

3

Experimentos con un solo factor: el análisis de varianza

En el capítulo 2 se analizaron los métodos para comparar dos condiciones o tratamientos. Por ejemplo, el experimento de la fuerza de la tensión de adhesión del cemento portland incluyó dos formulaciones diferentes del mortero. Otra forma de describir este experimento es como un experimento con un solo factor, con dos niveles del factor, donde el factor es la formulación del mortero y los dos niveles son los dos métodos diferentes para hacer la formulación. Muchos experimentos de este tipo involucran más de dos niveles del factor. En este capítulo se presentan los métodos para el diseño y el análisis de los experimentos con un solo factor con a niveles del mismo (o a tratamientos). Se supondrá que el experimento se ha aleatorizado completamente.

3-1 UN EJEMPLO

Un ingeniero de desarrollo de productos tiene interés en investigar la resistencia a la tensión de una fibra sintética nueva que se usará para hacer tela de camisas para caballero. El ingeniero sabe por experiencia previa que la resistencia a la tensión se afecta por el peso porcentual del algodón utilizado en la mezcla de materiales de la fibra. Además, sospecha que al aumentar el contenido de algodón se incrementará la resistencia, al menos en un principio. Sabe asimismo que el contenido de algodón deberá variar entre 10 y 40 por ciento para que el producto final tenga otras características de calidad que se desean (como la capacidad de ser sometido a un tratamiento de planchado permanente). El ingeniero decide probar ejemplares en cinco niveles del peso porcentual del algodón: 15, 20, 25, 30 y 35 por ciento. También decide probar cinco ejemplares en cada nivel del contenido de algodón.

Se trata de un ejemplo de un experimento con un solo factor con $a = 5$ niveles del factor y $n = 5$ réplicas. Las 25 corridas deberán realizarse de manera aleatoria. Para ilustrar cómo puede aleatorizarse el orden de las corridas, suponga que las corridas se numeran de la siguiente manera:

Peso porcentual del algodón	Número de corrida experimental				
15	1	2	3	4	5
20	6	7	8	9	10
25	11	12	13	14	15
30	16	17	18	19	20
35	21	22	23	24	25

Ahora se selecciona un número aleatorio entre 1 y 25. Suponga que este número es 8. Entonces la observación número 8 (20% de algodón) se corre primero. Este proceso se repetiría hasta que las 25 observaciones tengan asignada una posición en la secuencia de prueba.¹ Muchos paquetes de software de computadora para ayudar a los experimentadores a seleccionar y construir un diseño, aleatorizan el orden de las corridas utilizando números aleatorios de esta manera.

Suponga que la secuencia de prueba obtenida es

Secuencia de prueba	Número de corrida	Peso porcentual del algodón
1	8	20
2	18	30
3	10	20
4	23	35
5	17	30
6	5	15
7	14	25
8	6	20
9	15	25
10	20	30
11	9	20
12	4	15
13	12	25
14	7	20
15	1	15
16	24	35
17	21	35
18	11	25
19	2	15
20	13	25
21	22	35
22	16	30
23	25	35
24	19	30
25	3	15

Esta secuencia de prueba aleatorizada es necesaria para evitar que los efectos de variables perturbadoras desconocidas —las cuales quizá varíen fuera de control durante el experimento— contaminen los resultados. Para ilustrar esto, suponga que las 25 corridas de prueba tuvieran que realizarse en el orden original no aleatorizado (es decir, primero se prueban los cinco ejemplares con 15% de algodón, después se prue-

¹ La única restricción sobre la aleatorización en este caso, es que si se saca de nuevo el mismo número (es decir, 8), se descarta. Se trata de una restricción secundaria y se ignora.

Tabla 3-1 Datos (en lb/pulgada²) del experimento de la resistencia a la tensión

Peso porcentual del algodón	Observaciones					Total	Promedio
	1	2	3	4	5		
15	7	7	15	11	9	49	9.8
20	12	17	12	18	18	77	15.4
25	14	18	18	19	19	88	17.6
30	19	25	22	19	23	108	21.6
35	7	10	11	15	11	54	10.8
						376	15.04

ban los cinco ejemplares con 20% de algodón, etc.). Si la máquina empleada para probar la resistencia a la tensión presenta un efecto de calentamiento tal que entre más tiempo esté funcionando sean menores las lecturas de la resistencia a la tensión observadas, el efecto del calentamiento contaminará potencialmente los datos de la resistencia a la tensión y destruirá la validez del experimento.

Suponga que el ingeniero corre la prueba en el orden aleatorio que se ha determinado. En la tabla 3-1 se muestran las observaciones que obtiene para la resistencia a la tensión.

Siempre es una buena idea examinar **gráficamente** los datos experimentales. En la figura 3-1 se muestran los **diagramas de caja** para la resistencia a la tensión con cada nivel del peso porcentual de algodón, y en la figura 3-2 se ilustra un **diagrama de dispersión** de la resistencia a la tensión contra el peso porcentual del algodón. En la figura 3-2, los puntos rellenos son las observaciones individuales y los círculos huecos son los promedios de la resistencia a la tensión observada. Ambas gráficas indican que la resistencia a la tensión se incrementa cuando el contenido de algodón se incrementa, hasta cerca de 30% de algodón. Después de 30% de algodón, hay un marcado descenso de la resistencia a la tensión. No hay evidencia sólida que sugiera que la variabilidad de la resistencia a la tensión alrededor del promedio dependa del peso porcentual del algodón. Con base en este análisis gráfico simple, se tienen firmes sospechas de que 1) el contenido de algodón afecta la resistencia a la tensión y 2) alrededor de 30% de algodón produce la resistencia máxima.

Suponga que se quiere ser más **objetivo** en el análisis de los datos. Específicamente, imagine que quieren probarse las diferencias entre las resistencias a la tensión promedio con todos los niveles $\alpha = 5$ del

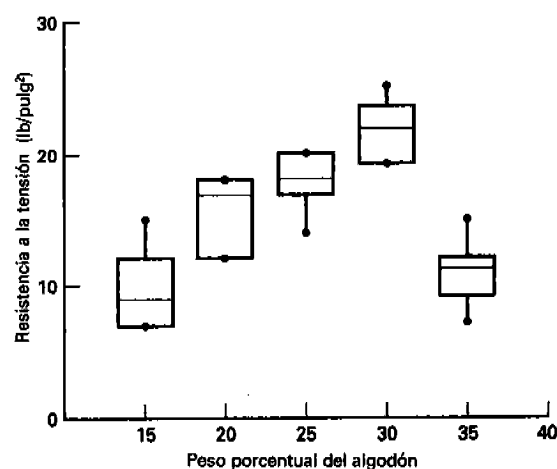


Figura 3-1 Diagramas de caja de la resistencia a la tensión contra el peso porcentual del algodón.

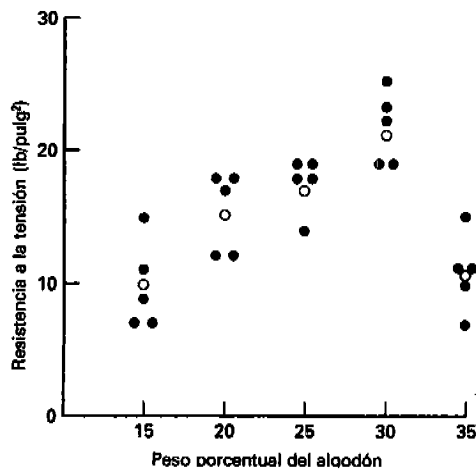


Figura 3-2 Diagrama de dispersión de la resistencia a la tensión contra el peso porcentual del algodón.

peso porcentual del algodón. Por lo tanto, el interés se centra en probar la igualdad de las cinco medias. Pudiera parecer que este problema se resolvería realizando una prueba t para todos los pares de medias posibles. Sin embargo, no es ésta la mejor solución de este problema, porque llevaría a una distorsión considerable en el error tipo I. Por ejemplo, suponga que quiere probarse la igualdad de las cinco medias usando comparaciones por pares. Hay 10 pares posibles, y si la probabilidad de aceptar correctamente la hipótesis nula en cada prueba individual es de $1 - \alpha = 0.95$, la probabilidad de aceptar correctamente la hipótesis nula en las 10 pruebas es de $(0.95)^{10} = 0.60$ si las pruebas son independientes. Por lo tanto, ha ocurrido un incremento sustancial en el error tipo I.

El procedimiento correcto para probar la igualdad de varias medias es el **análisis de varianza**. Sin embargo, el análisis de varianza tiene un rango de aplicaciones mucho más amplio que el problema anterior. Probablemente sea la técnica más útil en el campo de la inferencia estadística.

3-2 EL ANÁLISIS DE VARIANZA

Suponga que se tienen a **tratamientos** o **niveles** diferentes de **un solo factor** que quieren compararse. La respuesta observada de cada uno de los a tratamientos es una variable aleatoria. Los datos aparecerían como en la tabla 3-2. Una entrada de la tabla 3-2 (por ejemplo, y_{ij}) representa la observación j -ésima to-

Tabla 3-2 Datos típicos de un experimento de un solo factor

Tratamiento (nivel)	Observaciones				Totales	Promedios
1	y_{11}	y_{12}	...	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	...	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
\vdots	\vdots	\vdots	...	\vdots	\vdots	\vdots
a	y_{a1}	y_{a2}	...	y_{an}	$y_{a.}$	$\bar{y}_{a.}$
					$y_{..}$	$\bar{y}_{..}$

mada bajo el nivel del factor o tratamiento i . Habrá, en general, n observaciones bajo el tratamiento i -ésimo. Observe que la tabla 3-2 es el caso general de los datos del experimento de la resistencia a la tensión de la tabla 3-1.

Modelos para los datos

Se encontrará útil describir las observaciones de un experimento con un modelo. Una manera de escribir este modelo es

$$y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad (3-1)$$

donde y_{ij} es la observación ij -ésima, μ_i es la media del nivel del factor o tratamiento i -ésimo, y ε_{ij} es un componente del **error aleatorio** que incorpora todas las demás fuentes de variabilidad del experimento, incluyendo las mediciones, la variabilidad que surge de factores no controlados, las diferencias entre las unidades experimentales (como los materiales de prueba, etc.) a las que se aplican los tratamientos, y el ruido de fondo general en el proceso (ya sean la variabilidad con el tiempo, los efectos de variables ambientales, etc.). Es conveniente considerar que los errores tienen media cero, de tal modo que $E(y_{ij}) = \mu_i$.

A la ecuación 3-1 se le llama el **modelo de las medias**. Una forma alternativa de escribir un modelo de los datos es definiendo

$$\mu_i = \mu + \tau_i, \quad i = 1, 2, \dots, a$$

de tal modo que la ecuación 3-1 se convierte en

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad (3-2)$$

En esta forma del modelo, μ es un parámetro común a todos los tratamientos al que se llama la **media global**, y τ_i es un parámetro único del tratamiento i -ésimo al que se le llama el **efecto del tratamiento i -ésimo**. A la ecuación 3-2 se le llama por lo general el **modelo de los efectos**.

Tanto el modelo de las medias como el de los efectos son **modelos estadísticos lineales**; es decir, la variable de respuesta y_{ij} es una función lineal de los parámetros del modelo. Aun cuando ambas formas del modelo son útiles, el modelo de los efectos se encuentra con mayor frecuencia en la literatura del diseño experimental. Tiene cierto atractivo intuitivo por cuanto μ es una constante y los efectos de los tratamientos τ_i representan desviaciones de esta constante cuando se aplican los tratamientos específicos.

A la ecuación 3-2 (o a la 3-1) se le llama también el modelo del **análisis de varianza simple** o de un **solo factor** (o **dirección**), porque únicamente se investiga un factor. Además, será un requisito que el experimento se lleve a cabo en orden aleatorio para que el ambiente en el que se apliquen los tratamientos (llamados con frecuencia **unidades experimentales**) sea lo más uniforme posible. Por lo tanto, el diseño experimental es un **diseño completamente aleatorizado**. Los objetivos serán probar las hipótesis apropiadas acerca de las medias de los tratamientos y estimarlas. Para probar las hipótesis, se supone que los errores del modelo son variables aleatorias que siguen una distribución normal e independiente con media cero y varianza σ^2 . Se supone asimismo que la varianza σ^2 es constante para todos los niveles del factor. Esto implica que las observaciones

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

y que las observaciones son mutuamente independientes.

¿Factor fijo o aleatorio?

