk} mide las desviaciones de los valores y_{ijk} observados en la (ij) -ésima celda a partir de la media de la población μ_{ij} . Si $(\alpha\beta)_{ij}$ denota el efecto de la interacción del i -ésimo nivel del factor A y el j -ésimo nivel del factor B , α_i el efecto del i -ésimo nivel del factor A , β_j el efecto del j -ésimo nivel del factor B , y μ la media conjunta, escribimos

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij},$$

y, entonces,

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

a las que se imponen las restricciones

$$\sum_{i=1}^a \alpha_i = 0, \quad \sum_{j=1}^b \beta_j = 0, \quad \sum_{i=1}^a (\alpha\beta)_{ij} = 0, \quad \sum_{j=1}^b (\alpha\beta)_{ij} = 0.$$

Las 3 hipótesis por probar son las siguientes:

1. H'_0 : $\alpha_1 = \alpha_2 = \dots = \alpha_a = 0$,
 H'_1 : Al menos una de las α_i no es igual a 0.
2. H''_0 : $\beta_1 = \beta_2 = \dots = \beta_b = 0$,
 H''_1 : Al menos una de las β_j no es igual a 0.

3. $H_0''': (\alpha\beta)_{11} = (\alpha\beta)_{12} = \dots = (\alpha\beta)_{ab} = 0,$
 $H_1''':$ Al menos una de las $(\alpha\beta)_{ij}$ no es igual a 0.

Se alerta al lector acerca del problema del enmascaramiento de los efectos principales cuando la interacción contribuye de manera importante en el modelo. Se recomienda considerar primero el resultado de la prueba de interacción y, luego, la interpretación de la prueba del efecto principal; la naturaleza de la conclusión científica depende de si se encontró interacción. Si ésta se descarta, entonces se pueden probar las hipótesis 1 y 2 y la interpretación es muy sencilla. Sin embargo, si se descubre que hay interacción, la interpretación puede ser más complicada, como se vio al analizar el tiempo de secado y la temperatura en la sección previa. La estructura de las pruebas de hipótesis 1, 2 y 3 se estudiará en las secciones siguientes. En el análisis del ejemplo 14.1 se incluirá la interpretación de los resultados.

Las pruebas de las hipótesis anteriores se basarán en la comparación de estimados independientes de σ^2 , obtenidos al separar la suma de cuadrados total de los datos en 4 componentes mediante la siguiente identidad.

Partición de la variabilidad en el caso de dos factores

Teorema 14.1: Identidad de la suma de cuadrados

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2 &= bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 + an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &\quad + n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2 \end{aligned}$$

Simbólicamente, la identidad de la suma de cuadrados se escribe como

$$SCT = SCA + SCB + SC(AB) + SCE$$

donde a SCA y SCB se les denomina la suma de cuadrados para los efectos principales A y B , respectivamente, $SC(AB)$ recibe el nombre de suma de cuadrados de la interacción para A y B , y SCE es la suma de cuadrados del error. La partición de los grados de libertad se efectúa de acuerdo con la identidad

$$abn - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1) + ab(n - 1).$$

Formación de los cuadrados medios

Si dividimos cada una de las sumas de cuadrados en el lado derecho de la identidad de la suma de cuadrados entre su número correspondiente de grados de libertad, obtenemos los cuatro estadísticos

$$S_1^2 = \frac{SCA}{a - 1}, \quad S_2^2 = \frac{SCB}{b - 1}, \quad S_3^2 = \frac{SC(AB)}{(a - 1)(b - 1)}, \quad S^2 = \frac{SCE}{ab(n - 1)}.$$

Todos estos estimados de la varianza son estimados independientes de σ^2 , siempre que no haya efectos α_i , β_j ni, por supuesto, $(\alpha\beta)_{ij}$. Si las sumas de cuadrados se interpretan

como funciones de las variables aleatorias independientes $y_{111}, y_{112}, \dots, y_{abn}$, no es difícil comprobar que

$$\begin{aligned} E(S_1^2) &= E\left[\frac{SCA}{a-1}\right] = \sigma^2 + \frac{nb}{a-1} \sum_{i=1}^a \alpha_i^2, \\ E(S_2^2) &= E\left[\frac{SCB}{b-1}\right] = \sigma^2 + \frac{na}{b-1} \sum_{j=1}^b \beta_j^2, \\ E(S_3^2) &= E\left[\frac{SC(AB)}{(a-1)(b-1)}\right] = \sigma^2 + \frac{n}{(a-1)(b-1)} \sum_{i=1}^a \sum_{j=1}^b (\alpha\beta)_{ij}^2, \\ E(S^2) &= E\left[\frac{SCE}{ab(n-1)}\right] = \sigma^2, \end{aligned}$$

a partir de lo cual se observa de inmediato que los 4 estimados de σ^2 son no sesgados cuando H'_0 , H''_0 , y H'''_0 son verdaderas.

