

## 14-2.3 Método exacto

En situaciones en que los métodos aproximados no son apropiados, como cuando ocurren celdas vacías (algunas  $n_{ij} = 0$ ) o cuando las  $n_{ij}$  presentan diferencias radicales, el experimentador debe usar un análisis exacto. El enfoque utilizado para desarrollar las sumas de cuadrados para probar los efectos principales y las interacciones consiste en representar el modelo del análisis de varianza como un **modelo de regresión**, ajustar ese modelo a los datos y usar el enfoque de la prueba general de significación de la regresión. Sin embargo, existen varias formas en que puede hacerse esto, y estos métodos pueden producir valores diferentes para las sumas de cuadrados. Además, las hipótesis que se están probando no siempre son análogos directos de las del caso balanceado, y su interpretación tampoco es siempre sencilla. Para mayor información al respecto, ver el material suplementario del texto de éste capítulo. Otras buenas referencias son Searle [99a]; Speed y Hocking [105]; Hocking y Speed [58]; Hocking, Hackney y Speed [57]; Speed, Hocking y Hackney [106]; Searle, Speed y Henderson [102]; Searle [99c]; y Milliken y Johnson [79]. El software de estadística SAS proporciona un excelente enfoque del análisis de datos no balanceados a través del procedimiento PROC GLM.

## 14-3 ANÁLISIS DE COVARIANZA

En los capítulos 2 y 4 se introdujo el uso del principio de la formación de bloques para mejorar la precisión con la que se hacen comparaciones entre tratamientos. La prueba  $t$  pareada fue el procedimiento ilustrado en el capítulo 2, mientras que en el capítulo 4 se presentó el diseño de bloques aleatorizados. En general, el **principio de la formación de bloques** puede usarse para eliminar el efecto de los factores perturbadores controlables. El **análisis de covarianza** es otra técnica que en ocasiones es útil para mejorar la precisión de un experimento. Suponga que en un experimento con una variable de respuesta y existe otra variable, por ejemplo  $x$ , y que  $y$  se relaciona linealmente con  $x$ . Además, suponga que  $x$  no puede ser controlada por el experimentador, pero puede observarse junto con  $y$ . A la variable  $x$  se le llama **variable concomitante** o **covariable**. El análisis de covarianza implica ajustar la variable de respuesta observada para el efecto de la variable concomitante. Si no se hace este ajuste, la variable concomitante podría inflar el cuadrado medio del error y hacer que sean más difíciles de detectar las verdaderas diferencias en la respuesta debidas a los tratamientos. Por lo tanto, el análisis de covarianza es un método de ajuste para los efectos de una variable perturbadora no controlable. Como se verá, el procedimiento es una combinación del análisis de varianza y del análisis de regresión.

Como un ejemplo de un experimento en el que puede emplearse el análisis de covarianza, considere el estudio realizado para determinar si existe una diferencia en la resistencia de una fibra de monofila-

Tabla 14-8 Datos de la resistencia a la ruptura ( $y$  = resistencia en libras y  $x$  = diámetro en  $10^{-3}$  pulgadas)

Máquina 1		Máquina 2		Máquina 3	
$y$	$x$	$y$	$x$	$y$	$x$
36	20	40	22	35	21
41	25	48	28	37	23
39	24	39	22	42	26
42	25	45	30	34	21
49	32	44	28	32	15
207	126	216	130	180	106

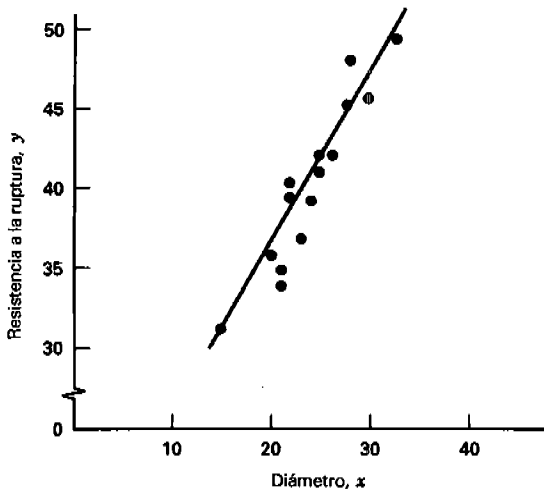


Figura 14-3 Resistencia a la ruptura ( $y$ ) contra el diámetro de la fibra ( $x$ ).

mento producida por tres máquinas diferentes. Los datos de este experimento se muestran en la tabla 14-8. En la figura 14-3 se presenta un diagrama de dispersión de la resistencia ( $y$ ) contra el diámetro (o grosor) de la muestra. Evidentemente, la resistencia de la fibra también se afecta por su grosor; por consiguiente, una fibra más gruesa será por lo general más resistente que una delgada. El análisis de covarianza podría usarse para eliminar el efecto del grosor ( $x$ ) sobre la resistencia ( $y$ ) cuando se prueban las diferencias en la resistencia entre las máquinas.

### 14-3.1 Descripción del procedimiento

A continuación se describe el procedimiento básico para el análisis de covarianza, ilustrándolo para un experimento de un solo factor con una covariable. Suponiendo que existe una relación lineal entre la respuesta y la covariable, un modelo estadístico apropiado es

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad (14-15)$$

donde  $y_{ij}$  es la observación  $j$ -ésima de la variable de respuesta tomada bajo el tratamiento o nivel  $i$ -ésimo del único factor,  $x_{ij}$  es la medición hecha de la covariable o variable concomitante correspondiente a  $y_{ij}$  (es decir, la corrida  $ij$ -ésima),  $\bar{x}_{..}$  es la media de los valores  $x_{ij}$ ,  $\mu$  es la media global,  $\tau_i$  es el efecto del tratamiento  $i$ -ésimo,  $\beta$  es el coeficiente de regresión lineal que indica la dependencia de  $y_{ij}$  de  $x_{ij}$  y  $\varepsilon_{ij}$  es un componente del error aleatorio. Se supone que los errores  $\varepsilon_{ij}$  son NID( $0, \sigma^2$ ), que la pendiente  $\beta \neq 0$  y que la verdadera relación entre  $y_{ij}$  y  $x_{ij}$  es lineal, que los coeficientes de regresión de cada tratamiento son idénticos, que la suma de los efectos de los tratamientos es cero ( $\sum_{i=1}^a \tau_i = 0$ ) y que la variable concomitante  $x_{ij}$  no se afecta por los tratamientos.

Observe, por la ecuación 14-15, que el modelo del análisis de covarianza es una combinación de los modelos lineales empleados en el análisis de varianza y regresión. Es decir, se tienen efectos de los tratamientos  $\{\tau_i\}$ , como en un análisis de varianza de un solo factor, y un coeficiente de regresión  $\beta$ , como en una ecuación de regresión. La variable concomitante de la ecuación 14-15 se expresa como  $(x_{ij} - \bar{x}_{..})$

en lugar de  $x_{ij}$ , para que el parámetro  $\mu$  se preserve como la media global. El modelo pudo haberse escrito como

$$y_{ij} = \mu' + \tau_i + \beta x_{ij} + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad (14-16)$$

donde  $\mu'$  es una constante diferente de la media global, que para este modelo es  $\mu' + \beta\bar{x}_{..}$ . Es más común encontrar la ecuación 14-15 en la literatura sobre el tema.