El modelo estadístico (ecuación 3-2) describe dos situaciones diferentes con respecto a los efectos de los tratamientos. Primera, los a tratamientos pudieron ser elegidos expresamente por el experimentador. En esta situación quieren probarse hipótesis acerca de las medias de los tratamientos, y las conclusiones se aplicarán únicamente a los niveles del factor considerados en el análisis. Las conclusiones no pueden extenderse a tratamientos similares que no fueron considerados explícitamente. También se podría querer estimar los parámetros del modelo (μ, τ_i, σ^2) . A éste se le llama el **modelo con efectos fijos**. De manera alternativa, los a tratamientos podrían ser una **muestra aleatoria** de una población más grande de tratamientos. En esta situación sería deseable poder extender las conclusiones (las cuales se basan en la muestra de los tratamientos) a la totalidad de los tratamientos de la población, sea que se hayan considerado explícitamente en el análisis o no. Aquí las τ_i son variables aleatorias, y el conocimiento de las τ_i particulares que se investigaron es relativamente inútil. Más bien, se prueban hipótesis acerca de la variabilidad de las τ_i y se intenta estimar su variabilidad. A éste se le llama el **modelo con efectos aleatorios** o **modelo de los componentes de la varianza**. La revisión de experimentos con factores aleatorios se pospondrá hasta el capítulo 12.

3-3 ANÁLISIS DEL MODELO CON EFECTOS FIJOS

En esta sección se desarrolla el análisis de varianza de un solo factor para el modelo con efectos fijos. Recuerde que $y_{i.}$ representa el total de las observaciones bajo el tratamiento i -ésimo. Sea que $\bar{y}_{i.}$ represente el promedio de las observaciones bajo el tratamiento i -ésimo. De manera similar, sea que $y_{..}$ represente el gran total de todas las observaciones y que $\bar{y}_{..}$ represente el gran promedio de todas las observaciones. Expresado simbólicamente,

$$\begin{aligned} y_{i.} &= \sum_{j=1}^n y_{ij}, & \bar{y}_{i.} &= y_{i.} / n & i &= 1, 2, \dots, a \\ y_{..} &= \sum_{i=1}^a \sum_{j=1}^n y_{ij} & \bar{y}_{..} &= y_{..} / N \end{aligned} \quad (3-3)$$

donde $N = an$ es el número total de observaciones. Se nota que el subíndice "punto" implica la operación suma sobre el subíndice que reemplaza.

El interés se encuentra en probar la igualdad de las a medias de los tratamientos; es decir, $E(y_{ij}) = \mu + \tau_i = \mu_i$, $i = 1, 2, \dots, a$. Las hipótesis apropiadas son

$$\begin{aligned} H_0: \mu_1 &= \mu_2 = \dots = \mu_a \\ H_1: \mu_i &\neq \mu_j \quad \text{para al menos un par } (i, j) \end{aligned}$$

En el modelo de los efectos, la media μ_i del tratamiento i -ésimo se descompone en dos componentes tales que $\mu_i = \mu + \tau_i$. Por lo general, μ se considera como una media global, de tal modo que

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

Esta definición implica que

$$\sum_{i=1}^a \tau_i = 0$$

Es decir, los efectos del tratamiento o factor pueden considerarse como desviaciones de la media global.² Por consiguiente, una forma equivalente de escribir las hipótesis anteriores es en términos de los efectos de los tratamientos τ_i , por ejemplo

$$\begin{aligned} H_0: \tau_1 = \tau_2 = \dots \tau_a = 0 \\ H_1: \tau_i \neq 0 \quad \text{para al menos una } i \end{aligned}$$

Por lo tanto, se habla de probar la igualdad de las medias de los tratamientos o de probar que los efectos de los tratamientos (las τ_i) son cero. El procedimiento apropiado para probar la igualdad de las medias de los a tratamientos es el análisis de varianza.

3-3.1 Descomposición de la suma de cuadrados total

El nombre **análisis de varianza** se deriva de la partición de la variabilidad total en sus partes componentes. La suma de cuadrados total corregida

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

se usa como una medida de la variabilidad global de los datos. Intuitivamente, esto es razonable porque, si SS_T tuviera que dividirse por el número apropiado de grados de libertad (en este caso, $an - 1 = N - 1$), se obtendría la **varianza muestral** de las y . La varianza muestral es, desde luego, una medida estándar de variabilidad.

Observe que la suma de cuadrados total corregida SS_T se puede escribir como

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n [(\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})]^2 \quad (3-4)$$

o

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 &= n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 \\ &\quad + 2 \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) \end{aligned} \quad (3-5)$$

Sin embargo, el término del producto cruzado de la ecuación 3-5 es cero, ya que

$$\sum_{j=1}^n (y_{ij} - \bar{y}_{i.}) = y_{i.} - n\bar{y}_{i.} = y_{i.} - n(y_{i.} / n) = 0$$

Se tiene, por lo tanto,

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 \quad (3-6)$$

La ecuación 3-6 establece que puede hacerse la partición de la variabilidad total de los datos, medida por la suma de cuadrados total corregida, en una suma de cuadrados de las diferencias **entre** los promedios de los tratamientos y el gran promedio, más una suma de cuadrados de las diferencias de las observaciones **dentro** de los tratamientos y el promedio de los tratamientos. Entonces, la diferencia entre los promedios

² Para más información sobre este tema, referirse al material suplementario del texto del capítulo 3.

de los tratamientos observados y el gran promedio es una medida de las diferencias entre las medias de los tratamientos, mientras que las diferencias de las observaciones dentro de un tratamiento y el promedio del tratamiento, pueden deberse únicamente al error aleatorio. Por lo tanto, la ecuación 3-6 puede escribirse simbólicamente como

$$SS_T = SS_{\text{Tratamientos}} + SS_E$$

donde a $SS_{\text{Tratamientos}}$ se le llama la suma de cuadrados debida a los tratamientos (es decir, entre los tratamientos), y a SS_E se le llama la suma de cuadrados debida al error (es decir, dentro de los tratamientos). Hay $an = N$ observaciones en total; por lo tanto, SS_T tiene $N - 1$ grados de libertad. Hay a niveles del factor (y medias de a tratamientos), de donde $SS_{\text{Tratamientos}}$ tiene $a - 1$ grados de libertad. Por último, dentro de cualquier tratamiento hay n réplicas que proporcionan $n - 1$ grados de libertad con los cuales estimar el error experimental. Puesto que hay a tratamientos, se tienen $a(n - 1) = an - a = N - a$ grados de libertad para el error.

Es útil examinar explícitamente los dos términos del lado derecho de la identidad fundamental del análisis de varianza (ecuación 3-6). Considere la suma de cuadrados del error

$$SS_E = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^a \left[\sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 \right]$$

En esta forma es fácil ver que el término entre corchetes, si se divide por $n - 1$, es la varianza muestral del tratamiento i -ésimo, o

$$S_i^2 = \frac{\sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{n - 1} \quad i = 1, 2, \dots, a$$

Ahora pueden combinarse a varianzas muestrales para obtener una sola estimación de la varianza poblacional común de la siguiente manera:

$$\begin{aligned} \frac{(n-1)S_1^2 + (n-1)S_2^2 + \dots + (n-1)S_a^2}{(n-1) + (n-1) + \dots + (n-1)} &= \frac{\sum_{i=1}^a \left[\sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 \right]}{\sum_{i=1}^a (n-1)} \\ &= \frac{SS_E}{(N-a)} \end{aligned}$$

Por lo tanto, $SS_E/(N - a)$ es una estimación combinada de la varianza común dentro de cada uno de los a tratamientos.

De manera similar, si no hubiera diferencias entre las medias de los a tratamientos, podría usarse la variación de los promedios de los tratamientos y el gran promedio para estimar σ^2 . Específicamente

$$\frac{SS_{\text{Tratamientos}}}{a-1} = \frac{n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2}{a-1}$$

es una estimación de σ^2 si las medias de los tratamientos son iguales. La razón de esto puede verse de manera intuitiva de la siguiente manera. La cantidad $\sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2 / (a - 1)$ estima σ^2/n , la varianza de los pro-

medios de los tratamientos, de donde $n \sum_{i=1}^a (\bar{y}_i - \bar{y})^2 / (a - 1)$ debe estimar σ^2 si no hay diferencias en las medias de los tratamientos.

Se observa que la identidad del análisis de varianza (ecuación 3-6) nos proporciona dos estimaciones de σ^2 : una basada en la variabilidad inherente dentro de los tratamientos y una basada en la variabilidad entre los tratamientos. Si no hay diferencias en las medias de los tratamientos, estas dos estimaciones deberán ser muy similares, y si no lo son, se sospecha que la diferencia observada puede ser causada por diferencias en las medias de los tratamientos. Aun cuando se ha usado un razonamiento intuitivo para desarrollar este resultado, puede adoptarse un enfoque un tanto más formal.

A las cantidades

$$MS_{\text{Tratamientos}} = \frac{SS_{\text{Tratamientos}}}{a - 1}$$

y

$$MS_E = \frac{SS_E}{N - a}$$

se les llama **cuadrados medios**. Se examinarán ahora los **valores esperados** de estos cuadrados medios. Considere

$$\begin{aligned} E(MS_E) &= E\left(\frac{SS_E}{N - a}\right) = \frac{1}{N - a} E\left[\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2\right] \\ &= \frac{1}{N - a} E\left[\sum_{i=1}^a \sum_{j=1}^n (y_{ij}^2 - 2y_{ij}\bar{y}_i + \bar{y}_i^2)\right] \\ &= \frac{1}{N - a} E\left[\sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - 2n \sum_{i=1}^a \bar{y}_i^2 + n \sum_{i=1}^a \bar{y}_i^2\right] \\ &= \frac{1}{N - a} E\left[\sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{1}{n} \sum_{i=1}^a y_i^2\right] \end{aligned}$$

Al sustituir el modelo (ecuación 3-1) en esta ecuación se obtiene

$$E(MS_E) = \frac{1}{N - a} E\left[\sum_{i=1}^a \sum_{j=1}^n (\mu + \tau_i + \varepsilon_{ij})^2 - \frac{1}{n} \sum_{i=1}^a \left(\sum_{j=1}^n \mu + \tau_i + \varepsilon_{ij}\right)^2\right]$$

Entonces, al elevar al cuadrado y tomar el valor esperado de la cantidad entre corchetes, se observa que los términos que incluyen a ε_{ij}^2 y ε_i^2 son reemplazados por σ^2 y $n\sigma^2$, respectivamente, debido a que $E(\varepsilon_{ij}) = 0$. Además, todos los productos cruzados que incluyen a ε_{ij} tienen valor esperado cero. Por lo tanto, después de elevar al cuadrado y tomar el valor esperado, la última ecuación se convierte en

$$E(MS_E) = \frac{1}{N - a} \left[N\mu^2 + n \sum_{i=1}^a \tau_i^2 + N\sigma^2 - N\mu^2 - n \sum_{i=1}^a \tau_i^2 - a\sigma^2 \right]$$

o

$$E(MS_E) = \sigma^2$$

Aplicando un enfoque similar puede demostrarse también que³

$$E(MS_{\text{Tratamientos}}) = \sigma^2 + \frac{n \sum_{i=1}^a \tau_i^2}{a-1}$$

Por lo tanto, como se argumentó heurísticamente, $MS_E = SS_E/(N-a)$ estima σ^2 y, si no hay diferencias en las medias de los tratamientos (lo cual implica que $\tau_i = 0$), $MS_{\text{Tratamientos}} = SS_{\text{Tratamientos}}/(a-1)$ también estima σ^2 . Sin embargo, observe que si las medias de los tratamientos difieren, el valor esperado del cuadrado medio de los tratamientos es mayor que σ^2 .

Parece claro que es posible realizar una prueba de la hipótesis de que no hay diferencias en las medias de los tratamientos comparando $MS_{\text{Tratamientos}}$ y MS_E . Se considera ahora cómo puede hacerse esta comparación.

3-3.2 Análisis estadístico

Se investiga ahora cómo puede llevarse a cabo una prueba formal de la hipótesis de que no hay diferencias en las medias de los tratamientos ($H_0: \mu_1 = \mu_2 = \dots = \mu_a$, o de manera equivalente, $H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$). Puesto que se ha supuesto que los errores ε_{ij} siguen una distribución normal e independiente con media cero y varianza σ^2 , las observaciones y_{ij} tienen una distribución normal e independiente con media $\mu + \tau_i$ y varianza σ^2 . Por lo tanto, SS_T es una suma de cuadrados de variables aleatorias con una distribución normal; por consiguiente, puede demostrarse que SS_T/σ^2 tiene una distribución ji-cuadrada con $N-1$ grados de libertad. Además, puede demostrarse que SS_E/σ^2 es una variable ji-cuadrada con $N-a$ grados de libertad y que $SS_{\text{Tratamientos}}/\sigma^2$ es una variable ji-cuadrada con $a-1$ grados de libertad si la hipótesis nula $H_0: \tau_i = 0$ es verdadera. Sin embargo, las tres sumas de cuadrados no son necesariamente independientes, ya que la suma de $SS_{\text{Tratamientos}}$ y SS_E es SS_T . El siguiente teorema, que es una forma especial de un teorema atribuido a William Cochran, es útil para establecer la independencia de SS_E y $SS_{\text{Tratamientos}}$.