Para probar la hipótesis H'_0 , de que los efectos de los factores A son todos iguales a cero, se calcula la siguiente razón:

Prueba F para
el factor A

$$f_1 = \frac{s_1^2}{s^2},$$

que es un valor de la variable aleatoria F_1 , el cual tiene la distribución F con $a-1$ y $ab(n-1)$ grados de libertad cuando H'_0 es verdadera. La hipótesis nula se rechaza al nivel de significancia α cuando $f_1 > f_\alpha[a-1, ab(n-1)]$.

De manera similar, para probar la hipótesis H''_0 , de que todos los efectos del factor B son iguales a cero, se calcula la razón:

Prueba F para
el factor B

$$f_2 = \frac{s_2^2}{s^2},$$

que es un valor de la variable aleatoria F_2 que tiene la distribución F con $b-1$ y $ab(n-1)$ grados de libertad cuando H''_0 es verdadera. Esta hipótesis se rechaza al nivel de significancia α cuando $f_2 > f_\alpha[b-1, ab(n-1)]$.

Por último, para probar la hipótesis H'''_0 , de que todos los efectos de interacción son iguales a 0, se calcula la razón siguiente:

Prueba F para
la interacción

$$f_3 = \frac{s_3^2}{s^2},$$

que es un valor de la variable aleatoria F_3 , el cual tiene la distribución F con $(a-1)(b-1)$ y $ab(n-1)$ grados de libertad cuando H'''_0 es verdadera. Concluimos que, a un nivel de significancia α , hay interacción cuando $f_\alpha[(a-1)(b-1), ab(n-1)]$.

Como se indicó en la sección 14.2, se recomienda interpretar la prueba para la interacción antes de tratar de hacer inferencias sobre los efectos principales. Si la interacción no es significativa, entonces hay evidencia de que las pruebas sobre los efectos principales son interpretables. El rechazo de la hipótesis 1 de la página 566 implica que las medias de la respuesta en los niveles del factor A difieren significativamente, mientras que

el rechazo de la hipótesis 2 implica una condición similar para las medias en los niveles del factor B . Sin embargo, una interacción significativa podría muy bien implicar que los datos se deberían analizar de una manera un poco diferente, **quizá observando el efecto del factor A en niveles fijos del factor B** , y así sucesivamente.

Los cálculos en un problema de análisis de varianza para un experimento de 2 factores con n réplicas suelen resumirse como se ilustra en la tabla 14.2.

Tabla 14.2: Análisis de varianza para el experimento de 2 factores con n réplicas

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Efecto principal				
A	SCA	$a - 1$	$s_1^2 = \frac{SCA}{a-1}$	$f_1 = \frac{s_1^2}{s^2}$
B	SCB	$b - 1$	$s_2^2 = \frac{SCB}{b-1}$	$f_2 = \frac{s_2^2}{s^2}$
Interacciones de 2 factores				
AB	$SC(AB)$	$(a - 1)(b - 1)$	$s_3^2 = \frac{SC(AB)}{(a-1)(b-1)}$	$f_3 = \frac{s_3^2}{s^2}$
Error	SCE	$ab(n - 1)$	$s^2 = \frac{SCE}{ab(n-1)}$	
Total	STC	$abn - 1$		

Ejemplo 14.1: En un experimento realizado para determinar cuál de 3 sistemas de misiles distintos es preferible, se midió la tasa de combustión del propulsor para 24 arranques estáticos. Se emplearon 4 tipos de combustible diferentes y el experimento generó observaciones duplicadas de las tasas de combustión para cada combinación de los tratamientos.

Los datos, ya codificados, se presentan en la tabla 14.3. Pruebe las siguientes hipótesis: $a) H'_0$: no hay diferencia en las tasas medias de combustión del propulsor cuando se emplean diferentes sistemas de misiles, $b) H''_0$: no existe diferencia en las tasas medias de combustión de los 4 tipos de propulsor, $c) H'''_0$: no hay interacción entre los distintos sistemas de misiles y los diferentes tipos de propulsor.

Tabla 14.3: Tasas de combustión del propulsor

Sistema de misiles	Tipo de propulsor			
	b_1	b_2	b_3	b_4
a_1	34.0	30.1	29.8	29.0
	32.7	32.8	26.7	28.9
a_2	32.0	30.2	28.7	27.6
	33.2	29.8	28.1	27.8
a_3	28.4	27.3	29.7	28.8
	29.3	28.9	27.3	29.1

Solución: 1. $a) H'_0$: $\alpha_1 = \alpha_2 = \alpha_3 = 0$.
 $b) H''_0$: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$.
 $c) H'''_0$: $(\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{34} = 0$.

2. a) H'_1 : Al menos una de las α_i no es igual a 0.
 b) H''_1 : Al menos una de las β_j no es igual a 0.
 c) H'''_1 : Al menos una de las $(\alpha\beta)_{ij}$ no es igual a 0.

Se utiliza la fórmula de la suma de cuadrados que se describió en el teorema 14.1. En la tabla 14.4 se presenta el análisis de varianza.