Para describir el análisis, se introduce la siguiente notación:

$$S_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{an} \quad (14-17)$$

$$S_{xx} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n x_{ij}^2 - \frac{x_{..}^2}{an} \quad (14-18)$$

$$S_{xy} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) = \sum_{i=1}^a \sum_{j=1}^n x_{ij} y_{ij} - \frac{(x_{..})(y_{..})}{an} \quad (14-19)$$

$$T_{yy} = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 = \frac{1}{n} \sum_{i=1}^a y_{i.}^2 - \frac{y_{..}^2}{an} \quad (14-20)$$

$$T_{xx} = n \sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..})^2 = \frac{1}{n} \sum_{i=1}^a x_{i.}^2 - \frac{x_{..}^2}{an} \quad (14-21)$$

$$T_{xy} = n \sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..}) = \frac{1}{n} \sum_{i=1}^a (x_{i.})(y_{i.}) - \frac{(x_{..})(y_{..})}{an} \quad (14-22)$$

$$E_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 = S_{yy} - T_{yy} \quad (14-23)$$

$$E_{xx} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2 = S_{xx} - T_{xx} \quad (14-24)$$

$$E_{xy} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}) = S_{xy} - T_{xy} \quad (14-25)$$

Observe que, en general,  $S = T + E$ , donde los símbolos  $S$ ,  $T$  y  $E$  se usan para denotar las sumas de cuadrados y los productos cruzados del total, los tratamientos y el error, respectivamente. Las sumas de cuadrados de  $x$  y  $y$  deben ser no negativas; sin embargo, las sumas de los productos cruzados ( $xy$ ) pueden ser negativas.

A continuación se indica la forma en que el análisis de covarianza ajusta la variable de respuesta para el efecto de la covariable. Considere el modelo completo (ecuación 14-15). Los estimadores de mínimos cuadrados de  $\mu$ ,  $\tau_i$  y  $\beta$  son  $\hat{\mu} = \bar{y}_{..}$ ,  $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})$ , y

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} \quad (14-26)$$

La suma de cuadrados del error en este modelo es

$$SS_E = E_{yy} - (E_{xy})^2 / E_{xx} \quad (14-27)$$

con  $a(n - 1) - 1$  grados de libertad. La varianza del error experimental se estima con

$$MS_E = \frac{SS_E}{a(n-1)-1}$$

Suponga ahora que no hay ningún efecto de los tratamientos. El modelo (ecuación 14-15) sería entonces

$$y_{ij} = \mu + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad (14-28)$$

y puede demostrarse que los estimadores de mínimos cuadrados de  $\mu$  y  $\beta$  son  $\hat{\mu} = \bar{y}_{..}$  y  $\hat{\beta} = S_{xy}/S_{xx}$ . La suma de cuadrados del error en este modelo reducido es

$$SS'_E = S_{yy} - (S_{xy})^2 / S_{xx} \quad (14-29)$$

con  $an - 2$  grados de libertad. En la ecuación 14-29, la cantidad  $(S_{xy})^2/S_{xx}$  es la reducción de la suma de cuadrados de  $y$  obtenida a través de la regresión lineal de  $y$  sobre  $x$ . Además, observe que  $SS'_E$  es menor que  $SS_E$  [ya que el modelo (ecuación 14-15) contiene los parámetros adicionales  $\{\tau_i\}$ ] y que la cantidad  $SS'_E - SS_E$  es una reducción en la suma de cuadrados debida a las  $\{\tau_i\}$ . Por lo tanto, la diferencia entre  $SS'_E$  y  $SS_E$ , es decir,  $SS'_E - SS_E$ , proporciona una suma de cuadrados con  $a - 1$  grados de libertad para probar la hipótesis de que no hay ningún efecto de los tratamientos. Por consiguiente, para probar  $H_0: \tau_i = 0$ , se calcula

$$F_0 = \frac{(SS'_E - SS_E) / (a - 1)}{SS_E / [a(n - 1) - 1]} \quad (14-30)$$

que, si la hipótesis nula es verdadera, se distribuye como  $F_{a-1, a(n-1)-1}$ . Por lo tanto,  $H_0: \tau_i = 0$  se rechaza si  $F_0 > F_{\alpha, a-1, a(n-1)-1}$ . También podría usarse el enfoque del valor  $P$ .

Es instructivo examinar la presentación de la tabla 14-9. En ella el análisis de covarianza se ha presentado como un análisis de varianza "ajustado". En la columna de la fuente de variación, la variabilidad total se mide por  $S_{yy}$ , con  $an - 1$  grados de libertad. La fuente de variación "regresión" tiene la suma de cuadrados  $(S_{xy})^2/S_{xx}$  con un grado de libertad. Si no hubiera ninguna variable concomitante, se tendría  $S_{xy} = S_{xx} = E_{xx} = 0$ . Entonces la suma de cuadrados del error sería simplemente  $E_{yy}$  y la suma de cuadrados de los tratamientos sería  $S_{yy} - E_{yy} = T_{yy}$ . Sin embargo, debido a la presencia de la variable concomitante,  $S_{yy}$  y  $E_{yy}$  deben "ajustarse" para la regresión de  $y$  sobre  $x$ , como se muestra en la tabla 14-9. La suma de cuadrados del error ajustada tiene  $a(n - 1) - 1$  grados de libertad en lugar de  $a(n - 1)$  grados de libertad debido a que se ajusta un parámetro adicional (la pendiente  $\beta$ ) a los datos.

Los cálculos suelen presentarse en una tabla del análisis de covarianza como la tabla 14-10. Se emplea esta presentación porque resume de manera conveniente todas las sumas de cuadrados y los productos cruzados requeridos, así como las sumas de cuadrados para probar las hipótesis acerca de los efectos de los tratamientos. Además de utilizarla para probar la hipótesis de que no hay diferencias en los efectos de los tratamientos, con frecuencia esta tabla se encuentra útil en la interpretación de los datos para presentar las medias de los tratamientos ajustadas. Estas medias ajustadas se calculan de acuerdo con

$$\bar{y}_{i, \text{ajustada}} = \bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \quad i = 1, 2, \dots, a \quad (14-31)$$

donde  $\hat{\beta} = E_{xy} / E_{xx}$ . Esta media de los tratamientos ajustada es el estimador de mínimos cuadrados de  $\mu + \tau_i$ ,  $i = 1, 2, \dots, a$ , en el modelo (ecuación 14-15). El error estándar de cualquier media ajustada de los tratamientos es

$$S_{\bar{y}_{i, \text{ajustada}}} = \left[ MS_E \left( \frac{1}{n} + \frac{(\bar{x}_{i.} - \bar{x}_{..})^2}{E_{xx}} \right) \right]^{1/2} \quad (14-32)$$

Tabla 14-9 El análisis de covarianza como un análisis de varianza "ajustado"

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	$F_0$
Regresión	$(S_y)^2 / S_x$	1		
Tratamientos	$SS'_E - SS_E = S_y - (S_y)^2 / S_x - [E_y - (E_y)^2 / E_x]$	$a - 1$	$\frac{SS'_E - SS_E}{a - 1}$	$\frac{(SS'_E - SS_E) / (a - 1)}{MS_E}$
Error	$SS_E = E_y - (E_y)^2 / E_x$	$a(n - 1) - 1$	$MS_E = \frac{SS_E}{a(n - 1) - 1}$	
Total	$S_y$	$an - 1$		

Tabla 14-10 Análisis de covarianza de un experimento de un solo factor con una covariable