TEOREMA 3-1

Teorema de Cochran

Sea Z_i igual a NID(0, 1) para $i = 1, 2, \dots, v$ y

$$\sum_{i=1}^v Z_i^2 = Q_1 + Q_2 + \dots + Q_s$$

donde $s \leq v$, y Q_i tiene ν_i grados de libertad ($i = 1, 2, \dots, s$). Entonces Q_1, Q_2, \dots, Q_s son variables aleatorias ji-cuadrada independientes con $\nu_1, \nu_2, \dots, \nu_s$ grados de libertad, respectivamente, si y sólo si

$$v = \nu_1 + \nu_2 + \dots + \nu_s$$

.....

Puesto que los grados de libertad de $SS_{\text{Tratamientos}}$ y SS_E suman $N-1$, el número total de grados de libertad, el teorema de Cochran implica que $SS_{\text{Tratamientos}}/\sigma^2$ y SS_E/σ^2 son variables aleatorias ji-cuadrada con

³ Referirse al material suplementario del texto del capítulo 3.

Tabla 3-3 Tabla de análisis de varianza para el modelo con un solo factor y efectos fijos

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0
Entre los tratamientos	$SS_{\text{Tratamientos}} = n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2$	$a-1$	$MS_{\text{Tratamientos}}$	$F_0 = \frac{MS_{\text{Tratamientos}}}{MS_E}$
Error (dentro de los tratamientos)	$SS_E = SS_T - SS_{\text{Tratamientos}}$	$N-a$	MS_E	
Total	$SS_T = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$	$N-1$		

una distribución independiente. Por lo tanto, si la hipótesis nula de que no hay diferencias en las medias de los tratamientos es verdadera, el cociente

$$F_0 = \frac{SS_{\text{Tratamientos}} / (a-1)}{SS_E / (N-a)} = \frac{MS_{\text{Tratamientos}}}{MS_E} \quad (3-7)$$

se distribuye como F con $a-1$ y $N-a$ grados de libertad. La ecuación 3-7 es el **estadístico de prueba** para la hipótesis de que no hay diferencias en las medias de los tratamientos.

Por los cuadrados medios esperados se observa que, en general, MS_E es un estimador insesgado de σ^2 . Asimismo, bajo la hipótesis nula, $MS_{\text{Tratamientos}}$ es un estimador insesgado de σ^2 . Sin embargo, si la hipótesis nula es falsa, el valor esperado de $MS_{\text{Tratamientos}}$ es mayor que σ^2 . Por lo tanto, bajo la hipótesis alternativa, el valor esperado del numerador del estadístico de prueba (ecuación 3-7) es mayor que el valor esperado del denominador, y H_0 deberá rechazarse para valores del estadístico de prueba que son muy grandes. Esto implica una región crítica de una sola cola superior. Por lo tanto, H_0 deberá rechazarse y concluirse que hay diferencias en las medias de los tratamientos si

$$F_0 > F_{\alpha, a-1, N-a}$$

donde F_0 se calcula con la ecuación 3-7. De manera alternativa, podría usarse el enfoque del valor P para tomar una decisión.

Es posible obtener fórmulas para calcular estas sumas de cuadrados reescribiendo y simplificando las definiciones de $SS_{\text{Tratamientos}}$ y SS_T en la ecuación 3-6. Se obtiene así

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{N} \quad (3-8)$$

y

$$SS_{\text{Tratamientos}} = \frac{1}{n} \sum_{i=1}^a y_i^2 - \frac{y_{..}^2}{N} \quad (3-9)$$

La suma de cuadrados del error se obtiene por sustracción como

$$SS_E = SS_T - SS_{\text{Tratamientos}} \quad (3-10)$$

El procedimiento de prueba se resume en la tabla 3-3. Se le conoce como **tabla del análisis de varianza**.

EJEMPLO 3-1

El experimento de la resistencia a la tensión

Para ilustrar el análisis de varianza, se retoma al ejemplo que empezó a comentarse en la sección 3-1. Recuerde que al ingeniero de desarrollo de productos le interesa determinar si el peso porcentual del algo-

dón en una fibra sintética afecta la resistencia a la tensión, y ha llevado a cabo un experimento completamente aleatorizado con cinco niveles del peso porcentual del algodón y cinco réplicas. Por conveniencia, a continuación se repiten los datos de la tabla 3-1:

Peso porcentual del algodón	Resistencia a la tensión observada (lb/pulg ²)					Totales y_i	Promedios \bar{y}_i
	1	2	3	4	5		
15	7	7	15	11	9	49	9.8
20	12	17	12	18	18	77	15.4
25	14	18	18	19	19	88	17.6
30	19	25	22	19	23	108	21.6
35	7	10	11	15	11	54	10.8
						$y_{..} = 376$	$\bar{y}_{..} = 15.04$

Se usará el análisis de varianza para probar $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ contra la hipótesis alternativa H_1 : algunas medias son diferentes. Las sumas de cuadrados requeridas se calculan como sigue:

$$\begin{aligned}
 SS_T &= \sum_{i=1}^5 \sum_{j=1}^5 y_{ij}^2 - \frac{y_{..}^2}{N} \\
 &= (7)^2 + (7)^2 + (15)^2 + \dots + (15)^2 + (11)^2 - \frac{(376)^2}{25} = 636.96 \\
 SS_{\text{Tratamientos}} &= \frac{1}{n} \sum_{i=1}^5 y_i^2 - \frac{y_{..}^2}{N} \\
 &= \frac{1}{5} [(49)^2 + \dots + (54)^2] - \frac{(376)^2}{25} = 475.76 \\
 SS_E &= SS_T - SS_{\text{Tratamientos}} \\
 &= 636.96 - 475.76 = 161.20
 \end{aligned}$$

Generalmente estos cálculos se realizarían en una computadora, utilizando paquetes de software con la capacidad de analizar datos de experimentos diseñados.

En la tabla 3-4 se resume el análisis de varianza. Observe que el cuadrado medio entre los tratamientos (118.94) es varias veces mayor que el cuadrado medio dentro de los tratamientos o cuadrado medio del error (8.06). Esto indica que no es posible que las medias de los tratamientos sean iguales. En términos más formales, puede calcularse el cociente F , $F_0 = 118.94/8.06 = 14.76$, y comparar este valor con un punto porcentual apropiado de la cola superior de la distribución $F_{4, 20}$. Suponga que el experimentador ha seleccionado $\alpha = 0.05$. En la tabla IV del apéndice se encuentra que $F_{0.05, 4, 20} = 2.87$. Puesto que $F_0 = 14.76 > 2.87$, se rechaza H_0 y se concluye que las medias de los tratamientos difieren; es decir, el peso porcentual del algodón en la fibra afecta de manera significativa la resistencia a la tensión media. También

Tabla 3-4 Análisis de varianza de los datos de la resistencia a la tensión

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Peso porcentual del algodón	475.76	4	118.94	$F_0 = 14.76$	< 0.01
Error	161.20	20	8.06		
Total	636.96	24			

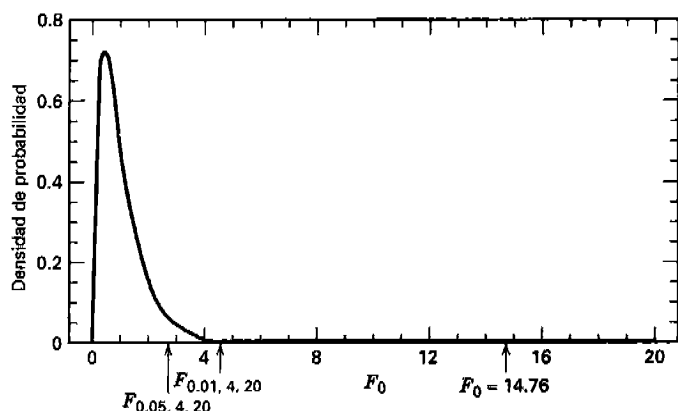


Figura 3-3 La distribución de referencia ($F_{4,20}$) para el estadístico de prueba F_0 en el ejemplo 3-1.

podría calcularse un valor P para este estadístico de prueba. En la figura 3-3 se muestra la distribución de referencia ($F_{4,20}$) para el estadístico de prueba F_0 . Evidentemente, el valor P es muy pequeño en este caso. Puesto que $F_{0.01,4,20} = 4.43$ y $F_0 > 4.43$, puede concluirse que un límite superior del valor P es 0.01; es decir, $P < 0.01$ (el valor P exacto es $P = 9.11 \times 10^{-6}$).

Cálculos manuales

Posiblemente el lector haya notado que la suma de cuadrados se definió en términos de promedios; es decir, por la ecuación 3-6,

$$SS_{\text{Tratamientos}} = n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2$$

pero las fórmulas de cálculo se desarrollaron utilizando los totales. Por ejemplo, para calcular $SS_{\text{Tratamientos}}$, se usaría la ecuación 3-9:

$$SS_{\text{Tratamientos}} = \frac{1}{n} \sum_{i=1}^a y_i^2 - \frac{y_{..}^2}{N}$$

La razón principal de esto es por conveniencia; además, los totales y_i y $y_{..}$ están menos sujetos al error de redondeo que los promedios \bar{y}_i y $\bar{y}_{..}$.

En general, no deberá prestarse demasiada atención a los cálculos, ya que se cuenta con una amplia variedad de programas de computadora para realizarlos. Estos programas de computadora son también útiles para realizar muchos otros análisis asociados con el diseño experimental (como el análisis residual y la verificación de la adecuación del modelo). En muchos casos, estos programas también ayudarán al experimentador a establecer el diseño.

Cuando es necesario realizar los cálculos manualmente, en ocasiones es útil codificar las observaciones. Esto se ilustra en el ejemplo siguiente.

EJEMPLO 3-2

Codificación de observaciones

Los cálculos del análisis de varianza pueden hacerse con frecuencia de manera más precisa o simplificada codificando las observaciones. Por ejemplo, considere los datos de la resistencia a la tensión del ejemplo

Tabla 3-5 Datos codificados de la resistencia a la tensión del ejemplo 3-2

Peso porcentual del algodón	Observaciones					Totales y_i
	1	2	3	4	5	
15	-8	-8	0	-4	-6	-26
20	-3	2	-3	3	3	2
25	-1	3	3	4	4	13
30	4	10	7	4	8	33
35	-8	-5	-4	0	-4	-21

3-1. Suponga que se resta 15 de cada observación. Los datos codificados se muestran en la tabla 3-5. Es sencillo verificar que

$$SS_T = (-8)^2 + (-8)^2 + \dots + (-4)^2 - \frac{(1)^2}{25} = 636.96$$

$$SS_{\text{Tratamientos}} = \frac{(-26)^2 + (2)^2 + \dots + (-21)^2}{5} - \frac{(1)^2}{25} = 475.76$$

y

$$SS_E = 161.20$$

Al comparar estas sumas de cuadrados con las que se obtuvieron en el ejemplo 3-1, se observa que al restar una constante de los datos originales las sumas de cuadrados no se modifican.

Suponga ahora que cada una de las observaciones del ejemplo 3-1 se multiplica por 2. Es sencillo verificar que las sumas de cuadrados de los datos transformados son $SS_T = 2547.84$, $SS_{\text{Tratamientos}} = 1903.04$ y $SS_E = 644.80$. Estas sumas de cuadrados parecen diferir considerablemente de las que se obtuvieron en el ejemplo 3-1. Sin embargo, si se dividen por 4 (es decir, 2^2), los resultados son idénticos. Por ejemplo, para la suma de cuadrados de los tratamientos, $1903.04/4 = 475.76$. Asimismo, para los datos codificados, el cociente F es $F = (1903.04/4)/(644.80/20) = 14.76$, que es idéntico al cociente F de los datos originales. Por lo tanto, los análisis de varianza son equivalentes.