Tabla 14.4: Análisis de varianza para los datos de la tabla 14.3

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Sistema de misiles	14.52	2	7.26	5.84
Tipo de propulsor	40.08	3	13.36	10.75
Interacción	22.16	6	3.69	2.97
Error	14.91	12	1.24	
Total	91.68	23		

Se remite al lector al procedimiento de modelos lineales generales (GLM, por sus siglas en inglés) del SAS para el análisis de los datos de la tasa de combustión de la figura 14.2. Observe la forma en que al principio se prueba el “modelo” (11 grados de libertad), y por separado se prueban el sistema, el tipo y el sistema por tipo de interacción. La prueba f en el modelo ($P = 0.0030$) prueba la acumulación de los 2 efectos principales y la interacción.

- a) Rechace H'_0 y concluya que los distintos sistemas de misiles resultan en diferentes tasas medias de combustión del propulsor. El valor P es de aproximadamente 0.0169.
 b) Rechace H''_0 y concluya que las tasas medias de combustión del propulsor no son las mismas para los 4 tipos de propulsores. El valor P es de aproximadamente 0.0010.
 c) La interacción es apenas insignificante al nivel 0.05, pero el valor P de aproximadamente 0.0513 indicaría que la interacción debe tomarse en serio.

En este momento debemos hacer algún tipo de interpretación de la interacción. Debe destacarse que la significancia estadística de un efecto principal tan sólo implica que las *medias marginales son significativamente diferentes*. Sin embargo, considere la tabla de promedios de 2 factores de la tabla 14.5.


Tabla 14.5: Interpretación de la interacción

	b_1	b_2	b_3	b_4	Promedio
a_1	33.35	31.45	28.25	28.95	30.50
a_2	32.60	30.00	28.40	27.70	29.68
a_3	28.85	28.10	28.50	28.95	28.60
Promedio	31.60	29.85	28.38	28.53	

Es evidente que hay más información importante en el cuerpo de la tabla, tendencias que son inconsistentes con la tendencia que describe los promedios marginales. La tabla 14.5 sugiere con certeza que el efecto del tipo de propulsor depende del sistema que se utiliza. Por ejemplo, para el sistema 3, el efecto del tipo de propulsor no parece ser

The GLM Procedure					
Dependent Variable: rate					
		Sum of			
Source	DF	Squares	Mean Square	F Value	Pr > F
Model	11	76.76833333	6.97893939	5.62	0.0030
Error	12	14.91000000	1.24250000		
Corrected Total	23	91.67833333			
R-Square	Coeff Var	Root MSE	rate Mean		
0.837366	3.766854	1.114675	29.59167		
Source	DF	Type III SS	Mean Square	F Value	Pr > F
system	2	14.52333333	7.26166667	5.84	0.0169
type	3	40.08166667	13.36055556	10.75	0.0010
system*type	6	22.16333333	3.69388889	2.97	0.0512

Figura 14.2: Salida de resultados del SAS para el análisis de los datos de la tasa de combustión del propulsor de la tabla 14.3.

importante, aunque tiene un efecto grande si se utiliza el sistema 1 o el 2. Esto explica la interacción “significativa” entre esos 2 factores. Más adelante se revelará más información acerca de esta interacción. 

Ejemplo 14.2: Remítase al ejemplo 14.1 y elija 2 contrastes ortogonales para dividir la suma de cuadrados del sistema de misiles en componentes con un solo grado de libertad, los cuales utilizará para comparar los sistemas 1 y 2 con el 3, y el sistema 1 contra el sistema 2.

Solución: El contraste para comparar los sistemas 1 y 2 con el 3 es

$$w_1 = \mu_1. + \mu_2. - 2\mu_{3.}.$$

Un segundo contraste, ortogonal a w_1 , para comparar el sistema 1 con el 2, es dado por $w_2 = \mu_{1.} - \mu_{2.}$. Las sumas de cuadrados con un solo grado de libertad son

$$SCw_1 = \frac{[244.0 + 237.4 - (2)(228.8)]^2}{(8)[(1)^2 + (1)^2 + (-2)^2]} = 11.80$$

y

$$SCw_2 = \frac{(244.0 - 237.4)^2}{(8)[(1)^2 + (-1)^2]} = 2.72.$$

Observe que $SCw_1 + SCw_2 = SCA$, como se esperaba. Los valores f calculados correspondientes a w_1 y w_2 son, respectivamente,

$$f_1 = \frac{11.80}{1.24} = 9.5 \quad \text{y} \quad f_2 = \frac{2.72}{1.24} = 2.2.$$

Al comparar con el valor crítico $f_{0.05}(1, 12) = 4.75$, se encuentra que f_1 es significativo. De hecho, el valor P es menor que 0.01. Así, el primer contraste indica que se rechaza la hipótesis

$$H_0: \frac{1}{2}(\mu_1. + \mu_2.) = \mu_3.$$

Como $f_2 < 4.75$, las tasas medias de combustión del primer y segundo sistemas no son significativamente diferentes. ■

Impacto de la interacción significativa en el ejemplo 14.1

Si la hipótesis de que no hay interacción en el ejemplo 14.1 es verdadera, podríamos hacer las comparaciones *generales* del ejemplo 14.2 relacionado con los sistemas de misiles, en lugar de comparaciones separadas para cada propulsor. De manera similar, se podrían realizar comparaciones generales entre los propulsores, en vez de comparar por separado cada sistema de misiles. Por ejemplo, se podrían comparar los propulsores 1 y 2 con el 3 y 4, y también el 1 contra el 2. Las razones f resultantes, cada una con 1 y 12 grados de libertad, resultan ser de 24.81 y 7.39, respectivamente, y ambas son muy significativas al nivel 0.05.