Fuente de variación	Grados de libertad	Sumas de cuadrados y productos			Ajustados para la regresión	
		$x$	$xy$	$y$	$y$	Cuadrado medio
Tratamientos	$a - 1$	$T_x$	$T_{xy}$	$T_y$		
Error	$a(n - 1)$	$E_x$	$E_{xy}$	$E_y$	$SS_E = E_y - (E_y)^2 / E_x$	$MS_E = \frac{SS_E}{a(n - 1) - 1}$
Total	$an - 1$	$S_x$	$S_{xy}$	$S_y$	$SS'_E = S_y - (S_y)^2 / S_x$	$an - 2$
Tratamientos ajustados					$SS'_E - SS_E$	$\frac{SS'_E - SS_E}{a - 1}$

Por último, cabe recordar que se ha supuesto que el coeficiente de regresión  $\beta$  del modelo (ecuación 14-15) es diferente de cero. La hipótesis  $H_0: \beta = 0$  puede probarse utilizando el estadístico de prueba

$$F_0 = \frac{(E_{xy})^2 / E_{xx}}{MS_E} \quad (14-33)$$

que bajo la hipótesis nula se distribuye como  $F_{1, a(n-1)-1}$ . Por lo tanto,  $H_0: \beta = 0$  se rechaza si  $F_0 > F_{\alpha, 1, a(n-1)-1}$ .

#### EJEMPLO 14-4 .....

Considere el experimento descrito al principio de la sección 14-3. Tres máquinas producen una fibra de monofilamento en una fábrica textil. El ingeniero del proceso tiene interés en determinar si existe alguna diferencia en la resistencia a la ruptura de la fibra producida por las tres máquinas. Sin embargo, la resistencia de una fibra se relaciona con su diámetro, con las fibras más gruesas, siendo éstas, en general, más resistentes que las más delgadas. Se selecciona una muestra aleatoria de cinco ejemplares de prueba de fibra de cada máquina. En la tabla 14-8 se muestra la resistencia de la fibra ( $y$ ) y el diámetro correspondiente ( $x$ ) de cada ejemplar.

El diagrama de dispersión de la resistencia a la ruptura contra el diámetro de la fibra (figura 14-3) indica una clara tendencia a una relación lineal entre la resistencia a la ruptura y el diámetro, y parece apropiado eliminar el efecto del diámetro sobre la resistencia mediante un análisis de covarianza. Suponiendo que la relación lineal entre la resistencia a la ruptura y el diámetro es apropiada, el modelo es

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, 3 \\ j = 1, 2, \dots, 5 \end{cases}$$

Utilizando las ecuaciones 14-17 a 14-25, pueden calcularse

$$S_{yy} = \sum_{i=1}^3 \sum_{j=1}^5 y_{ij}^2 - \frac{y_{..}^2}{an} = (36)^2 + (41)^2 + \dots + (32)^2 - \frac{(603)^2}{(3)(5)} = 346.40$$

$$S_{xx} = \sum_{i=1}^3 \sum_{j=1}^5 x_{ij}^2 - \frac{x_{..}^2}{an} = (20)^2 + (25)^2 + \dots + (15)^2 - \frac{(362)^2}{(3)(5)} = 261.73$$

$$S_{xy} = \sum_{i=1}^3 \sum_{j=1}^5 x_{ij} y_{ij} - \frac{(x_{..})(y_{..})}{an} = (20)(36) + (25)(41) + \dots + (15)(32) - \frac{(362)(603)}{(3)(5)} = 282.60$$

$$T_{yy} = \frac{1}{n} \sum_{i=1}^3 y_i^2 - \frac{y_{..}^2}{an} = \frac{1}{5} [(207)^2 + (216)^2 + (180)^2] - \frac{(603)^2}{(3)(5)} = 140.40$$

$$T_{xx} = \frac{1}{n} \sum_{i=1}^3 x_i^2 - \frac{x_{..}^2}{an} = \frac{1}{5} [(126)^2 + (130)^2 + (106)^2] - \frac{(362)^2}{(3)(5)} = 66.13$$

$$T_{xy} = \frac{1}{n} \sum_{i=1}^3 x_i y_i - \frac{(x_{..})(y_{..})}{an} = \frac{1}{5} [(126)(207) + (130)(216) + (106)(184)] - \frac{(362)(603)}{(3)(5)} = 96.00$$

$$\begin{aligned}
 E_{yy} &= S_{yy} - T_{yy} = 346.40 - 140.40 = 206.00 \\
 E_{xx} &= S_{xx} - T_{xx} = 261.73 - 66.13 = 195.60 \\
 E_{xy} &= S_{xy} - T_{xy} = 282.60 - 96.00 = 186.60
 \end{aligned}$$

Por la ecuación 14-29 se encuentra

$$\begin{aligned}
 SS'_E &= S_{yy} - (S_{xy})^2 / S_{xx} \\
 &= 346.40 - (186.60)^2 / 261.73 \\
 &= 41.27
 \end{aligned}$$

con  $an - 2 = (3)(5) - 2 = 13$  grados de libertad; y por la ecuación 14-27 se encuentra

$$\begin{aligned}
 SS_E &= E_{yy} - (E_{xy})^2 / E_{xx} \\
 &= 206.00 - (186.60)^2 / 195.60 \\
 &= 27.99
 \end{aligned}$$

con  $a(n - 1) - 1 = 3(5 - 1) - 1 = 11$  grados de libertad.

La suma de cuadrados para probar  $H_0: \tau_1 = \tau_2 = \tau_3 = 0$  es

$$\begin{aligned}
 SS'_E - SS_E &= 41.27 - 27.99 \\
 &= 13.28
 \end{aligned}$$

con  $a - 1 = 3 - 1 = 2$  grados de libertad. Estos cálculos se resumen en la tabla 14-11.

Para probar la hipótesis de que las máquinas difieren en la resistencia a la ruptura de la fibra producida, es decir,  $H_0: \tau_i = 0$ , por la ecuación 14-30 el estadístico de la prueba se calcula como

$$\begin{aligned}
 F_0 &= \frac{(SS'_E - SS_E) / (a - 1)}{SS_E / [a(n - 1) - 1]} \\
 &= \frac{13.28 / 2}{27.99 / 11} = \frac{6.64}{2.54} = 2.61
 \end{aligned}$$

Al comparar este valor con  $F_{0.10, 2, 11} = 2.86$ , se encuentra que no puede rechazarse la hipótesis nula. El valor  $P$  de este estadístico de la prueba es  $P = 0.1181$ . Por lo tanto, no hay evidencia sólida de que las fibras producidas por las tres máquinas difieran en la resistencia a la ruptura.

La estimación del coeficiente de regresión se calcula con la ecuación 14-26 como

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} = \frac{186.60}{195.60} = 0.9540$$

La hipótesis  $H_0: \beta = 0$  puede probarse usando la ecuación 14-33. El estadístico de prueba es

$$F_0 = \frac{(E_{xy})^2 / E_{xx}}{MS_E} = \frac{(186.60)^2 / 195.60}{2.54} = 70.08$$

y puesto que  $F_{0.01, 1, 11} = 9.65$ , se rechaza la hipótesis de que  $\beta = 0$ . Por lo tanto, existe una relación lineal entre la resistencia a la ruptura y el diámetro, y el ajuste proporcionado por el análisis de covarianza fue necesario.