Pruebas de aleatorización y análisis de varianza

En el desarrollo del análisis de varianza con la prueba F , se ha utilizado el supuesto de que los errores aleatorios ε_{ij} son variables aleatorias que siguen una distribución normal e independiente. También es posible justificar la prueba F como la aproximación de una **prueba de aleatorización**. Para ilustrar esto, suponga que se tienen cinco observaciones de cada uno de dos tratamientos y que quiere probarse la igualdad de las medias de los tratamientos. Los datos aparecerían así:

Tratamiento 1	Tratamiento 2
y_{11}	y_{21}
y_{12}	y_{22}
y_{13}	y_{23}
y_{14}	y_{24}
y_{15}	y_{25}

Podría usarse el análisis de varianza con la prueba F para probar $H_0: \mu_1 = \mu_2$. De manera alternativa, podría recurrirse a un enfoque un tanto diferente. Suponga que se consideran todas las formas posibles de

asignar los 10 números de la muestra anterior a los dos tratamientos. Hay $10!/5!5! = 252$ arreglos posibles de las 10 observaciones. Si no hay ninguna diferencia en las medias de los tratamientos, los 252 arreglos son igualmente posibles. Para cada uno de los 252 arreglos, se calcula el valor del estadístico F usando la ecuación 3-7. A la distribución de estos valores F se le llama **distribución de aleatorización**, y un valor grande de F indica que los datos no son consistentes con la hipótesis $H_0: \mu_1 = \mu_2$. Por ejemplo, si el valor de F que se observó realmente fue excedido sólo por 5 de los valores de la distribución de aleatorización, esto correspondería con el rechazo de $H_0: \mu_1 = \mu_2$ con un nivel de significación de $\alpha = 5/252 = 0.0198$ (o 1.98%). Observe que no es necesario ningún supuesto de normalidad en este enfoque.

La dificultad con este enfoque es que, incluso en problemas relativamente pequeños, los cálculos requeridos hacen inviable la enumeración de la distribución de aleatorización exacta. Sin embargo, numerosos estudios han demostrado que la distribución F común de la teoría normal es una buena aproximación de la distribución de aleatorización exacta. Por lo tanto, incluso sin el supuesto de normalidad, la prueba F puede considerarse como una aproximación de la prueba de aleatorización. Para más detalles sobre las pruebas de aleatorización en el análisis de varianza, ver Box, Hunter y Hunter [18].

3-3.3 Estimación de los parámetros del modelo

Se presentan ahora los estimadores de los parámetros del modelo con un solo factor

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

y los intervalos de confianza para las medias de los tratamientos. Más adelante se demostrará que estimadores razonables de la media global y de los efectos de los tratamientos están dados por

$$\begin{aligned}\hat{\mu} &= \bar{y}_{..} \\ \hat{\tau}_i &= \bar{y}_{i.} - \bar{y}_{..}, \quad i = 1, 2, \dots, a\end{aligned}\tag{3-11}$$

Estos estimadores poseen un considerable atractivo intuitivo; observe que la media global se estima con el gran promedio de las observaciones y que el efecto de cualquier tratamiento no es sino la diferencia entre el promedio del tratamiento y el gran promedio.

Es posible determinar con facilidad una estimación del **intervalo de confianza** de la media del tratamiento i -ésimo. La media del tratamiento i -ésimo es

$$\mu_i = \mu + \tau_i$$

Un estimador puntual de μ_i sería $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i.}$. Ahora bien, si se supone que los errores siguen una distribución normal, cada $\bar{y}_{i.}$ es una $NID(\mu_i, \sigma^2/n)$. Por lo tanto, si σ^2 fuera conocida, podría usarse la distribución normal para definir el intervalo de confianza. Al utilizar MS_E como estimador de σ^2 , el intervalo de confianza se basaría en la distribución t . Por lo tanto, un intervalo de confianza de $100(1 - \alpha)$ por ciento para la media μ_i del tratamiento i -ésimo es

$$\bar{y}_{i.} - t_{\alpha/2, N-a} \sqrt{\frac{MS_E}{n}} \leq \mu_i \leq \bar{y}_{i.} + t_{\alpha/2, N-a} \sqrt{\frac{MS_E}{n}}\tag{3-12}$$

Un intervalo de confianza de $100(1 - \alpha)$ por ciento para la diferencia en las medias de dos tratamientos cualesquiera, por ejemplo $\mu_i - \mu_j$, sería

$$\bar{y}_{i.} - \bar{y}_{j.} - t_{\alpha/2, N-a} \sqrt{\frac{2MS_E}{n}} \leq \mu_i - \mu_j \leq \bar{y}_{i.} - \bar{y}_{j.} + t_{\alpha/2, N-a} \sqrt{\frac{2MS_E}{n}}\tag{3-13}$$

EJEMPLO 3-3

Utilizando los datos del ejemplo 3-1 pueden encontrarse las estimaciones de la media global y de los efectos de los tratamientos como $\hat{\mu} = 376/25 = 15.04$ y

$$\hat{\tau}_1 = \bar{y}_1 - \bar{y}_{..} = 9.80 - 15.04 = -5.24$$

$$\hat{\tau}_2 = \bar{y}_2 - \bar{y}_{..} = 15.40 - 15.04 = +0.36$$

$$\hat{\tau}_3 = \bar{y}_3 - \bar{y}_{..} = 17.60 - 15.04 = -2.56$$

$$\hat{\tau}_4 = \bar{y}_4 - \bar{y}_{..} = 21.60 - 15.04 = +6.56$$

$$\hat{\tau}_5 = \bar{y}_5 - \bar{y}_{..} = 10.80 - 15.04 = -4.24$$

Un intervalo de confianza de 95% para la media del tratamiento 4 (30% de algodón) se calcula con la ecuación 3-12 como

$$21.60 - 2.086\sqrt{\frac{8.06}{5}} \leq \mu_4 \leq 21.60 + 2.086\sqrt{\frac{8.06}{5}}$$

o

$$21.60 - 2.65 \leq \mu_4 \leq 21.60 + 2.65$$

Por lo tanto, el intervalo de confianza de 95% deseado es $18.95 \leq \mu_4 \leq 24.25$.

.....

Intervalos de confianza simultáneos

Las expresiones para los intervalos de confianza dados en las ecuaciones 3-12 y 3-13 son intervalos de confianza **uno a la vez**. Es decir, el nivel de confianza $1 - \alpha$ sólo se aplica a una estimación particular. Sin embargo, en muchos problemas, el experimentador tal vez quiera calcular varios intervalos de confianza, uno para cada una de varias medias o diferencias entre medias. Si hay interés en r de estos intervalos de confianza de $100(1 - \alpha)$ por ciento, la probabilidad de que los r intervalos sean correctos **simultáneamente** es al menos $1 - r\alpha$. A la probabilidad $r\alpha$ se le llama con frecuencia **índice de error en el modo del experimento** o coeficiente de confianza global. El número de intervalos r no tiene que ser muy grande antes de que el conjunto de intervalos de confianza se vuelva relativamente falto de información. Por ejemplo, si hay $r = 5$ intervalos y $\alpha = 0.05$ (una elección típica), el nivel de confianza simultáneo para el conjunto de los cinco intervalos de confianza es de al menos 0.75, y si $r = 10$ y $\alpha = 0.05$, el nivel de confianza simultáneo es de al menos 0.50.

Un enfoque para asegurarse de que el nivel de confianza simultáneo no sea demasiado pequeño es sustituir $\alpha/2$ con $\alpha/(2r)$ en las ecuaciones 3-12 y 3-13 del intervalo de confianza uno a la vez. A éste se le llama el **método de Bonferroni**, y le permite al experimentador construir un conjunto de r intervalos de confianza simultáneos para las medias de los tratamientos o las diferencias en las medias de los tratamientos para los que el nivel de confianza global es de al menos $100(1 - \alpha)$ por ciento. Cuando r no es muy grande, éste es un método muy atinado que produce intervalos de confianza razonablemente cortos. Para más información, referirse al material suplementario del texto del capítulo 3.

3-3.4 Datos no balanceados

En algunos experimentos con un solo factor, puede ser diferente el número de observaciones que se hacen dentro de cada tratamiento. Se dice entonces que el diseño es **no balanceado**. Sigue siendo posible aplicar el análisis de varianza descrito arriba, pero deben hacerse ligeras modificaciones en las fórmulas

de las sumas de cuadrados. Sea que se hagan n_i observaciones bajo el tratamiento i ($i = 1, 2, \dots, a$) y que $N = \sum_{i=1}^a n_i$. Las fórmulas para calcular manualmente SS_T y $SS_{\text{Tratamientos}}$ quedan como

$$SS_T = \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \quad (3-14)$$

y

$$SS_{\text{Tratamientos}} = \sum_{i=1}^a \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} \quad (3-15)$$

No se requieren más cambios en el análisis de varianza.

Hay dos ventajas al elegir un diseño balanceado. Primera, el estadístico de prueba es relativamente insensible a las desviaciones pequeñas del supuesto de la igualdad de las varianzas de los a tratamientos cuando los tamaños de las muestras son iguales. No es éste el caso cuando los tamaños de las muestras son diferentes. Segunda, la potencia de la prueba se maximiza cuando las muestras tienen el mismo tamaño.

3-4 VERIFICACIÓN DE LA ADECUACIÓN DEL MODELO

La descomposición de la variabilidad presente en las observaciones mediante la identidad del análisis de varianza (ecuación 3-6) es una relación puramente algebraica. Sin embargo, el uso de la partición para probar formalmente que no hay diferencias en las medias de los tratamientos requiere que se satisfagan ciertos supuestos. Específicamente, estos supuestos son que el modelo

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

describe de manera adecuada las observaciones, y que los errores siguen una distribución normal e independiente con media cero y varianza σ^2 constante pero desconocida. Si estos supuestos se satisfacen, el procedimiento del análisis de varianza es una prueba exacta de la hipótesis de que no hay diferencias en las medias de los tratamientos.

Sin embargo, es común que en la práctica estos supuestos no se satisfagan exactamente. Por consiguiente, en general no es prudente confiar en el análisis de varianza hasta haber verificado estos supuestos. Las violaciones de los supuestos básicos y la adecuación del modelo pueden investigarse con facilidad mediante el examen de los **residuales**. El residual de la observación j -ésima en el tratamiento i -ésimo se define como

$$e_{ij} = y_{ij} - \hat{y}_{ij} \quad (3-16)$$

donde \hat{y}_{ij} es una estimación de la observación y_{ij} correspondiente que se obtiene como sigue:

$$\begin{aligned} \hat{y}_{ij} &= \hat{\mu} + \hat{\tau}_i \\ &= \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) \\ &= \bar{y}_{i.} \end{aligned} \quad (3-17)$$

La ecuación 3-17 da el resultado intuitivamente claro de que la estimación de cualquier observación en el tratamiento i -ésimo no es sino el promedio del tratamiento correspondiente.

El examen de los residuales deberá ser una parte automática de cualquier análisis de varianza. Si el modelo es adecuado, los residuales deberán estar **sin estructura**; es decir, no deberán contener patrones obvios. A través de un estudio de los residuales, pueden descubrirse muchos tipos de inadecuaciones del modelo y violaciones de los supuestos subyacentes. En esta sección se indica cómo puede hacerse con facilidad la verificación de diagnóstico del modelo mediante el análisis gráfico de los residuales y cómo resolver varias anormalidades que ocurren comúnmente.

3-4.1 El supuesto de normalidad

La verificación del supuesto de normalidad podría hacerse graficando un histograma de los residuales. Si se satisface el supuesto de $NID(0, \sigma^2)$ para los errores, esta gráfica deberá aparecer como una muestra de una distribución normal con centro en cero. Desafortunadamente, cuando se trabaja con muestras pequeñas, suelen ocurrir fluctuaciones significativas, por lo que la aparición de una desviación moderada de la normalidad no implica necesariamente una violación seria de los supuestos. Las desviaciones marcadas de la normalidad son potencialmente serias y requieren análisis adicional.

Un procedimiento en extremo útil es construir una gráfica de probabilidad normal de los residuales. Recuerde que en el capítulo 2 se utilizó una gráfica de probabilidad normal de los datos originales para verificar el supuesto de normalidad cuando se usó la prueba t . En el análisis de varianza, por lo general es más eficaz (y directo) hacer lo mismo con los residuales. Si la distribución fundamental de los errores es normal, esta gráfica tendrá la apariencia de una línea recta. Para visualizar la línea recta, deberá prestarse más atención a los valores centrales de la gráfica que a los valores extremos.