Por los promedios de los propulsores, parece haber evidencia de que el 1 ofrece la tasa media de combustión más alta. Un experimentador prudente sería cauteloso al sacar conclusiones generales en un problema como éste, donde la razón f de la interacción está apenas por debajo del valor crítico de 0.05. Por ejemplo, la evidencia general, 31.60 contra 29.85 sobre el promedio para los 2 propulsores, indica con claridad que el 1 es superior al 2, en términos de una mayor tasa de combustión. Sin embargo, si nos restringimos al sistema 3, donde tenemos un promedio de 28.85 para el propulsor 1 en oposición a 28.10 para el propulsor 2, parece haber una diferencia mínima o incluso ninguna entre estos 2 propulsores. De hecho, parece que hay una estabilización de las tasas de combustión para los distintos propulsores si se opera con el sistema 3. Es claro que existe evidencia general que indica que el sistema 1 ofrece una tasa de combustión más alta que el sistema 3, pero parece que esta conclusión no se sostiene si nos restringimos al propulsor 4.

Para recabar evidencias concluyentes de que la interacción está *produciendo dificultades considerables en la obtención de conclusiones generales sobre los efectos principales*, el analista puede hacer una prueba t sencilla utilizando las tasas de combustión promedio del sistema 3. Considere una comparación del propulsor 1 contra el 2 usando únicamente el sistema 3. Se toma prestado un estimado de σ^2 del análisis general, es decir, se utiliza $s^2 = 1.24$ con 12 grados de libertad, y se obtiene

$$|t| = \frac{0.75}{\sqrt{2s^2/n}} = \frac{0.75}{\sqrt{1.24}} = 0.67,$$

que no se acerca a ser significativa. Esta ilustración sugiere que, en presencia de interacción, debería tenerse cautela con la interpretación estricta de los efectos principales.

Análisis gráfico para el problema de dos factores del ejemplo 14.1

Muchos de los mismos tipos de ilustraciones gráficas que se sugirió emplear en los problemas de un factor también se aplican en el caso de 2 factores. Las gráficas en 2 dimensiones de las medias de las celdas o de las medias de las combinaciones de tratamientos ofrecen información sobre la presencia de interacciones entre los 2 factores.

Además, una gráfica de los residuales contra los valores ajustados bien podría indicar si se cumple o no la suposición de la varianza homogénea. Por supuesto, es frecuente que una violación de la suposición de varianza homogénea implique un aumento en la varianza del error conforme *los números de la respuesta se vuelven más grandes*. Como resultado, esta gráfica podría resaltar la violación.

La figura 14.3 presenta la gráfica de las medias de las celdas para el caso del propulsor de los sistemas de misiles del ejemplo 14.1. Observe gráficamente (en este caso) cuánta falta de paralelismo hay. Note el aplanamiento de la parte de la figura que indica el efecto del propulsor para el sistema 3. Esto ilustra la interacción entre los factores. La figura 14.4 muestra la gráfica de los residuales contra los valores ajustados para los mismos datos. Al parecer no hay dificultades con la suposición de la varianza homogénea.

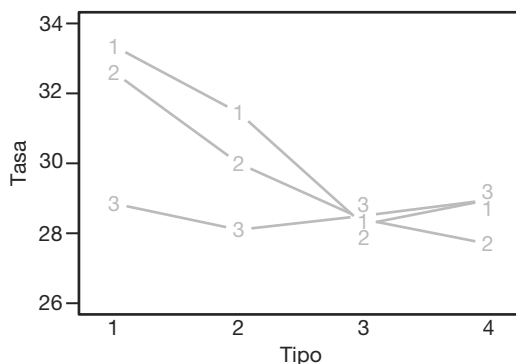


Figura 14.3: Gráfica de las medias de las celdas para los datos del ejemplo 14.1. Los números representan los sistemas de misiles.

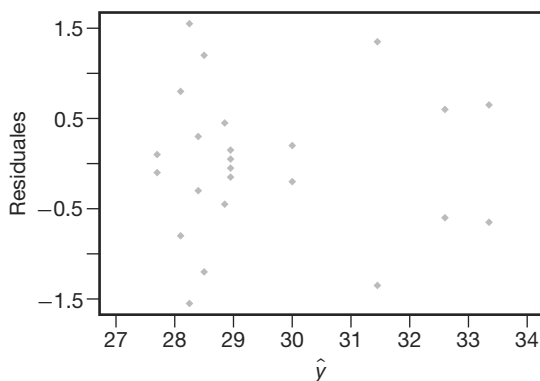


Figura 14.4: Gráfica de los residuales de los datos del ejemplo 14.1.