**Tabla 14-11** Análisis de covarianza de los datos de la resistencia a la ruptura

Fuente de variación	Grados de libertad	Sumas de cuadrados y productos			Ajustados para la región			
		$x$	$xy$	$y$	$y$	Grados de libertad	Cuadrado medio	$F_0$ Valor $P$
Máquinas	2	66.13	96.00	140.40				
Error	12	195.60	186.60	206.00	27.99	11	2.54	
Total	14	261.73	282.60	346.40	41.27	13		
Máquinas ajustadas					13.28	2	6.64	2.61 0.1181



Las medias de los tratamientos ajustadas pueden calcularse con la ecuación 14-31. Estas medias ajustadas son

$$\begin{aligned}\bar{y}_1 \text{ ajustada} &= \bar{y}_1 - \hat{\beta}(\bar{x}_1 - \bar{x}_..) \\ &= 41.40 - (0.9540)(25.20 - 24.13) = 40.38\end{aligned}$$

$$\begin{aligned}\bar{y}_2 \text{ ajustada} &= \bar{y}_2 - \hat{\beta}(\bar{x}_2 - \bar{x}_..) \\ &= 43.20 - (0.9540)(26.00 - 24.13) = 41.42\end{aligned}$$

y

$$\begin{aligned}\bar{y}_3 \text{ ajustada} &= \bar{y}_3 - \hat{\beta}(\bar{x}_3 - \bar{x}_..) \\ &= 36.00 - (0.9540)(21.20 - 24.13) = 38.80\end{aligned}$$

Al comparar las medias ajustadas con las medias no ajustadas de los tratamientos (las  $\bar{y}_i$ ), se observa que las medias ajustadas se encuentran mucho más próximas entre sí, una indicación más de que el análisis de covarianza fue necesario.

Un supuesto básico en el análisis de covarianza es que los tratamientos no influyen en la covariable  $x$ , ya que la técnica elimina el efecto de las variaciones en las  $\bar{x}_i$ . Sin embargo, si la variabilidad en las  $\bar{x}_i$  se debe en parte a los tratamientos, entonces el análisis de covarianza elimina parte del efecto de los tratamientos. Por lo tanto, deberá tenerse una seguridad razonable de que los tratamientos no afectan los valores  $x_{ij}$ . En algunos experimentos esto puede ser obvio a partir de la naturaleza de la covariable, mientras que en otros puede ser más dudoso. En el ejemplo tratado aquí puede haber una diferencia en el diámetro de la fibra ( $x_{ij}$ ) entre las tres máquinas. En tales casos, Cochran y Cox [26] sugieren la posible utilidad de un análisis de varianza de los valores  $x_{ij}$  para determinar la validez de este supuesto. Para el problema tratado aquí, con este procedimiento se obtiene

$$F_0 = \frac{66.13/2}{195.60/12} = \frac{33.07}{16.30} = 2.03$$

que es menor que  $F_{0.10,2,12} = 2.81$ , por lo que no hay razón para creer que las máquinas producen fibras con diámetros diferentes.

.....

La verificación del diagnóstico del modelo de covarianza se basa en el análisis residual. Para el modelo de covarianza, los residuales son

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

donde los valores ajustados son

$$\begin{aligned}\hat{y}_{ij} &= \hat{\mu} + \hat{\tau}_i + \hat{\beta}(x_{ij} - \bar{x}_..) = \bar{y}_i + [\bar{y}_i - \bar{y}_.. - \hat{\beta}(\bar{x}_i - \bar{x}_..)] \\ &\quad + \hat{\beta}(x_{ij} - \bar{x}_..) = \bar{y}_i + \hat{\beta}(x_{ij} - \bar{x}_i)\end{aligned}$$

Por lo tanto,

$$e_{ij} = y_{ij} - \bar{y}_i - \hat{\beta}(x_{ij} - \bar{x}_i) \quad (14-34)$$

Para ilustrar el uso de la ecuación 14-34, el residual de la primera observación de la primera máquina del ejemplo 14-4 es

$$\begin{aligned}e_{11} &= y_{11} - \bar{y}_1 - \hat{\beta}(x_{11} - \bar{x}_1) = 36 - 41.4 - (0.9540)(20 - 25.2) \\ &= 36 - 36.4392 = -0.4392\end{aligned}$$

En la tabla siguiente se presenta una lista completa de las observaciones, los valores ajustados y los residuales:

Valor observado $y_{ij}$	Valor ajustado $\hat{y}_{ij}$	Residual $e_{ij} = y_{ij} - \hat{y}_{ij}$
36	36.4392	-0.4392
41	41.2092	-0.2092
39	40.2552	-1.2552
42	41.2092	0.7908
49	47.8871	1.1129
40	39.3840	0.6160
48	45.1079	2.8921
39	39.3840	-0.3840
45	47.0159	-2.0159
44	45.1079	-1.1079
35	35.8092	-0.8092
37	37.7171	-0.7171
42	40.5791	1.4209
34	35.8092	-1.8092
32	30.0852	1.9148

Los residuales se grafican contra los valores ajustados  $\hat{y}_{ij}$  en la figura 14-4, contra la covariable  $x_{ij}$  en la figura 14-5 y contra las máquinas en la figura 14-6. En la figura 14-7 se muestra la gráfica de probabilidad normal de los residuales. Estas gráficas no revelan ninguna desviación importante de los supuestos, por lo que se concluye que el modelo de covarianza (ecuación 14-15) es apropiado para los datos de la resistencia a la ruptura.

Es interesante observar lo que habría ocurrido en este experimento si no se hubiera realizado el análisis de covarianza, es decir, si los datos de la resistencia a la ruptura ( $y$ ) se hubieran analizado como un experimento de un solo factor en el que se ignorara la covariable  $x$ . En la tabla 14-12 se muestra el análisis de varianza de los datos de la resistencia a la ruptura. Se concluiría, con base en este análisis, que las máquinas difieren significativamente en la resistencia de la fibra producida. Es exactamente *la conclusión*

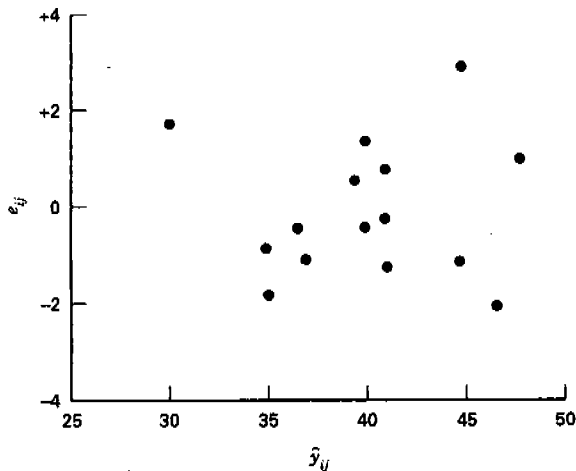


Figura 14-4 Gráfica de los residuales contra los valores ajustados del ejemplo 14-4.

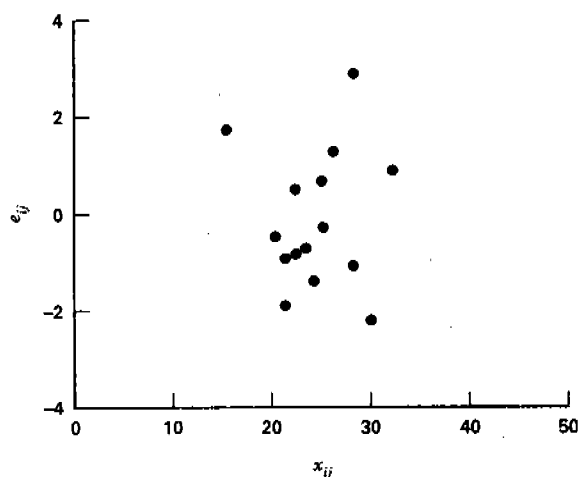


Figura 14-5 Gráfica de los residuales contra el diámetro  $x$  de la fibra en el ejemplo 14-4.