En la tabla 3-6 se muestran los datos originales y los residuales de los datos de la resistencia a la tensión del ejemplo 3-1. La gráfica de probabilidad normal se muestra en la figura 3-4. La impresión general que surge al examinar esta representación es que la distribución de los errores puede tener un ligero sesgo, con la cola derecha siendo más larga que la izquierda. La tendencia de la gráfica de probabilidad normal a curvarse hacia abajo ligeramente del lado izquierdo, implica que la cola izquierda de la distribución de los errores sea un tanto más *delgada* de lo que se anticiparía con una distribución normal; es decir, los residuales negativos no son tan grandes (en valor absoluto) como se esperaba. Sin embargo, esta gráfica no muestra una desviación marcada de la distribución normal.

En general, las desviaciones moderadas de la normalidad no son motivo de gran preocupación en el análisis de varianza de efectos fijos (recuerde el análisis de las pruebas de aleatorización de la sección 3-3.2). Una distribución de los errores que tiene colas considerablemente más gruesas o delgadas que la distribución normal es motivo de mayor preocupación que una distribución sesgada. Puesto que la prueba F sólo se afecta ligeramente, se dice que el análisis de varianza (y los procedimientos relacionados como las comparaciones múltiples) es **robusto** con respecto al supuesto de normalidad. Las desviaciones de la normalidad hacen por lo general que tanto el verdadero nivel de significación como la verdadera potencia difieran ligeramente de los valores anunciados, con la potencia siendo generalmente más baja. El modelo de los efectos aleatorios que se revisará en el capítulo 12 se afecta en forma más severa por la no normalidad.

Tabla 3-6 Datos y residuales del ejemplo 3-1^a

Peso porcentual del algodón	Observaciones (j)					$\hat{y}_i = \bar{y}_i$
	1	2	3	4	5	
15	<div>-2.8 7 (15)</div>	<div>-2.8 7 (19)</div>	<div>5.2 15 (25)</div>	<div>1.2 11 (12)</div>	<div>-0.8 9 (6)</div>	9.8
20	<div>-3.4 12 (8)</div>	<div>1.6 17 (14)</div>	<div>-3.4 12 (1)</div>	<div>2.6 18 (11)</div>	<div>2.6 19 (3)</div>	15.4
25	<div>-3.6 14 (18)</div>	<div>0.4 18 (13)</div>	<div>0.4 18 (20)</div>	<div>1.4 19 (7)</div>	<div>1.4 19 (9)</div>	17.6
30	<div>-2.6 19 (22)</div>	<div>3.4 25 (5)</div>	<div>0.4 22 (2)</div>	<div>-2.6 19 (24)</div>	<div>1.4 23 (10)</div>	21.6
35	<div>-3.8 7 (17)</div>	<div>-0.8 10 (21)</div>	<div>0.2 11 (4)</div>	<div>4.2 15 (16)</div>	<div>0.2 11 (23)</div>	10.8

^aLos residuales se indican en el recuadro de cada celda. Los números entre paréntesis indican el orden en que se recolectaron los datos.

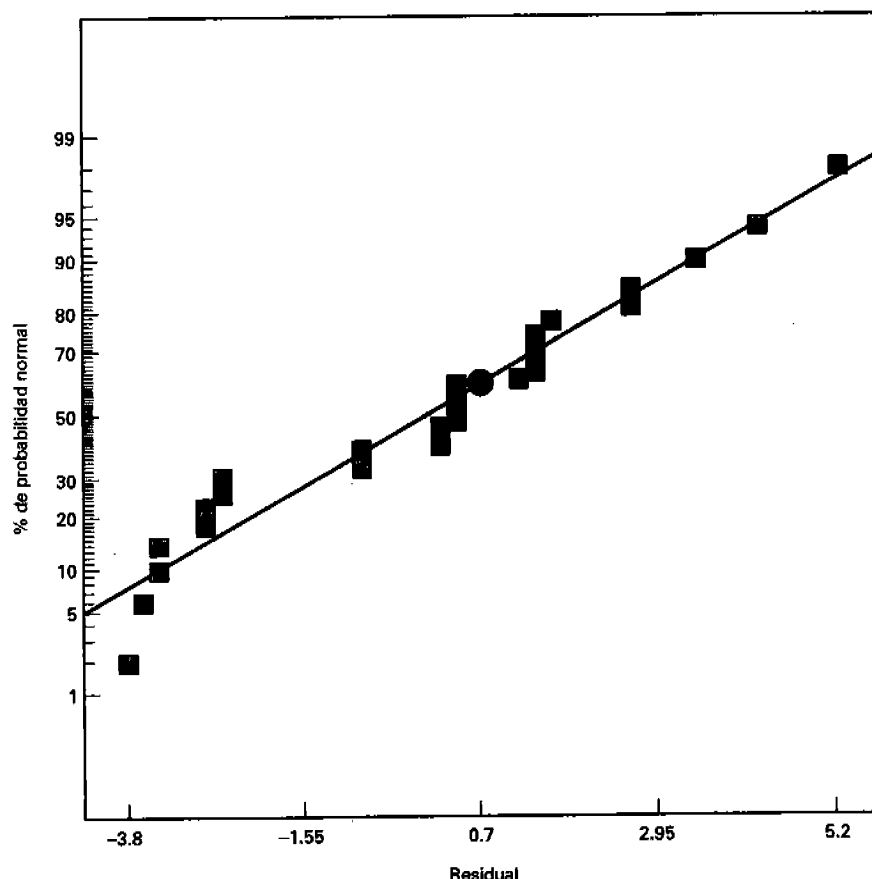


Figura 3-4 Gráfica de probabilidad normal de los residuales del ejemplo 3-1.

Una anomalía muy común que suele ponerse de manifiesto en las gráficas de probabilidad normal es un residual que es mucho más grande que cualquier otro. A un residual así se le llama con frecuencia **punto atípico**. La presencia de uno o más puntos atípicos puede introducir serias distorsiones en el análisis de varianza, por lo que cuando se localiza un punto atípico potencial, se requiere una investigación atenta. En muchas ocasiones, la causa del punto atípico es un error en los cálculos o un error al codificar o copiar los datos. Si no es ésta la causa, las circunstancias experimentales que rodean esta corrida particular deben estudiarse con atención. Si la respuesta atípica ocurre en un valor particularmente deseable (alta resistencia, costo bajo, etc.), el punto atípico puede ser más informativo que el resto de los datos. Deberá tenerse cuidado de no rechazar o descartar una observación atípica a menos que se tengan razones estadísticas de peso para hacerlo. En el peor de los casos, puede terminarse con dos análisis; uno con el punto atípico y uno sin él.

Existen varios procedimientos estadísticos formales para detectar puntos atípicos (por ejemplo, ver Barnett y Lewis [8], John y Prescott [60] y Stefansky [107]). Puede hacerse una verificación aproximada de los puntos atípicos examinando los **residuales estandarizados**

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}} \quad (3-18)$$

Si los errores ε_{ij} son $N(0, \sigma^2)$, los residuales estandarizados deberán ser aproximadamente normales con media cero y varianza unitaria. Por lo tanto, cerca de 68% de los residuales estandarizados deberán estar incluidos dentro de los límites ± 1 , cerca de 95% de ellos deberán estar incluidos dentro de ± 2 y virtualmente todos ellos deberán estar incluidos dentro de ± 3 . Un residual mayor que 3 o 4 desviaciones estándar a partir de cero es un punto atípico potencial.

Para los datos de la resistencia a la tensión del ejemplo 3-1, la gráfica de probabilidad normal no produce indicio alguno de puntos atípicos. Además, el residual estandarizado mayor es

$$d_{13} = \frac{e_{13}}{\sqrt{MS_E}} = \frac{5.2}{\sqrt{8.06}} = \frac{5.2}{2.84} = 1.83$$

el cual no deberá ser motivo de preocupación.

3-4.2 Gráfica de los residuales en secuencia en el tiempo

La graficación de los residuales en el orden temporal de la recolección de los datos es útil para detectar **correlaciones** entre los residuales. Una tendencia a tener corridas de residuales positivos y negativos indica una correlación positiva. Esto implicaría que el **supuesto de independencia** de los errores ha sido violado. Se trata de un problema potencialmente serio y cuya solución es difícil, por lo que de ser posible es importante evitar el problema cuando se colecten los datos. La aleatorización adecuada del experimento es un paso importante para conseguir la independencia.

En ocasiones las habilidades del experimentador (o los sujetos) pueden cambiar conforme el experimento avanza, o el proceso bajo estudio puede “vagar sin rumbo” o volverse más errático. Esto producirá con frecuencia un cambio en la varianza del error con el tiempo. Esta condición suele llevar a una gráfica de los residuales contra el tiempo que muestra una dispersión mayor en uno de sus extremos que en el otro. Una varianza no constante es un problema potencialmente serio. En las secciones 3-4.3 y 3-4.4 se abundará sobre el tema.

En la tabla 3-6 se muestran los residuales y la secuencia en el tiempo de la recolección de los datos para el experimento de la resistencia a la tensión. En la figura 3-5 se presenta una gráfica de estos residua-

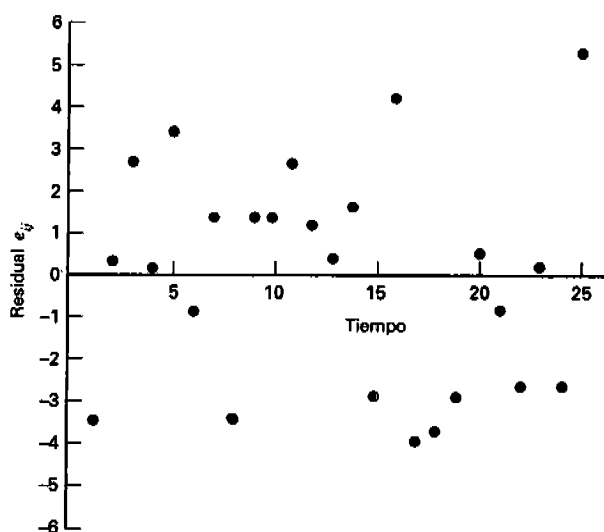


Figura 3-5 Gráfica de los residuales contra el tiempo.

les contra el tiempo. No hay razón para sospechar cualquier violación de los supuestos de independencia o de una varianza constante.

3-4.3 Gráfica de los residuales contra los valores ajustados

Si el modelo es correcto y se satisfacen los supuestos, los residuales deberán estar sin estructura; en particular, no deberán estar relacionados con ninguna otra variable, incluyendo la respuesta predicha. Una verificación simple es graficar los residuales contra los valores ajustados \hat{y}_{ij} . (Para el modelo de un experimento con un solo factor, recuerde que $\hat{y}_{ij} = \bar{y}_i$, el promedio del tratamiento i -ésimo.) Esta gráfica no deberá mostrar ningún patrón obvio. En la figura 3-6 se grafican los residuales contra los valores ajustados para los datos de la resistencia a la tensión del ejemplo 3-1. No es evidente ninguna estructura inusual.

Un defecto que sale a relucir en ocasiones en esta gráfica es una **varianza no constante**. En ocasiones la varianza de las observaciones se incrementa cuando la magnitud de la observación se incrementa. Éste sería el caso si el error o ruido de fondo del experimento fuera un porcentaje constante de la magnitud de la observación. (Esto ocurre comúnmente con muchos instrumentos de medición; el error es un porcentaje de la escala de medición.) Si éste fuera el caso, los residuales se harían mayores conforme y_{ij} se hiciera más grande, y la gráfica de los residuales contra y_{ij} se vería como un embudo o un megáfono con la boca hacia afuera. Una varianza no constante también surge en los casos en que los datos siguen una distribución no normal, sesgada, porque en las distribuciones sesgadas la varianza tiende a ser una función de la media.

Si se viola el supuesto de la homogeneidad de las varianzas, la prueba F sólo resulta afectada ligeramente en el modelo balanceado (el mismo tamaño de la muestra en todos los tratamientos) con efectos fijos. Sin embargo, en diseños no balanceados o en casos en que una de las varianzas es considerablemente más grande que las demás, el problema es más grave. Específicamente, si los niveles del factor que tienen las varianzas mayores corresponden también con los tamaños de las muestras más pequeños, el índice de error tipo I real es mayor que lo previsto (o los intervalos de confianza tienen niveles de confianza reales más bajos que los que fueron especificados). Recíprocamente, si los niveles del factor con las varianzas mayores tienen también los tamaños de las muestras mayores, los niveles de significación son mucho menores que lo anticipado (los niveles de confianza son más altos). Ésta es una buena razón para escoger **tamaños de las muestras iguales** siempre que sea posible. Para los modelos con efectos aleatorios, las

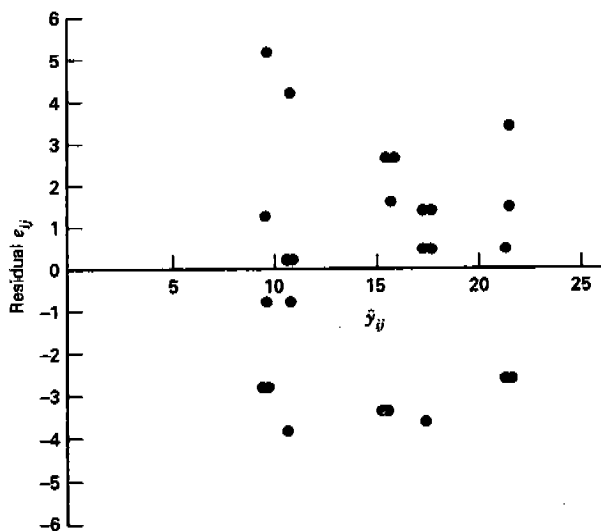


Figura 3-6 Gráfica de los residuales contra los valores ajustados.

varianzas del error diferentes pueden introducir alteraciones significativas en las inferencias sobre los componentes de la varianza, incluso cuando se usan diseños balanceados.