Ejemplo 14.3: Un ingeniero eléctrico investiga un proceso de grabado con plasma que se emplea en la fabricación de semiconductores. Es de interés estudiar los efectos de 2 factores, la cantidad de flujo (A) del gas C_2F_6 y la potencia aplicada al cátodo (B). La respuesta es la velocidad de grabado. Cada factor se aplica a 3 niveles y se hacen 2 corridas experimentales sobre la velocidad de grabado para cada una de las 9 combinaciones. El plan representa un diseño completamente aleatorizado. En la tabla 14.6 se presentan los datos. La velocidad de grabado se expresa en A°/min .

Tabla 14.6: Datos para el ejemplo 14.3

Cantidad de flujo del C_2F_6	Potencia suministrada		
	1	2	3
1	288	488	670
	360	465	720
2	385	482	692
	411	521	724
3	488	595	761
	462	612	801

Los niveles de los factores están en orden ascendente, donde el nivel 1 es el más bajo y el 3 el más alto.

- a) Elabore una tabla de análisis de varianza y saque conclusiones; empiece con la prueba de interacción.
- b) Haga pruebas sobre los efectos principales y saque conclusiones.

Solución: En la figura 14.5 se muestra una salida de resultados por computadora del SAS. De ese listado se concluye lo siguiente.

The GLM Procedure					
Dependent Variable: etchrates					
Sum of					
Source	DF	Squares	Mean Square	F Value	Pr > F
Model	8	379508.7778	47438.5972	61.00	<.0001
Error	9	6999.5000	777.7222		
Corrected Total	17	386508.2778			
R-Square					
0.981890					
Coeff Var					
5.057714					
Root MSE					
27.88767					
etchrates Mean					
551.3889					
Source					
DF					
Type III SS					
Mean Square					
F Value					
Pr > F					
c2f6	2	46343.1111	23171.5556	29.79	0.0001
power	2	330003.4444	165001.7222	212.16	<.0001
c2f6*power	4	3162.2222	790.5556	1.02	0.4485

Figura 14.5: Una salida de resultados por computadora del SAS para el ejemplo 14.3.

- a) El valor P para la prueba de interacción es 0.4485. Se concluye que la interacción no es significativa.
- b) Existe una diferencia significativa en la velocidad media de grabado para los 3 niveles de la velocidad de flujo del C_2F_6 . Una prueba de Duncan muestra que la velocidad media de grabado para el nivel 3 es significativamente mayor que para el nivel 2, y la

velocidad para el nivel 2 es significativamente mayor que para el nivel 1. Véase la figura 14.6a.

Existe una diferencia significativa en la velocidad media de grabado basada en el nivel de potencia al cátodo. Una prueba de Duncan revela que la velocidad de grabado para el nivel 3 es significativamente más alta que para el 2, y que la velocidad para el nivel 2 es significativamente más alta que para el 1. Véase la figura 14.6b.

Duncan Grouping	Mean	N	c2f6	Duncan Grouping	Mean	N	power
A	619.83	6	3	A	728.00	6	3
B	535.83	6	2	B	527.17	6	2
C	498.50	6	1	C	399.00	6	1
(a)				(b)			

Figura 14.6: Una salida de resultados por computadora del SAS para el ejemplo 14.3. a) Prueba de Duncan de la cantidad de flujo del gas; b) Prueba de Duncan de la potencia.

Ejercicios

14.1 Se realizó un experimento para estudiar los efectos de la temperatura y el tipo de horno sobre la vida de un componente en particular. En el experimento se utilizaron 4 tipos de horno y 3 niveles de temperatura. Se asignaron 24 piezas al azar, 2 para cada combinación de tratamientos y se registraron los siguientes resultados.

Temperatura (°F)	Horno			
	O ₁	O ₂	O ₃	O ₄
500	227	214	225	260
	221	259	236	229
550	187	181	232	246
	208	179	198	273
600	174	198	178	206
	202	194	213	219

A un nivel de significancia de 0.05 pruebe las hipótesis de que

- las diferentes temperaturas no tienen efecto en la vida del componente;
- los diferentes hornos no tienen efecto en la vida del componente;
- no hay interacción entre el tipo de horno y la temperatura.

14.2 El Departamento de Nutrición Humana y Alimentos del Virginia Tech realizó un estudio titulado *Vitamin C Retention in Reconstituted Frozen Orange Juice* sobre la estabilidad de la vitamina C en el concentrado de jugo de naranja congelado reconstituido y almacenado en un refrigerador durante un periodo de hasta una semana. Se probaron 3 tipos de concentrado de jugo de naranja congelado en 3 periodos distintos, los cuales se refieren al número de días transcurridos desde que se mezcló el jugo hasta que se probó. Se

registraron los resultados en miligramos de ácido ascórbico por litro. Utilice un nivel de significancia de 0.05 para probar las hipótesis de que

- no hay diferencias en el contenido de ácido ascórbico entre las diferentes marcas de concentrado de jugo de naranja;
- no existen diferencias en el contenido de ácido ascórbico para distintos periodos;
- no hay interacción entre las marcas de concentrado de jugo de naranja y el número de días transcurridos desde que el jugo se mezcló hasta que se probó.