*opuesta* del análisis de covarianza. Si se sospechara que las máquinas difieren significativamente en su efecto sobre la resistencia de la fibra, entonces se intentaría igualar la resistencia producida por las tres máquinas. Sin embargo, en este problema las máquinas no difieren en la resistencia de la fibra producida después de que se elimina el efecto lineal del diámetro. Sería conveniente reducir la variabilidad del diámetro de la fibra dentro de las máquinas, ya que con esto probablemente se reduciría la variabilidad de la resistencia de la fibra.

### 14-3.2 Solución por computadora

Se cuenta con varios paquetes de software que pueden realizar el análisis de covarianza. En la tabla 14-13 se muestra la salida del procedimiento General Linear Models (modelos lineales generales) de Minitab

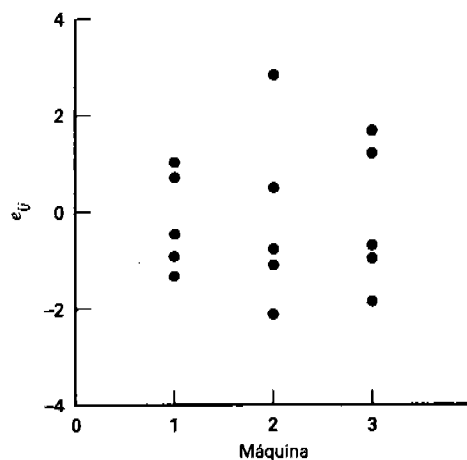


Figura 14-6 Gráfica de los residuales contra las máquinas.

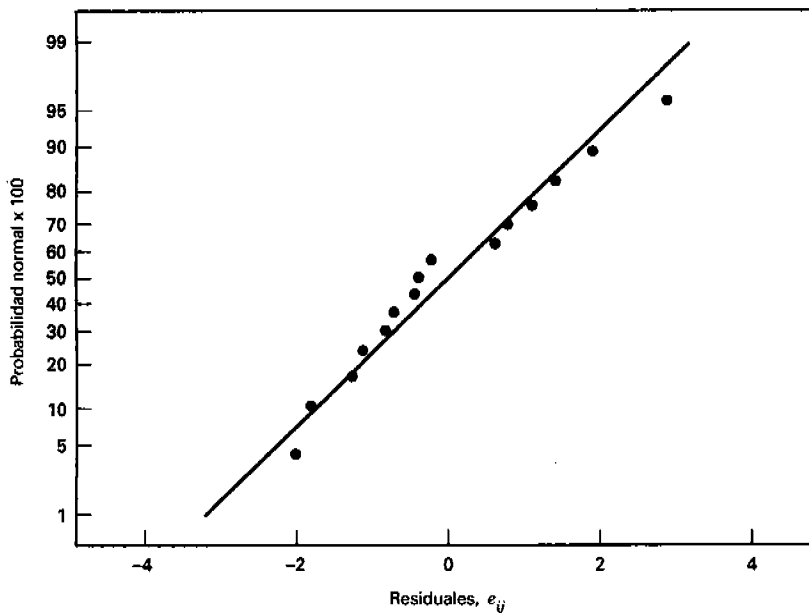


Figura 14-7 Gráfica de probabilidad normal de los residuales del ejemplo 14-4.

para los datos del ejemplo 14-4. Esta salida es muy similar a las que se presentaron anteriormente. En la sección de la salida bajo el encabezado “Análisis de varianza” (“Analysis of variance”), “SS Seq” corresponde a la partición “secuencial” de la suma de cuadrados del modelo global, es decir,

$$\begin{aligned} SS(\text{Modelo}) &= SS(\text{Diámetro}) + SS(\text{Máquina} | \text{Diámetro}) \\ &= 305.13 + 13.28 \\ &= 318.41 \end{aligned}$$

mientras que “SS ajustada” corresponde a la suma de cuadrados “extra” para cada factor, es decir,

$$SS(\text{Máquina} | \text{Diámetro}) = 13.28$$

y

$$SS(\text{Diámetro} | \text{Máquina}) = 178.01$$

Observe que  $SS(\text{Máquina} | \text{Diámetro})$  es la suma de cuadrados que deberá usarse para probar que no hay ningún efecto de la máquina, y que  $SS(\text{Diámetro} | \text{Máquina})$  es la suma de cuadrados corregida que deberá usarse para probar la hipótesis de que  $\beta = 0$ . Los estadísticos de la prueba de la tabla 14-13 difieren ligeramente de los que se calcularon manualmente debido al redondeo.

Tabla 14-12 Análisis incorrecto de los datos de la resistencia a la ruptura como un experimento de un solo factor

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	$F_0$	Valor $P$
Máquinas	140.40	2	70.20	4.09	0.0442
Error	206.00	12	17.17		
Total	346.40	14			

Tabla 14-13 Salida de Minitab (análisis de covarianza) del ejemplo 14-4

Modelo lineal general

Factor	Type	Levels	Values
Machine	fixed	3	1 2 3

## Analysis of Variance for Strength, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Diameter	1	305.13	178.01	178.01	69.97	0.000
Machine	2	13.28	13.28	6.64	2.61	0.118
Error	11	27.99	27.99	2.54		
Total	14	346.40				

Term	Coef	StDev	T	P
Constant	17.177	2.783	6.17	0.000
Diameter	0.9540	0.1140	8.36	0.000
Machine				
1	0.1824	0.5950	0.31	0.765
2	1.2192	0.6201	1.97	0.075

## Means for Covariates

Covariate	Mean	StDev
Diameter	24.13	4.324

## Least Squares Means for Strength

Machine	Mean	StDev
1	40.38	0.7236
2	41.42	0.7444
3	38.80	0.7879

El programa calcula también las medias de los tratamientos ajustadas con la ecuación 14-31 (Minitab hace referencia a éstas como las medias de mínimos cuadrados en la salida muestral) y los errores estándar. El programa comparará asimismo todos los pares de medias de tratamientos utilizando los procedimientos de comparación múltiple por pares estudiados en el capítulo 3.

## 14-3.3 Desarrollo mediante la prueba general de significación de la regresión

Es posible desarrollar formalmente el procedimiento para probar  $H_0: \tau_i = 0$  en el modelo de covarianza

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad (14-35)$$

utilizando la prueba general de significación de la regresión. Considere la estimación de los parámetros del modelo (ecuación 14-15) por mínimos cuadrados. La función de mínimos cuadrados es

$$L = \sum_{i=1}^a \sum_{j=1}^n [y_{ij} - \mu - \tau_i - \beta(x_{ij} - \bar{x}_{..})]^2 \quad (14-36)$$

y a partir de  $\partial L / \partial \mu = \partial L / \partial \tau_i = \partial L / \partial \beta = 0$ , se obtienen las ecuaciones normales

$$\mu: an\hat{\mu} + n \sum_{i=1}^a \hat{\tau}_i = y_{..} \quad (14-37a)$$

$$\tau_i: n\hat{\mu} + n\hat{\tau}_i + \hat{\beta} \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) = y_{i.} \quad i = 1, 2, \dots, a \quad (14-37b)$$

$$\beta: \sum_{i=1}^a \hat{\tau}_i \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) + \hat{\beta} S_{xx} = S_{xy} \quad (14-37c)$$

Al sumar las  $a$  ecuaciones de la ecuación 14-37b, se obtiene la ecuación 14-37a porque  $\sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) = 0$ , por lo que existe una dependencia lineal en las ecuaciones normales. Por lo tanto, es necesario aumentar las ecuaciones 14-37 con una ecuación linealmente independiente para obtener una solución. Una condición lógica es  $\sum_{i=1}^a \hat{\tau}_i = 0$ .