El enfoque usual para abordar el problema de una varianza no constante que ocurre por las razones expuestas antes consiste en aplicar una **transformación para estabilizar la varianza** para correr después el análisis de varianza en los datos transformados. En este enfoque, deberá tenerse presente que las conclusiones del análisis de varianza se aplican a las poblaciones *transformadas*.

Se han dedicado considerables esfuerzos de investigación a la selección de una transformación adecuada. Si los experimentadores conocen la distribución teórica de las observaciones, pueden hacer uso de esta información para elegir la transformación. Por ejemplo, si las observaciones siguen la distribución de Poisson, se usaría la **transformación de la raíz cuadrada** $y_{ij}^* = \sqrt{y_{ij}}$ o $y_{ij}^* = \sqrt{1 + y_{ij}}$. Si los datos siguen la distribución lognormal, la **transformación logarítmica** $y_{ij}^* = \log y_{ij}$ es adecuada. Para datos binomiales expresados como fracciones, la **transformación arcsen** $y_{ij}^* = \arcsen \sqrt{y_{ij}}$ es útil. Cuando no hay una transformación obvia, el experimentador realizará por lo general la búsqueda *empírica* de una transformación que iguale la varianza independientemente del valor de la media. Al final de esta sección se ofrecen algunos consejos al respecto. En los experimentos factoriales, los cuales se introducen en el capítulo 5, otro enfoque es seleccionar una transformación que minimice el cuadrado medio de las interacciones, siendo el resultado un experimento cuya interpretación es más sencilla. En el capítulo 14 se revisan con mayor detalle los métodos para seleccionar analíticamente la forma de la transformación. Las transformaciones que se hacen para la desigualdad de la varianza afectan también la forma de la distribución del error. En la mayoría de los casos, la transformación hace que la distribución del error esté más cerca de la distribución normal. Para más detalles sobre las transformaciones, referirse a Bartlett [7], Box y Cox [15], Dolby [38] y Draper y Hunter [39].

Pruebas estadísticas para la igualdad de la varianza

Aun cuando es frecuente el uso de las gráficas residuales para diagnosticar la desigualdad de la varianza, se han propuesto también varias pruebas estadísticas. Estas pruebas pueden considerarse como pruebas formales de las hipótesis

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$$

$$H_1: \text{el enunciado anterior no es verdadero para al menos una } \sigma_i^2$$

Un procedimiento muy utilizado es la **prueba de Bartlett**. El procedimiento incluye el cálculo de un estadístico cuya distribución de muestreo está aproximada muy de cerca por la distribución ji-cuadrada con $a - 1$ grados de libertad cuando las a muestras aleatorias son de poblaciones normales independientes. El estadístico de prueba es

$$\chi_0^2 = 2.3026 \frac{q}{c} \quad (3-19)$$

donde

$$q = (N - a) \log_{10} S_p^2 - \sum_{i=1}^a (n_i - 1) \log_{10} S_i^2$$

$$c = 1 + \frac{1}{3(a-1)} \left(\sum_{i=1}^a (n_i - 1)^{-1} - (N - a)^{-1} \right)$$

$$S_p^2 = \frac{\sum_{i=1}^a (n_i - 1) S_i^2}{N - a}$$

y S_i^2 es la varianza muestral de la población i -ésima.

La cantidad q es grande cuando la diferencia entre las varianzas muestrales S_i^2 es considerablemente grande, y es igual a cero cuando todas las S_i^2 son iguales. Por lo tanto, H_0 deberá rechazarse para los valores de χ_0^2 que sean muy grandes; es decir, se rechaza H_0 sólo cuando

$$\chi_0^2 > \chi_{\alpha, a-1}^2$$

donde $\chi_{\alpha, a-1}^2$ es el punto porcentual α superior de la distribución ji-cuadrada con $a - 1$ grados de libertad. También podría usarse el enfoque del valor P para tomar una decisión.

La prueba de Bartlett es muy sensible al supuesto de normalidad. Por consiguiente, cuando la validez de este supuesto está en duda, no deberá usarse la prueba de Bartlett.

EJEMPLO 3-4

Ya que el supuesto de normalidad no está en entredicho, la prueba de Bartlett puede aplicarse a los datos de la resistencia a la tensión del experimento del peso porcentual de algodón del ejemplo 3-1. Se calculan primero las varianzas muestrales de cada tratamiento y se encuentra que $S_1^2 = 11.2$, $S_2^2 = 9.8$, $S_3^2 = 4.3$, $S_4^2 = 6.8$ y $S_5^2 = 8.2$. Entonces

$$S_p^2 = \frac{4(11.2) + 4(9.8) + 4(4.3) + 4(6.8) + 4(8.2)}{20} = 8.06$$

$$q = 20 \log_{10} (8.06) - 4[\log_{10} 11.2 + \log_{10} 9.8 + \log_{10} 4.3 + \log_{10} 6.8 + \log_{10} 8.2] = 0.45$$

$$c = 1 + \frac{1}{3(4)} \left(\frac{5}{4} - \frac{1}{20} \right) = 1.10$$

y el estadístico de prueba es

$$\chi_0^2 = 2.3026 \frac{(0.45)}{(1.10)} = 0.93$$

Puesto que $\chi_{0.05, 4}^2 = 9.49$, no puede rechazarse la hipótesis nula y se concluye que las cinco varianzas son iguales. Se trata de la misma conclusión a la que se llegó al analizar la gráfica de los residuales contra los valores ajustados.

.....

Debido a que la prueba de Bartlett es sensible al supuesto de normalidad, puede haber situaciones en las que sería útil un procedimiento alternativo. Anderson y McLean [2] presentan una atinada revisión de algunas pruebas estadísticas de la igualdad de la varianza. La **prueba de Levene modificada** (ver Levene [72] y Conover, Johnson y Johnson [31]) es un procedimiento muy útil que es robusto en cuanto a las desviaciones de la normalidad. Para probar la hipótesis de que las varianzas son iguales en todos los tratamientos, la prueba de Levene modificada utiliza la desviación absoluta de las observaciones y_{ij} de cada tratamiento de la mediana de los tratamientos, por ejemplo $\bar{y}_{.j}$. Sea que estas desviaciones se denoten por

$$d_{ij} = |y_{ij} - \bar{y}_{.j}| \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n_i \end{cases}$$

La prueba de Levene modificada evalúa entonces si la media de estas desviaciones es igual o no para todos los tratamientos. Cuando las desviaciones medias son iguales, las varianzas de las observaciones de

Tabla 3-7 Datos de la descarga pico

Método de estimación	Observaciones						\bar{y}_i	\bar{y}_i	S_i
1	0.34	0.12	1.23	0.70	1.75	0.12	0.71	0.520	0.66
2	0.91	2.94	2.14	2.36	2.86	4.55	2.63	2.610	1.09
3	6.31	8.37	9.75	6.09	9.82	7.24	7.93	7.805	1.66
4	17.15	11.82	10.95	17.20	14.35	16.82	14.72	15.59	2.77

Método de estimación	Desviaciones d_{ij} para la prueba de Levene modificada								
1	0.18	0.40	0.71	0.18	1.23	0.40			
2	1.70	0.33	0.47	0.25	0.25	1.94			
3	1.495	0.565	1.945	1.715	2.015	0.565			
4	1.56	3.77	4.64	1.61	1.24	1.23			

todos los tratamientos serán iguales. El estadístico de prueba para la prueba de Levene es simplemente el estadístico F ANOVA usual para probar la igualdad de las medias que se aplica a las desviaciones absolutas.

EJEMPLO 3-5

Un ingeniero civil está interesado en determinar si cuatro métodos diferentes para estimar la frecuencia de las inundaciones producen estimaciones equivalentes de la descarga pico cuando se aplican a la misma cuenca. Cada procedimiento se usa seis veces en la cuenca, y los datos de las descargas resultantes (en pies cúbicos por segundo) se muestran en la parte superior de la tabla 3-7. El análisis de varianza de los datos, el cual se resume en la tabla 3-8, implica que hay una diferencia en las estimaciones de la descarga pico promedio obtenidas en los cuatro procedimientos. La gráfica de los residuales contra los valores ajustados, la cual se muestra en la figura 3-7, es preocupante porque la forma de embudo con la boca hacia afuera indica que no se satisface el supuesto de una varianza constante.

Se aplicará la prueba de Levene modificada a los datos de la descarga pico. La parte superior de la tabla 3-7 contiene las medianas de los tratamientos \bar{y}_i y la parte inferior contiene las desviaciones d_{ij} alrededor de las medianas. La prueba de Levene consiste en realizar un análisis de varianza estándar en las d_{ij} . El estadístico de prueba F que resulta en este caso es $F_0 = 4.55$, para el cual el valor P es $P = 0.0137$. Por lo tanto, la prueba de Levene rechaza la hipótesis nula de que las varianzas son iguales, coincidiendo en esencia con el diagnóstico que se hizo a partir del examen visual de la figura 3-7. Los datos de la descarga pico son un buen candidato para una transformación de datos.

Selección empírica de una transformación

Se señaló ya que si los experimentadores conocieran la relación entre la varianza de las observaciones y la media, podrían usar esta información como guía para la selección de la forma de la transformación. Se

Tabla 3-8 Análisis de varianza de los datos de la descarga pico

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Métodos	708.3471	3	236.1157	76.07	<0.001
Error	62.0811	20	3.1041		
Total	770.4282	23			

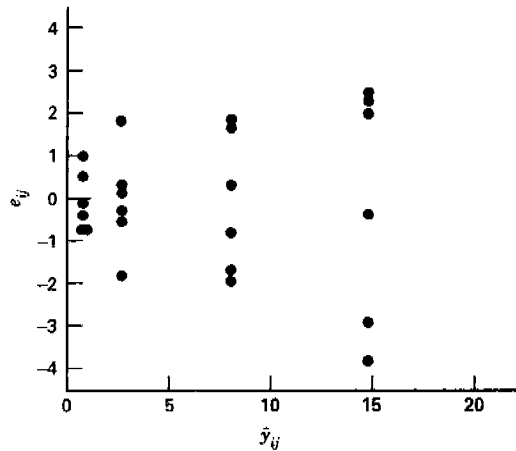


Figura 3-7 Gráfica de los residuales contra \hat{y}_{ij} para el ejemplo 3-5.

desarrolla ahora este punto y se presenta un método para seleccionar empíricamente la forma de la transformación requerida de los datos.

Sea $E(y) = \mu$ la media de y , y suponga que la desviación estándar de y es proporcional a una potencia de la media de y tal que

$$\sigma_y \propto \mu^\alpha$$

Quiere encontrarse una transformación de y que produzca una varianza constante. Suponga que la transformación es una potencia de los datos originales, por ejemplo

$$y^* = y^\lambda \quad (3-20)$$

Puede demostrarse entonces que

$$\sigma_{y^*} \propto \mu^{\lambda+\alpha-1} \quad (3-21)$$

Evidentemente, si se hace $\lambda = 1 - \alpha$, la varianza de los datos transformados y^* es constante.