Marca	Tiempo (días)					
	0		3		7	
Richfood	52.6	54.2	49.4	49.2	42.7	48.8
	49.8	46.5	42.8	53.2	40.4	47.6
Sealed-Sweet	56.0	48.0	48.8	44.0	49.2	44.0
	49.6	48.4	44.0	42.4	42.0	43.2
Minute Maid	52.5	52.0	48.0	47.0	48.5	43.3
	51.8	53.6	48.2	49.6	45.2	47.6

14.3 Se estudió el desempeño de 3 cepas de ratas en una prueba de laberintos en 2 condiciones ambientales diferentes. Se registraron las puntuaciones de error de las 48 ratas:

Ambiente	Cepa					
	Brillante		Mezclada		Torpe	
Libre	28	12	33	83	101	94
	22	23	36	14	33	56
	25	10	41	76	122	83
	36	86	22	58	35	23
Restringido	72	32	60	89	136	120
	48	93	35	126	38	153
	25	31	83	110	64	128
	91	19	99	118	87	140

Utilice un nivel de significancia de 0.01 para probar las hipótesis de que

- a) no hay diferencia en las puntuaciones de error para ambientes diferentes;
- b) no existe diferencia en las puntuaciones de error para cepas diferentes;
- c) no hay interacción entre los ambientes y las cepas de las ratas.

14.4 La fatiga por corrosión de los metales se define como la acción simultánea de tensión cíclica y ataque químico sobre una estructura metálica. Una técnica muy utilizada para minimizar el daño de la fatiga por corrosión en el aluminio requiere la aplicación de un recubrimiento protector. En un estudio efectuado por el Departamento de Ingeniería Mecánica de Virginia Tech se utilizaron 3 niveles diferentes de humedad:

Bajo: 20 a 25% de humedad relativa

Medio: 55 a 60% de humedad relativa

Alto: 86 a 91% de humedad relativa

y 3 tipos de recubrimiento:

No revestido: Sin recubrimiento

Anodizado: Recubrimiento de óxido anódico por ácido sulfúrico

Conversión: Recubrimiento por conversión química de cromato.

Los datos de fatiga por corrosión, expresados en miles de ciclos hasta que se presenta la falla, se registraron como sigue:

Recubrimiento	Humedad relativa					
	Baja		Media		Alta	
No revestido	361	469	314	522	1344	1216
	466	937	244	739	1027	1097
	1069	1357	261	134	1011	1011
Anodizado	114	1032	322	471	78	466
	1236	92	306	130	387	107
	533	211	68	398	130	327
Conversión	130	1482	252	874	586	524
	841	529	105	755	402	751
	1595	754	847	573	846	529

- a) Lleve a cabo un análisis de varianza con $\alpha = 0.05$ para probar si existen efectos principales y efectos de interacción.
- b) Utilice la prueba de Duncan de rango múltiple a un nivel de significancia de 0.05 para determinar cuáles niveles de humedad relativa dan como resultado daños distintos de fatiga por corrosión.

14.5 Para determinar cuáles músculos necesitan someterse a un programa de acondicionamiento para mejorar el rendimiento individual en el servicio tendido que se usa en el tenis, el Departamento de Salud, Educación Física y Recreación de Virginia Tech realizó

un estudio de 5 músculos diferentes:

- 1: deltoides anterior
- 2: pectoral mayor
- 3: deltoides posterior
- 4: deltoides medio
- 5: tríceps

los cuales se probaron en cada uno de 3 sujetos; el experimento se efectuó 3 veces para cada combinación de tratamiento. Los datos electromiográficos que se registraron durante el servicio se presentan a continuación.

Sujeto	Músculo				
	1	2	3	4	5
1	32	5	58	10	19
	59	1.5	61	10	20
	38	2	66	14	23
2	63	10	64	45	43
	60	9	78	61	61
	50	7	78	71	42
3	43	41	26	63	61
	54	43	29	46	85
	47	42	23	55	95

Utilice un nivel de significancia de 0.01 para probar las hipótesis de que

- a) diferentes sujetos tienen medidas electromiográficas iguales;
- b) los diferentes músculos no tienen un efecto en las medidas electromiográficas;
- c) no hay interacción entre los sujetos y los tipos de músculos.

14.6 Se realizó un experimento para determinar si los aditivos incrementan la adherencia de productos de caucho. Se elaboraron 16 productos con el aditivo nuevo y otros 16 sin dicho aditivo. Se registró la siguiente adherencia.