Al utilizar esta condición, de la ecuación 14-37a se obtiene

$$\hat{\mu} = \bar{y}_{..} \quad (14-38a)$$

y de la ecuación 14-37b

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \quad (14-38b)$$

La ecuación 14-37c puede reescribirse como

$$\sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..}) \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) - \hat{\beta} \sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..}) \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) + \hat{\beta} S_{xx} = S_{xy}$$

después de sustituir para  $\hat{\tau}_i$ . Pero se observa que

$$\sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..}) \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) = T_{xy}$$

y

$$\sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..}) \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) = T_{xx}$$

Por lo tanto, la solución de la ecuación 14-37c es

$$\hat{\beta} = \frac{S_{xy} - T_{xy}}{S_{xx} - T_{xx}} = \frac{E_{xy}}{E_{xx}}$$

que fue el resultado dado anteriormente en la sección 14-3.1 en la ecuación 14-26.

La reducción en la suma de cuadrados total debida al ajuste del modelo (ecuación 14-15) puede expresarse como

$$\begin{aligned} R(\mu, \tau, \beta) &= \hat{\mu} y_{..} + \sum_{i=1}^a \hat{\tau}_i y_{i.} + \hat{\beta} S_{xy} \\ &= (\bar{y}_{..}) y_{..} + \sum_{i=1}^a [\bar{y}_{i.} - \bar{y}_{..} - (E_{xy} / E_{xx})(\bar{x}_{i.} - \bar{x}_{..})] y_{i.} + (E_{xy} / E_{xx}) S_{xy} \\ &= y_{..}^2 / an + \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..}) y_{i.} - (E_{xy} / E_{xx}) \sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..}) y_{i.} + (E_{xy} / E_{xx}) S_{xy} \\ &= y_{..}^2 / an + T_{yy} - (E_{xy} / E_{xx})(T_{xy} - S_{xy}) \\ &= y_{..}^2 / an + T_{yy} + (E_{xy})^2 / E_{xx} \end{aligned}$$

Esta suma de cuadrados tiene  $a + 1$  grados de libertad porque el rango de las ecuaciones normales es  $a + 1$ . La suma de cuadrados del error de este modelo es

$$\begin{aligned}
 SS_E &= \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - R(\mu, \tau, \beta) \\
 &= \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - y_{..}^2 / an - T_{yy} - (E_{xy})^2 / E_{xx} \\
 &= S_{yy} - T_{yy} - (E_{xy})^2 / E_{xx} \\
 &= E_{yy} - (E_{xy})^2 / E_{xx}
 \end{aligned} \tag{14-39}$$

con  $an - (a + 1) = a(n - 1) - 1$  grados de libertad. Esta cantidad se obtuvo anteriormente como la ecuación 14-27.

Considere ahora el modelo restringido a la hipótesis nula, es decir, a  $H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$ . Este modelo reducido es

$$y_{ij} = \mu + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \tag{14-40}$$

Se trata de un modelo de regresión lineal simple, y las ecuaciones normales de mínimos cuadrados para este modelo son

$$an\hat{\mu} = y_{..} \tag{14-41a}$$

$$\hat{\beta}S_{xx} = S_{xy} \tag{14-41b}$$

Las soluciones de estas ecuaciones son  $\hat{\mu} = \bar{y}_{..}$  y  $\hat{\beta} = S_{xy}/S_{xx}$ ; y la reducción en la suma de cuadrados total debida al ajuste del modelo reducido es

$$\begin{aligned}
 R(\mu, \beta) &= \hat{\mu}y_{..} + \hat{\beta}S_{xy} \\
 &= (\bar{y}_{..})y_{..} + (S_{xy}/S_{xx})S_{xy} \\
 &= y_{..}^2 / an + (S_{xy})^2 / S_{xx}
 \end{aligned} \tag{14-42}$$

Esta suma de cuadrados tiene dos grados de libertad.

La suma de cuadrados apropiada para probar  $H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$  puede encontrarse como

$$\begin{aligned}
 R(\tau | \mu, \beta) &= R(\mu, \tau, \beta) - R(\mu, \beta) \\
 &= y_{..}^2 / an + T_{yy} + (E_{xy})^2 / E_{xx} - y_{..}^2 / an - (S_{xy})^2 / S_{xx} \\
 &= S_{yy} - (S_{xy})^2 / S_{xx} - [E_{yy} - (E_{xy})^2 / E_{xx}]
 \end{aligned} \tag{14-43}$$

utilizando  $T_{yy} = S_{yy} - E_{yy}$ . Observe que  $R(\tau | \mu, \beta)$  tiene  $a + 1 - 2 = a - 1$  grados de libertad y que es idéntica a la suma de cuadrados dada por  $SS'_E - SS_E$  en la sección 14-3.1. Por lo tanto, el estadístico de prueba para  $H_0: \tau_i = 0$  es

$$F_0 = \frac{R(\tau | \mu, \beta) / (a - 1)}{SS_E / [a(n - 1) - 1]} = \frac{(SS'_E - SS_E) / (a - 1)}{SS_E / [a(n - 1) - 1]} \tag{14-44}$$

expresión que se dio anteriormente como la ecuación 14-30. Por lo tanto, utilizando la prueba general de significación de la regresión, se ha justificado el desarrollo heurístico del análisis de covarianza de la sección 14-3.1.

### 14-3.4 Experimentos factoriales con covariables

El análisis de covarianza puede aplicarse a estructuras de tratamientos más complejas, como los diseños factoriales. Siempre que existan datos suficientes para cada combinación de tratamientos, prácticamente cualquier estructura de tratamientos compleja puede analizarse mediante el enfoque del análisis de covarianza. A continuación se indica cómo podría usarse el análisis de covarianza en la familia más común de diseños factoriales utilizados en la experimentación industrial, los factoriales  $2^k$ .

Al establecer el supuesto de que la covariable afecta a la variable de respuesta de manera idéntica para todas las combinaciones de tratamientos, podría construirse una tabla del análisis de covarianza similar al procedimiento dado en la sección 14-3.1. La única diferencia sería la suma de cuadrados de los tratamientos. Para un factorial  $2^2$  con  $n$  réplicas, la suma de cuadrados de los tratamientos ( $T_{yy}$ ) sería  $(1/n) \sum_{i=1}^2 \sum_{j=1}^2 y_{ij}^2 - y_{...}^2 / (2)(2)n$ . Esta cantidad es la suma de las sumas de cuadrados de los factores  $A$ ,  $B$  y la interacción  $AB$ . Entonces podría hacerse la partición de la suma de cuadrados ajustada de los tratamientos en componentes de los efectos individuales, es decir, la suma de cuadrados de los efectos principales ajustados  $SS_A$  y  $SS_B$ , y una suma de cuadrados de la interacción,  $SS_{AB}$ .