En la tabla 3-9 se resumen varias de las transformaciones comunes analizadas anteriormente. Observe que $\lambda = 0$ implica la transformación logarítmica. Estas transformaciones se enlistan en el orden de **fuerza** creciente. Por fuerza de una transformación se entiende la cantidad de curvatura que induce. Una transformación suave aplicada a datos que se extienden en un rango estrecho tiene escaso efecto sobre el

Tabla 3-9 Transformaciones para estabilizar la varianza

Relación entre σ_y y μ	α	$\lambda = 1 - \alpha$	Transformación	Comentario
$\sigma_y \propto \text{constante}$	0	1	Sin transformación	
$\sigma_y \propto \mu^{1/2}$	1/2	1/2	Raíz cuadrada	Datos (números) de Poisson
$\sigma_y \propto \mu$	1	0	Log	
$\sigma_y \propto \mu^{3/2}$	3/2	-1/2	Raíz cuadrada recíproca	
$\sigma_y \propto \mu^2$	2	-1	Recíproco	

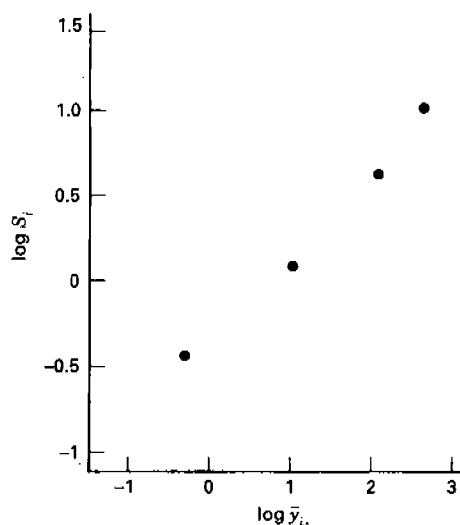


Figura 3-8 Gráfica de $\log S_i$ contra $\log \bar{y}_i$ para los datos de la descarga pico del ejemplo 3-5.

análisis, mientras que una transformación fuerte aplicada en un rango amplio puede tener resultados dramáticos. Con frecuencia las transformaciones tienen escaso efecto a menos que el cociente y_{\max}/y_{\min} sea mayor que 2 o 3.

En muchas situaciones de diseño experimental en las que se usan réplicas, α puede estimarse empíricamente a partir de los datos. Puesto que la combinación i -ésimo de los tratamientos $\sigma_{y_i} \propto \mu_i^\alpha = \theta \mu_i^\alpha$, donde θ es una constante de proporcionalidad, pueden tomarse logaritmos para obtener

$$\log \sigma_{y_i} = \log \theta + \alpha \log \mu_i \quad (3-22)$$

Por lo tanto, una gráfica de $\log \sigma_{y_i}$ contra $\log \mu_i$ sería una línea recta con pendiente α . Puesto que no se conocen σ_{y_i} y μ_i , pueden sustituirse estimaciones razonables de ellos en la ecuación 3-22 y usar la pendiente del ajuste de la línea recta resultante como estimación de α . De manera típica, se usaría la desviación estándar S_i y el promedio \bar{y}_i del tratamiento i -ésimo (o, en términos más generales, la combinación i -ésima de los tratamientos o conjunto de condiciones experimentales) para estimar σ_{y_i} y μ_i .

Para investigar la posibilidad de usar una transformación para estabilizar la varianza en los datos de la descarga pico del ejemplo 3-5, en la figura 3-8 se grafica $\log S_i$ contra $\log \bar{y}_i$. La pendiente de la recta que pasa por estos cuatro puntos está cerca de $1/2$ y, por la tabla 3-9, esto implica que la transformación de la raíz cuadrada puede ser apropiada. El análisis de varianza de los datos transformados $y^* = \sqrt{y}$ se presenta en la tabla 3-10, y en la figura 3-9 se muestra una gráfica de los residuales contra la respuesta predicha. Esta gráfica residual muestra una mejoría sensible en comparación con la figura 3-7, por lo que se conclu-

Tabla 3-10 Análisis de varianza de los datos transformados de la descarga pico, $y^* = \sqrt{y}$

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Métodos	32.6842	3	10.8947	76.99	<0.001
Error	2.6884	19	0.1415		
Total	35.3726	22			

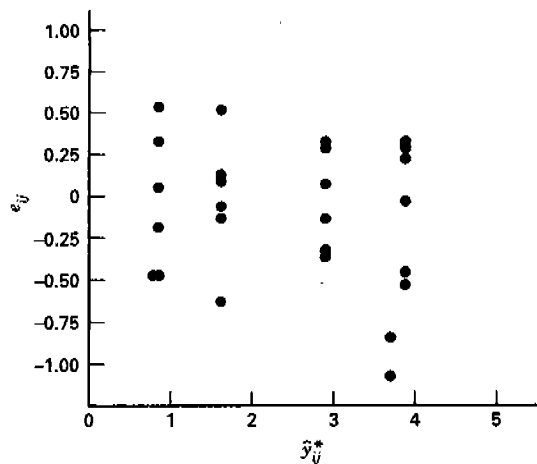


Figura 3-9 Gráfica de los residuales de los datos transformados contra \hat{y}_{ij}^* para los datos de la descarga pico del ejemplo 3-5.

ye que la transformación de la raíz cuadrada ha sido útil. Observe que en la tabla 3-10 se han reducido los grados de libertad del error en 1 para tomar en consideración el uso de los datos para estimar el parámetro de la transformación α .

En la práctica, muchos experimentadores seleccionan la forma de la transformación probando varias alternativas y observando el efecto de cada transformación en la gráfica de los residuales contra la respuesta predicha. Entonces se selecciona la transformación que produjo la gráfica residual más satisfactoria.

3-4.4 Gráficas de los residuales contra otras variables

Si se han recolectado datos de cualesquiera otras variables que posiblemente pudieran afectar la respuesta, los residuales deberán graficarse contra estas variables. Por ejemplo, en el experimento de la resistencia a la tensión del ejemplo 3-1, la resistencia puede ser afectada de manera significativa por el espesor de la fibra, por lo que los residuales deberán graficarse contra el espesor de la fibra. Si se usaron diferentes máquinas de prueba para recolectar los datos, los residuales deberán graficarse contra las máquinas. Los patrones en tales gráficas residuales implican que la variable afecta la respuesta. Esto sugiere que la variable debería controlarse con mayor atención en experimentos futuros o que debería incluirse en el análisis.

3-5 INTERPRETACIÓN PRÁCTICA DE LOS RESULTADOS

Después de realizar el experimento, llevar a cabo el análisis estadístico e investigar los supuestos fundamentales, el experimentador está listo para sacar conclusiones prácticas acerca del problema bajo estudio. Muchas veces esto es relativamente fácil, y ciertamente en los experimentos sencillos que se han considerado hasta este punto, esto podría hacerse de manera un tanto informal, tal vez mediante la inspección de las representaciones gráficas, como los diagramas de caja y el diagrama de dispersión de las figuras 3-1 y 3-2. Sin embargo, en algunos casos es necesario aplicar técnicas más formales. En esta sección se presentarán algunas de ellas.

3-5.1 Un modelo de regresión

Los factores que intervienen en un experimento pueden ser **cuantitativos** o **cualitativos**. Un factor cuantitativo es aquel cuyos niveles pueden asociarse con puntos en una escala numérica, como la temperatura, la presión o el tiempo. Los factores cualitativos, por otra parte, son aquellos cuyos niveles no pueden ordenarse por magnitud. Los operadores, los lotes de materia prima y los cambios de turno son factores cualitativos típicos, ya que no existe ninguna razón para ordenarlos bajo algún criterio numérico particular.

En lo que se refiere al diseño inicial y al análisis del experimento, ambos tipos de factores se tratan de manera idéntica. El experimentador está interesado en determinar las diferencias, en caso de haberlas, entre los niveles de los factores. Si el factor es cualitativo, como los operadores, no tiene sentido considerar la respuesta de una corrida subsecuente en un nivel intermedio del factor. Sin embargo, con un factor cuantitativo como el tiempo, el experimentador tiene interés por lo general en el rango completo de los valores usados, particularmente la respuesta de una corrida subsecuente con un nivel intermedio del factor. Es decir, si se usan los niveles 1.0, 2.0 y 3.0 en el experimento, tal vez quiera predecir la respuesta de 2.5 horas. Por lo tanto, el experimentador tiene con frecuencia interés en desarrollar una ecuación de interpolación para la variable de respuesta del experimento. Esta ecuación es un **modelo empírico** del proceso que se ha estudiado.

Al enfoque general para ajustar modelos empíricos se le llama **análisis de regresión**, el cual se analiza en detalle en el capítulo 10. Véase también el material suplementario del texto para este capítulo. En esta sección se ilustra brevemente la técnica utilizando los datos de la resistencia a la tensión del ejemplo 3-1.

En la figura 3-10 se presenta el diagrama de dispersión de la resistencia a la tensión y contra el peso porcentual del algodón x en la tela para el experimento del ejemplo 3-1. Los círculos huecos de la gráfica son las resistencias de tensión promedio para cada valor x del peso porcentual del algodón. A partir del examen del diagrama de dispersión, es evidente que la relación entre la resistencia a la tensión y el peso

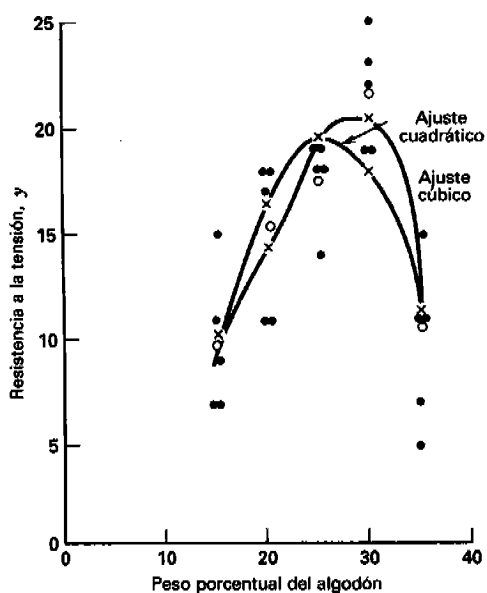


Figura 3-10 Diagrama de dispersión para los datos de la resistencia a la tensión del ejemplo 3-1.

porcentual del algodón no es lineal. Como una primera aproximación, podría intentarse ajustar un **modelo cuadrático** para los datos, por ejemplo

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

donde β_0, β_1 y β_2 son parámetros desconocidos que deberán estimarse y ε es un término del error aleatorio. El método que se usa con mayor frecuencia para estimar los parámetros en un modelo como éste es el método de mínimos cuadrados. Éste consiste en elegir estimaciones de las β tales que minimicen la suma de cuadrados de los errores (las ε). El ajuste de mínimos cuadrados en el ejemplo que se considera aquí es

$$\hat{y} = -39.9886 + 4.596x - 0.0886x^2$$

(Si el lector no está familiarizado con los métodos de regresión, vea el capítulo 10 y el material suplementario del texto para este capítulo.)

En la figura 3-10 se muestra este modelo cuadrático. No parece muy satisfactorio, ya que subestima de manera drástica las respuestas para $x = 30\%$ de algodón y sobrestima las respuestas para $x = 25\%$. Quizá pueda lograrse un mejoramiento agregando un término cúbico en x . El ajuste con el **modelo cúbico** resultante es

$$\hat{y} = 62.6114 - 9.0114x + 0.4814x^2 - 0.0076x^3$$

Este ajuste cúbico se ilustra también en la figura 3-10. El modelo cúbico parece mejor que el cuadrático porque proporciona un ajuste mejor para $x = 25$ y $x = 30\%$ de algodón.

En general, sería preferible hacer el ajuste con el polinomio de orden menor que describa adecuadamente el sistema o proceso. En este ejemplo, el polinomio cúbico parece un mejor ajuste que el cuadrático, por lo que la complejidad adicional del modelo cúbico se justifica. Sin embargo, seleccionar el orden del polinomio de aproximación no siempre es fácil, y es relativamente sencillo excederse en el ajuste, es decir, agregar polinomios de orden superior que no mejoran en realidad el ajuste pero que incrementan la complejidad del modelo y con frecuencia demeritan su utilidad como predictor o ecuación de interpolación.

En este ejemplo, el modelo empírico podría usarse para predecir la resistencia a la tensión media para los valores del peso porcentual del algodón dentro de la región de experimentación. En otros casos, el modelo empírico podría usarse para la **optimización del proceso**, es decir, para encontrar los niveles de las variables del diseño que dan como resultado los mejores valores de la respuesta. Más adelante se analizarán e ilustrarán en detalle estos problemas.

3-5.2 Comparaciones entre las medias de los tratamientos

Suponga que al hacer el análisis de varianza en el modelo con efectos fijos se rechaza la hipótesis nula. Por lo tanto, hay diferencias entre las medias de los tratamientos, pero no se especifica exactamente *cuáles* medias difieren. En ocasiones pueden ser de utilidad en esta situación las comparaciones y los análisis adicionales entre **grupos** de las medias de los tratamientos. La media del tratamiento i -ésimo se define como $\mu_i = \mu + \tau_i$ y μ_i se estima con \bar{y}_i . Las comparaciones entre las medias de los tratamientos se hacen ya sea en términos de los totales de los tratamientos $\{y_i\}$ o bien de los promedios de los tratamientos $\{\bar{y}_i\}$. Es común llamar a los procedimientos para hacer estas comparaciones **métodos de comparaciones múltiples**. En varias de las secciones siguientes se analizan los métodos para hacer comparaciones entre las medias de los tratamientos individuales o de grupos de estas medias.