	Temperatura (°C)			
	50	60	70	80
Sin el aditivo	2.3	3.4	3.8	3.9
	2.9	3.7	3.9	3.2
	3.1	3.6	4.1	3.0
	3.2	3.2	3.8	2.7
Con el aditivo	4.3	3.8	3.9	3.5
	3.9	3.8	4.0	3.6
	3.9	3.9	3.7	3.8
	4.2	3.5	3.6	3.9

Haga un análisis de varianza para probar la existencia de efectos principales y de interacción significativos.

14.7 Se sabe que la velocidad de extracción de cierto polímero depende de la temperatura de reacción y de la cantidad de catalizador empleada. Se hizo un experimento en 4 niveles de temperatura y 5 niveles de catalizador, y se registró la velocidad de extracción en la siguiente tabla:

	Cantidad de catalizador				
	0.5%	0.6%	0.7%	0.8%	0.9%
50°C	38	45	57	59	57
	41	47	59	61	58
60°C	44	56	70	73	61
	43	57	69	72	58
70°C	44	56	70	73	61
	47	60	67	61	59
80°C	49	62	70	62	53
	47	65	55	69	58

Realice un análisis de varianza. Pruebe si hay efectos principales y de interacción significativos.

14.8 En Myers, Montgomery y Anderson-Cook (2009) se estudia un escenario donde se describe un proceso de laminado por prensado. La respuesta es el espesor del material. Los factores que podrían afectar el espesor incluyen la cantidad de níquel (*A*) y el pH (*B*). Se diseñó un experimento con 2 factores. El plan es un diseño completamente aleatorizado en el que las prensas individuales se asignan al azar a las combinaciones de factores. En el experimento se utilizan 3 niveles de pH y 2 niveles de contenido de níquel. Los espesores, en $\text{cm} \times 10^{-3}$, son los siguientes:

Contenido de níquel (gramos)	pH		
	6	5.5	6
18	250	211	221
	195	172	150
	188	165	170
10	115	88	69
	165	112	101
	142	108	72

- Elabore la tabla del análisis de varianza con pruebas para los efectos principales y de interacción. Incluya valores *P*.
- Saque conclusiones para ingeniería. ¿Qué aprendió del análisis de estos datos?
- Elabore una gráfica que ilustre la presencia o ausencia de interacción.

14.9 Un ingeniero está interesado en los efectos de la velocidad de corte y la geometría de la herramienta sobre las horas de vida de una máquina-herramienta. Se utilizan 2 velocidades de corte y 2 geometrías distintas. Se llevan a cabo 3 pruebas experimentales con cada una de las 4 combinaciones. Los datos son los siguientes:

Geometría de la herramienta	Velocidad de corte					
	Baja			Alta		
1	22	28	20	34	37	29
2	18	15	16	11	10	10

- Calcule la tabla del análisis de varianza con pruebas sobre los efectos principales y de interacción.
- Haga comentarios sobre el efecto que tiene la interacción sobre la prueba de la velocidad de corte.

- Efectúe pruebas secundarias que permitan al ingeniero conocer el verdadero impacto de la velocidad de corte.
- Construya una gráfica que ilustre el efecto de interacción.

14.10 En un experimento se estudiaron 2 factores de un proceso de manufactura de un circuito integrado. El propósito del experimento es conocer su efecto sobre la resistividad de las obleas de silicio. Los factores son la dosis del implante (2 niveles) y la posición de la caldera (3 niveles). El experimento es costoso, por lo que sólo se hizo una corrida con cada combinación. Los datos son los siguientes.

Dosis	Posición		
1	15.5	14.8	21.3
2	27.2	24.9	26.1

Se supone que no hay interacción entre esos 2 factores.

- Escriba el modelo y explique sus términos.
- Elabore la tabla de análisis de varianza.
- Explique los 2 grados de libertad del “error”.
- Use una prueba de Tukey para hacer pruebas de comparaciones múltiples sobre la posición de la caldera. Explique qué es lo que muestran los resultados.

14.11 Se realizó un estudio para determinar la influencia de 2 factores, el método de análisis y el laboratorio que hace el análisis, sobre el nivel de contenido de azufre del carbón. Se asignaron al azar 28 especímenes de carbón a 14 combinaciones de factores, la estructura de las unidades experimentales representada por las combinaciones de 7 laboratorios y 2 métodos de análisis con 2 especímenes por combinación de factores. Los datos, expresados en porcentaje de azufre, son los siguientes.

Laboratorio	Método			
	1		2	
1	0.109	0.105	0.105	0.108
2	0.129	0.122	0.127	0.124
3	0.115	0.112	0.109	0.111
4	0.108	0.108	0.117	0.118
5	0.097	0.096	0.110	0.097
6	0.114	0.119	0.116	0.122
7	0.155	0.145	0.164	0.160

(Los datos se tomaron de G. Taguchi, “Signal to Noise Ratio and Its Applications to Testing Material”, *Reports of Statistical Application Research*, Union of Japanese Scientists and Engineers, Vol. 18, Núm. 4, 1971).