El número de réplicas es un aspecto clave cuando se amplía la estructura de los tratamientos del diseño. Considere un arreglo factorial  $2^3$ . Se necesita un mínimo de dos réplicas para evaluar todas las combinaciones de tratamientos con una covariable separada para cada combinación de tratamientos (una covariable por interacción de tratamientos). Esto es equivalente a ajustar un modelo de regresión simple a cada combinación de tratamientos o celda del diseño. Con dos observaciones por celda, un grado de libertad se usa para estimar la ordenada al origen (el efecto del tratamiento), y el otro se usa para estimar la pendiente (el efecto de la covariable). Con este modelo saturado, no se cuenta con ningún grado de libertad para estimar el error. Por lo tanto, se necesitan al menos tres réplicas para un análisis de covarianza completo, suponiendo el caso más general. Este problema se agudiza cuando se incrementa el número de celdas distintas del diseño (combinaciones de tratamientos) y las covariables.

Si el número de réplicas está limitado, pueden hacerse varios supuestos para permitir un análisis útil. El supuesto más simple (y típicamente el peor) que puede hacerse es que la covariable no tiene ningún efecto. Si la covariable, incorrectamente, deja de tomarse en consideración, el análisis completo y las conclusiones subsecuentes podrían tener graves errores. Otra elección es suponer que no hay ningún tratamiento por interacción de la covariable. Aun cuando este supuesto sea incorrecto, el efecto promedio de la covariable en todos los tratamientos seguirá incrementando la precisión de la estimación y la prueba de los efectos de los tratamientos. Una desventaja de este supuesto es que si varios niveles de los tratamientos interactúan con la covariable, los diferentes términos pueden cancelarse entre sí y el término de la covariable, si se estima solo sin ninguna interacción, puede resultar no significativo. Una tercera elección sería suponer que algunos de los factores (como algunas interacciones de dos factores y de órdenes superiores) no son significativos. Esto permite usar parte de los grados de libertad para estimar el error. Sin embargo, este curso de acción deberá emprenderse con cuidado, y los modelos subsecuentes deberán evaluarse a profundidad, ya que la estimación del error será relativamente imprecisa a menos que se le asignen suficientes grados de libertad. Con dos réplicas, cada uno de estos supuestos liberará algunos grados de libertad para estimar el error y permitirá realizar pruebas de hipótesis útiles. El supuesto que se establecerá deberá ser dictado por la situación experimental y por el riesgo que el experimentador esté dispuesto a correr. Cabe hacer notar que en la estrategia de construcción del modelo de los efectos, si se elimina el factor de uno de los tratamientos, entonces las dos "réplicas" resultantes de cada factorial  $2^3$  original no son en realidad réplicas. Estas "réplicas ocultas" liberan grados de libertad para la estimación de parámetros, pero no deberán usarse como réplicas para estimar el error puro porque la ejecución del diseño original quizá no se haya aleatorizado para ello.



Tabla 14-14 Datos de la respuesta y la covariable para un diseño  $2^3$  con 2 réplicas

A	B	C	x	y
-1	-1	-1	4.05	-30.73
1	-1	-1	0.36	9.07
-1	1	-1	5.03	39.72
1	1	-1	1.96	16.30
-1	-1	1	5.38	-26.39
1	-1	1	8.63	54.58
-1	1	1	4.10	44.54
1	1	1	11.44	66.20
-1	-1	-1	3.58	-26.46
1	-1	-1	1.06	10.94
-1	1	-1	15.53	103.01
1	1	-1	2.92	20.44
-1	-1	1	2.48	-8.94
1	-1	1	13.64	73.72
-1	1	1	-0.67	15.89
1	1	1	5.13	38.57

Para ilustrar algunas de estas ideas, considere el diseño factorial  $2^3$  con dos réplicas y una covariable que se muestra en la tabla 14-14. Si la variable de respuesta y se analiza sin tomar en cuenta la covariable, resulta el siguiente modelo:

$$\hat{y} = 25.03 + 11.20A + 18.05B + 7.24C - 18.91AB + 14.80AC$$

El modelo global es significativo en el nivel de  $\alpha = 0.01$  con  $R^2 = 0.786$  y  $MS_E = 470.82$ . El análisis residual no indica problemas con este modelo, excepto porque la observación con  $y = 103.01$  es inusual.

Si se elige el segundo supuesto, que las pendientes son comunes con ningún tratamiento por interacción de la covariable, pueden estimarse el modelo de los efectos completo y el efecto de la covariable. La salida de Minitab (de la rutina General Linear Models) se muestra en la tabla 14-15. Observe que  $MS_E$  se ha reducido considerablemente al tomar en consideración la covariable. El análisis final resultante después de eliminar de manera secuencial cada interacción no significativa y el efecto principal  $C$  se muestra en la tabla 14-16. Este modelo reducido proporciona un  $MS_E$  todavía menor que el modelo completo con la covariable de la tabla 14-15.

Por último, podría considerarse un tercer curso de acción, suponiendo que ciertos términos de interacción son insignificantes. Se considera el modelo completo que permite pendientes diferentes entre los tratamientos y la interacción tratamiento por covariable. Se supone que no son significativas las interacciones de tres factores (tanto  $ABC$  como  $ABCx$ ) y se usan los grados de libertad asociados con ellas para estimar el error en el modelo de los efectos más general que pueda ajustarse. Éste es con frecuencia un supuesto práctico. Las interacciones de tres factores son por lo general insignificantes en la mayoría de los ambientes experimentales. La versión actual de Minitab no puede modelar covariables que interactúan con los tratamientos, por lo que se usa PROC GLM de SAS. Las sumas de cuadrados tipo III son las sumas de cuadrados ajustadas que se necesitan. En la tabla 14-17 se presentan los resultados de SAS para este modelo.

Con un modelo casi saturado, la estimación del error será bastante imprecisa. Incluso cuando unos cuantos términos son individualmente significativos en el nivel  $\alpha = 0.05$ , el sentido general es que este modelo es mejor que los dos escenarios previos (basados en  $R^2$  y el cuadrado medio del error). Debido a que el aspecto de los efectos de los tratamientos del modelo es de mayor interés, se eliminan de manera secuencial términos de la porción de la covariable del modelo a fin de agregar grados de libertad para esti-

Tabla 14-15 Análisis de covarianza de Minitab para el experimento de la tabla 14-14, suponiendo una pendiente común

Modelo lineal general

Factor	Type	Levels	Values
A	fixed	2	-1 1
B	fixed	2	-1 1
C	fixed	2	-1 1

Analysis of Variance for y, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
x	1	12155.9	2521.6	2521.6	28.10	0.001
A	1	1320.7	1403.8	1403.8	15.64	0.005
B	1	3997.6	4066.2	4066.2	45.31	0.000
C	1	52.7	82.3	82.3	0.92	0.370
A*B	1	3788.3	3641.0	3641.0	40.58	0.000
A*C	1	10.2	1.1	1.1	0.01	0.913
B*C	1	5.2	8.4	8.4	0.09	0.769
A*B*C	1	33.2	33.2	33.2	0.37	0.562
Error	7	628.1	628.1	89.7		
Total	15	21992.0				

Term	Coef	StDev	T	P
Constant	-1.016	5.454	-0.19	0.858
x	4.9245	0.9290	5.30	0.001

mar el error. Si se elimina secuencialmente el término  $ACx$  seguido de  $BCx$ , el  $MS_E$  decrece a 0.7336 y varios términos no son significativos. En la tabla 14-18 se muestra el modelo final después de eliminar secuencialmente  $Cx$ ,  $AC$  y  $BC$ .

Este ejemplo destaca la necesidad de contar con grados de libertad para estimar el error experimental a fin de incrementar la precisión de las pruebas de hipótesis asociadas con los términos individuales del modelo. Este proceso deberá hacerse de manera secuencial para evitar la eliminación de términos significativos enmascarados por una estimación pobre del error.

Tabla 14-16 Análisis de covarianza de Minitab, modelo reducido para el experimento de la tabla 14-14

Modelo lineal general

Factor	Type	Levels	Values
A	fixed	2	-1 1
B	fixed	2	-1 1

Analysis of Variance for y, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
x	1	12155.9	8287.9	8287.9	119.43	0.000
A	1	1320.7	1404.7	1404.7	20.24	0.001
B	1	3997.6	4097.7	4097.7	59.05	0.000
A*B	1	3754.5	3754.5	3754.5	54.10	0.000
Error	11	763.3	763.3	69.4		
Total	15	21992.0				

Term	Coef	StDev	T	P
Constant	-1.878	3.225	-0.58	0.572
x	5.0876	0.4655	10.93	0.000

Tabla 14-17 Salida de PROC GLM (análisis de covarianza) de SAS para el experimento de la tabla 14-14

Dependent Variable: Y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	21989.20828	1691.47756	1206.45	0.0008
Error	2	2.80406	1.40203		
Corrected Total	15	21992.01234			
R-Square = 0.999872					
C.V. = 4.730820					
Root MSE = 1.184074					
Y Mean = 25.02895					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	4.6599694	4.6599694	3.32	0.2099
B	1	13.0525319	13.0525319	9.31	0.0927
C	1	35.0087994	35.0087994	24.97	0.0378
AB	1	17.1013635	17.1013635	12.20	0.0731
AC	1	0.0277472	0.0277472	0.02	0.9010
BC	1	0.4437474	0.4437474	0.32	0.6304
X	1	49.2741287	49.2741287	35.14	0.0273
AX	1	33.9024288	33.9024288	24.18	0.0390
BX	1	95.7747490	95.7747490	68.31	0.0143
CX	1	0.1283784	0.1283784	0.09	0.7908
ABX	1	336.9732676	336.9732676	240.35	0.0041
ACX	1	0.0020997	0.0020997	0.00	0.9726
BCX	1	0.0672386	0.0672386	0.05	0.8470

Tabla 14-18 Salida de PROC GLM de SAS para el experimento de la tabla 14-14, modelo reducido

Dependent Variable: Y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	21986.33674	2748.29209	3389.61	0.0001
Error	7	5.67560	0.81080		
Corrected Total	15	21992.01234			
R-Square					
	0.999742	C.V.	Root MSE	Y Mean	
		3.597611	0.900444	25.02895	
Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	19.1597158	19.1597158	23.63	0.0018
B	1	38.0317496	38.0317496	46.91	0.0002
C	1	232.2435668	232.2435668	286.44	0.0001
AB	1	31.7635098	31.7635098	39.18	0.0004
X	1	240.8726525	240.8726525	297.08	0.0001
AX	1	233.3934567	233.3934567	287.86	0.0001
BX	1	550.1530561	550.1530561	678.53	0.0001
ABX	1	542.3268940	542.3268940	668.88	0.0001
Parameter	Estimate	T for H0: Parameter=0	Pr >  T	Std Error of Estimate	
Intercept	10.2438830	18.74	0.0001	0.54659908	
A	2.7850330	4.86	0.0018	0.57291820	
B	3.6596279	6.85	0.0002	0.53434356	
C	5.4560862	16.92	0.0001	0.32237858	
AB	-3.3636850	-6.26	0.0004	0.53741264	
X	2.0471937	17.24	0.0001	0.11877417	
AX	2.0632049	16.97	0.0001	0.12160595	
BX	3.0340997	26.05	0.0001	0.11647826	
ABX	-3.0342229	-25.86	0.0001	0.11732045	

Al revisar los resultados obtenidos de los tres enfoques, se observa que cada método mejora de manera sucesiva el ajuste del modelo en este ejemplo. Si hay una razón fundada para creer que la covariable no interactúa con los factores, quizá sea mejor establecer ese supuesto desde el principio del análisis. Esta opción también puede ser dictada por el software. Aun cuando los paquetes de software para diseños experimentales quizá sólo tengan capacidad para modelar covariables que no interactúan con los tratamientos, el analista puede tener una oportunidad razonable de identificar los factores principales que influyen en el proceso, incluso si hay alguna covariable por interacción de tratamientos. Se observa asimismo que las pruebas usuales de la adecuación del modelo siguen siendo apropiadas y se recomiendan enérgicamente como parte del proceso de construcción del modelo del análisis de covarianza ANCOVA.

#### 14-4 MEDICIONES REPETIDAS

En el trabajo experimental de las ciencias sociales y el comportamiento, así como en algunos aspectos de la ingeniería y las ciencias físicas, las unidades experimentales son con frecuencia personas. Debido a las discrepancias en experiencia, capacitación o formación, en algunas situaciones experimentales las diferencias en las respuestas de distintas personas al mismo tratamiento pueden ser muy grandes. A menos que esté controlada, esta variabilidad entre las personas se convertirá en parte del error experimental, y en algunos casos inflará significativamente el cuadrado medio del error, haciendo más difícil detectar las diferencias reales entre los tratamientos.

Es posible controlar esta variabilidad entre las personas utilizando un diseño en el que cada uno de los  $a$  tratamientos se usa en cada persona (o "sujeto"). A éste se le llama **diseño de mediciones repetidas**. En esta sección se ofrece una breve introducción a los experimentos de mediciones repetidas con un solo factor.

Suponga que un experimento incluye  $a$  tratamientos y que cada tratamiento se va a usar exactamente una sola vez en cada uno de  $n$  sujetos. Los datos aparecerían como en la tabla 14-19. Vea que la observación  $y_{ij}$  representa la respuesta del sujeto  $j$  al tratamiento  $i$  y que sólo se usan  $n$  sujetos. El modelo que se utiliza para este diseño es

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad (14-45)$$

donde  $\tau_i$  es el efecto del tratamiento  $i$ -ésimo y  $\beta_j$  es un parámetro asociado con el sujeto  $j$ -ésimo. Se supone que los tratamientos son fijos (de donde  $\sum_{i=1}^a \tau_i = 0$ ) y que los sujetos empleados son una muestra aleatoria de alguna población más grande de individuos potenciales. Por lo tanto, colectivamente los sujetos representan un efecto aleatorio, por lo que se supone que la media de  $\beta_j$  es cero y que la varianza de  $\beta_j$  es  $\sigma_\beta^2$ . Puesto que el término  $\beta_j$  es común a todas las  $a$  mediciones del mismo sujeto, la covarianza entre  $y_{ij}$  y  $y_{i'j}$  no

Tabla 14-19 Datos de un diseño de mediciones repetidas con un solo factor

Tratamiento	Sujeto				Totales de los tratamientos
	1	2	...	$n$	
1	$y_{11}$	$y_{12}$	...	$y_{1n}$	$y_{1.}$
2	$y_{21}$	$y_{22}$	...	$y_{2n}$	$y_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$a$	$y_{a1}$	$y_{a2}$		$y_{an}$	$y_{a.}$
Totales de los sujetos	$y_{.1}$	$y_{.2}$	...	$y_{.n}$	$y_{..}$