- Haga un análisis de varianza y exprese los resultados en la tabla correspondiente.
- ¿Es significativa la interacción? Si lo es, analice lo que significa para el científico. Utilice un valor *P* en sus conclusiones.

- c) ¿Son estadísticamente significativos los efectos principales individuales, el laboratorio y el método de análisis? Analice la información y lo que aprendió y base su respuesta en el contexto de cualquier interacción significativa.
- d) Dibuje una gráfica de interacción que ilustre el efecto de la interacción.
- e) Efectúe una prueba para comparar los métodos 1 y 2 en el laboratorio 1, y haga lo mismo para el laboratorio 7. Comente lo que revelan esos resultados.

14.12 En un experimento efectuado en el departamento de Ingeniería Civil de Virginia Tech se observó el crecimiento que cierto tipo de alga tenía en el agua, en función del tiempo y la dosis de cobre que se agregaba al líquido. Los datos se presentan a continuación. La respuesta se expresa en unidades de algas.

Cobre	Tiempo en días		
	5	12	18
1	0.30	0.37	0.25
	0.34	0.36	0.23
	0.32	0.35	0.24
2	0.24	0.30	0.27
	0.23	0.32	0.25
	0.22	0.31	0.25
3	0.20	0.30	0.27
	0.28	0.31	0.29
	0.24	0.30	0.25

- a) Haga un análisis de varianza y elabore la tabla correspondiente.
- b) Comente acerca de si los datos son suficientes para mostrar un efecto del tiempo sobre la concentración de algas.
- c) Haga lo mismo para el contenido de cobre. ¿El nivel de contenido de cobre tiene algún efecto sobre la concentración de algas?
- d) Comente los resultados de la prueba de interacción. ¿Cómo influye el tiempo sobre el efecto del contenido de cobre?

14.13 En *Classical and Modern Regression with Applications* (Duxbury Classic Series, 2a. ed., 1990), de Myers, se describe un experimento en el que la Agencia de Protección Ambiental busca determinar el efecto de 2 métodos de tratamiento de aguas sobre la absorción del magnesio. Se miden los niveles de magnesio, en gramos por centímetro cúbico (cc) y se incorporan 2 niveles diferentes de tiempo al experimento. Los datos son los siguientes:

Tiempo (horas)	Tratamiento					
	1			2		
1	2.19	2.15	2.16	2.03	2.01	2.04
2	2.01	2.03	2.04	1.88	1.86	1.91

- a) Dibuje una gráfica de la interacción. ¿Cuál es su impresión?

- b) Efectúe un análisis de varianza y presente pruebas para los efectos principales y de interacción.
- c) Mencione los hallazgos científicos acerca de cómo influyen el tiempo y el tratamiento en la absorción del magnesio.
- d) Ajuste el modelo de regresión adecuado usando el tratamiento como variable categórica. Incluya la interacción en el modelo.
- e) ¿La interacción es significativa en el modelo de regresión?

14.14 Considere los datos del ejercicio 14.12 y responda las siguientes preguntas.

- a) Ambos factores, el cobre y el tiempo, son cuantitativos. Como resultado, podría ser de interés un modelo de regresión. Describa cuál sería un modelo adecuado si se usa x_1 = contenido de cobre y x_2 = tiempo. Ajuste el modelo a los datos mostrando los coeficientes de regresión y haga una prueba t para cada uno.
- b) Ajuste el modelo

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \epsilon,$$

y compárelo con el que eligió en el inciso a. ¿Cuál es más apropiado? Como criterio utilice R^2_{ajus} .

14.15 El propósito del estudio *The Incorporation of a Chelating Agent into a Flame Retardant Finish of a Cotton Flannelette and the Evaluation of Selected Fabric Properties*, llevado a cabo en Virginia Tech, fue evaluar el uso de un agente quelante como parte del acabado retardante del fuego de la franela de algodón, determinando sus efectos en la inflamabilidad después de lavar la tela en condiciones específicas. Se utilizaron 2 tratamientos con 2 niveles; se prepararon 2 baños, uno con celulosa de carboximetilo (baño I) y otro sin ella (baño II). La mitad de la tela se lavó 5 veces y la otra mitad se lavó 10 veces. Hubo 12 pedazos de tela en cada combinación de baño/número de lavados. Después de los lavados se midieron las longitudes quemadas de la tela, así como los tiempos de combustión. Se registraron los siguientes tiempos de combustión (en segundos):

Lavados	Baño I			Baño II		
5	13.7	23.0	15.7	6.2	5.4	5.0
	25.5	15.8	14.8	4.4	5.0	3.3
	14.0	29.4	9.7	16.0	2.5	1.6
	14.0	12.3	12.3	3.9	2.5	7.1
10	27.2	16.8	12.9	18.2	8.8	14.5
	14.9	17.1	13.0	14.7	17.1	13.9
	10.8	13.5	25.5	10.6	5.8	7.3
	14.2	27.4	11.5	17.7	18.3	9.9

- a) Realice un análisis de varianza. ¿Existe un término de interacción significativo?
- b) ¿Se encontraron diferencias en los efectos principales? Analice la información.