3-5.3 Comparaciones gráficas de medias

Es muy sencillo desarrollar un procedimiento gráfico para la comparación de las medias después de un análisis de varianza. Suponga que el factor de interés tiene a niveles y que $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_a$ son los promedios de los tratamientos. Si se conoce σ , el promedio de cualquier tratamiento tendría una desviación estándar σ/\sqrt{n} . Por consiguiente, si todas las medias de los niveles del factor son idénticas, las medias muestrales observadas \bar{y}_i se comportarían como un conjunto de observaciones tomadas al azar de una distribución normal con media \bar{y} y desviación estándar σ/\sqrt{n} . Visualice una distribución normal con la capacidad de ser deslizada sobre un eje abajo del cual están graficadas $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_a$. Si todas las medias de los tratamientos son iguales, deberá haber una posición de esta distribución que haga evidente que los valores \bar{y}_i se sacaron de la misma distribución. Si no es éste el caso, los valores \bar{y}_i que no parecen haberse sacado de esta distribución se asocian con los niveles del factor que producen respuestas medias diferentes.

La única falla en esta lógica es que σ es desconocida. Sin embargo, puede sustituirse σ con $\sqrt{MS_E}$ del análisis de varianza y usar una distribución t con un factor de escalación $\sqrt{MS_E}/n$ en lugar de la distribución normal. En la figura 3-11 se muestra este arreglo para los datos de la resistencia a la tensión del ejemplo 3-1.

Para trazar la distribución t de la figura 3-11, simplemente se multiplica el valor de la abscisa t por el factor de escalación

$$\sqrt{MS_E}/n = \sqrt{8.06/5} = 1.27$$

y se grafica contra la ordenada de t en ese punto. Dada la gran similitud entre la distribución t y la normal, excepto porque la primera es un poco más plana cerca del centro y tiene colas más largas, por lo general este trazo se construye fácilmente a ojo. Si quiere obtenerse mayor precisión, en Box, Hunter y Hunter [18] se incluye una tabla de los valores de las abscisas t y las ordenadas correspondientes. La distribución puede tener un origen arbitrario, aun cuando por lo general es mejor uno que esté en la región de los valores \bar{y}_i que van a compararse. En la figura 3-11, el origen es 15 lb/pulg².

Visualice ahora el desplazamiento de la distribución t de la figura 3-11 sobre el eje horizontal y examine las cinco medias graficadas en la figura. Observe que no hay ninguna posición de la distribución tal que los cinco promedios puedan considerarse como observaciones típicas seleccionadas al azar de la distribución. Esto implica que las cinco medias no son iguales; por lo tanto, la figura es una representación gráfica de los resultados del análisis de varianza. La figura indica que 30% de algodón produce resistencias a la tensión mucho más altas que 20 o 25 por ciento de algodón (las cuales son aproximadamente iguales), y que 15 o 35 por ciento de algodón (las cuales son aproximadamente iguales) producirían resistencias a la tensión aún más bajas.

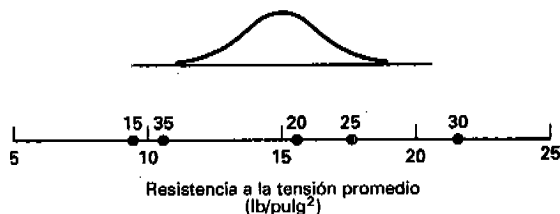


Figura 3-11 Promedio de la resistencia a la tensión del experimento del peso porcentual del algodón en relación con una distribución t con un factor de escalación $\sqrt{MS_E}/n = \sqrt{8.06/5} = 1.27$.

Este procedimiento simple es una técnica aproximada pero eficaz en muchos problemas de comparaciones múltiples. Sin embargo, existen métodos más formales. A continuación se presenta una breve revisión de algunos de estos procedimientos.

3-5.4 Contrastes

Muchos métodos de comparaciones múltiples utilizan el concepto de contraste. Considere el problema de la prueba de la fibra sintética del ejemplo 3-1. Puesto que se rechazó la hipótesis nula, se sabe que algunos pesos porcentuales del algodón producen resistencias a la tensión diferentes que otros, pero, ¿cuáles son los que causan en realidad esta diferencia? Al principio del experimento podría sospecharse que los niveles 4 y 5 del peso porcentual del algodón (30 y 35 por ciento) producen la misma resistencia a la tensión, lo cual implicaría que la hipótesis por probar sería

$$\begin{aligned}H_0: \mu_4 &= \mu_5 \\H_1: \mu_4 &\neq \mu_5\end{aligned}$$

o, de manera equivalente,

$$\begin{aligned}H_0: \mu_4 - \mu_5 &= 0 \\H_1: \mu_4 - \mu_5 &\neq 0\end{aligned}\tag{3-23}$$

Si desde el principio del experimento se hubiera sospechado que el *promedio* de los niveles más bajos del peso porcentual del algodón (1 y 2) no difería del *promedio* de los niveles más altos del peso porcentual del algodón (4 y 5), entonces la hipótesis habría sido

$$\begin{aligned}H_0: \mu_1 + \mu_2 &= \mu_4 + \mu_5 \\H_1: \mu_1 + \mu_2 &\neq \mu_4 + \mu_5\end{aligned}$$

o

$$\begin{aligned}H_0: \mu_1 + \mu_2 - \mu_4 - \mu_5 &= 0 \\H_1: \mu_1 + \mu_2 - \mu_4 - \mu_5 &\neq 0\end{aligned}\tag{3-24}$$

En general, un **contraste** es una combinación lineal de parámetros de la forma

$$\Gamma = \sum_{i=1}^a c_i \mu_i$$

donde las **constantes de los contrastes** c_1, c_2, \dots, c_a suman cero; es decir, $\sum_{i=1}^a c_i = 0$. Las dos hipótesis anteriores pueden expresarse en términos de contrastes:

$$\begin{aligned}H_0: \sum_{i=1}^a c_i \mu_i &= 0 \\H_1: \sum_{i=1}^a c_i \mu_i &\neq 0\end{aligned}\tag{3-25}$$

Las constantes de los contrastes para las hipótesis de la ecuación 3-23 son $c_1 = c_2 = c_3 = 0, c_4 = +1$ y $c_5 = -1$, mientras que para las hipótesis de la ecuación 3-24 son $c_1 = c_2 = +1, c_3 = 0$, y $c_4 = c_5 = -1$.

Las pruebas de hipótesis que incluyen contrastes pueden hacerse de dos maneras básicas. En el primer método se utiliza la prueba t . El contraste de interés se escribe en términos de los **totales de los tratamientos**, obteniéndose

$$C = \sum_{i=1}^a c_i y_i$$

La varianza de C es

$$V(C) = n\sigma^2 \sum_{i=1}^a c_i^2 \quad (3-26)$$

cuando los tamaños de las muestras de cada tratamiento son iguales. Si la hipótesis nula de la ecuación 3-25 es verdadera, el cociente

$$\frac{\sum_{i=1}^a c_i y_i}{\sqrt{n\sigma^2 \sum_{i=1}^a c_i^2}}$$

tiene la distribución $N(0, 1)$. Entonces se sustituiría la varianza desconocida σ^2 con su estimación, el error cuadrático medio MS_E , y se utilizaría el estadístico

$$t_0 = \frac{\sum_{i=1}^a c_i y_i}{\sqrt{nMS_E \sum_{i=1}^a c_i^2}} \quad (3-27)$$

para probar las hipótesis de la ecuación 3-25. La hipótesis nula se rechazaría si $|t_0|$ de la ecuación 3-27 excede $t_{\alpha/2, N-a}$.

En el segundo enfoque se utiliza la prueba F . Entonces, el cuadrado de una variable aleatoria t con ν grados de libertad es una variable aleatoria F con un grado de libertad en el numerador y ν grados de libertad en el denominador. Por lo tanto, puede obtenerse

$$F_0 = t_0^2 = \frac{\left(\sum_{i=1}^a c_i y_i \right)^2}{nMS_E \sum_{i=1}^a c_i^2} \quad (3-28)$$

como un estadístico F para probar la ecuación 3-25. La hipótesis nula se rechazaría si $F_0 > F_{\alpha, 1, N-a}$. Este estadístico de prueba de la ecuación 3-28 puede escribirse como

$$F_0 = \frac{MS_C}{MS_E} = \frac{SS_C / 1}{MS_E}$$

donde la suma de cuadrados de los contrastes con un solo grado de libertad es

$$SS_C = \frac{\left(\sum_{i=1}^a c_i y_i \right)^2}{n \sum_{i=1}^a c_i^2} \quad (3-29)$$

Intervalo de confianza para un contraste

En lugar de probar hipótesis acerca de un contraste, puede ser más útil construir un intervalo de confianza. Entonces el contraste suele expresarse en términos de los promedios de los tratamientos \bar{y}_i . Suponga que el contraste de interés es

$$\Gamma = \sum_{i=1}^a c_i \mu_i$$

Al sustituir las medias de los tratamientos con los promedios de los tratamientos se obtiene

$$C = \sum_{i=1}^a c_i \bar{y}_i$$

y

$$V(C) = \frac{\sigma^2}{n} \sum_{i=1}^a c_i^2$$

cuando los tamaños de las muestras son iguales. Si se usa MS_E para estimar σ^2 , el intervalo de confianza de $100(1 - \alpha)$ por ciento para el contraste $\sum_{i=1}^a c_i \mu_i$ es

$$\sum_{i=1}^a c_i \bar{y}_i - t_{\alpha/2, N-a} \sqrt{\frac{MS_E}{n} \sum_{i=1}^a c_i^2} \leq \sum_{i=1}^a c_i \mu_i \leq \sum_{i=1}^a c_i \bar{y}_i + t_{\alpha/2, N-a} \sqrt{\frac{MS_E}{n} \sum_{i=1}^a c_i^2} \quad (3-30)$$

Evidentemente, si este intervalo de confianza incluye al cero, no podría rechazarse la hipótesis nula en la ecuación 3-25.

Contraste estandarizado

Cuando hay interés en más de un contraste, con frecuencia es útil evaluarlos en la misma escala. Una forma de hacer esto es estandarizando el contraste para que su varianza sea σ^2 . Si el contraste $\sum_{i=1}^a c_i \mu_i$ se expresa en términos de los totales de los tratamientos como $\sum_{i=1}^a c_i y_i$, al dividirlo por $\sqrt{n \sum_{i=1}^a c_i^2}$ se obtendrá un contraste estandarizado con varianza σ^2 . Entonces el **contraste estandarizado** es en realidad

$$\sum_{i=1}^a c_i^* y_i$$

donde

$$c_i^* = \frac{c_i}{\sqrt{n \sum_{i=1}^a c_i^2}}$$

Tamaños de las muestras desiguales

Cuando los tamaños de las muestras de cada tratamiento son diferentes, se introducen modificaciones menores en los resultados anteriores. Primero, observe que la definición de un contraste requiere ahora que

$$\sum_{i=1}^a n_i c_i = 0$$

Otros cambios requeridos son directos. Por ejemplo, el estadístico t de la ecuación 3-27 queda como

$$t_0 = \frac{\sum_{i=1}^a c_i y_i}{\sqrt{MS_E \sum_{i=1}^a n_i c_i^2}}$$

y la suma de cuadrados de los contrastes de la ecuación 3-29 queda como

$$SS_c = \frac{\left(\sum_{i=1}^a c_i y_i \right)^2}{\sum_{i=1}^a n_i c_i^2}$$

3-5.5 Contrastes ortogonales

Un caso especial útil del procedimiento de la sección 3-5.4 es el de los **contrastes ortogonales**. Dos contrastes con coeficientes $\{c_i\}$ y $\{d_i\}$ son ortogonales si

$$\sum_{i=1}^a c_i d_i = 0$$

o, para un diseño no balanceado, si

$$\sum_{i=1}^a n_i c_i d_i = 0$$

Para a tratamientos, el conjunto de $a - 1$ contrastes ortogonales hace la partición de la suma de cuadrados debida a los tratamientos en $a - 1$ componentes independientes con un solo grado de libertad. Por lo tanto, las pruebas que se realizan en los contrastes ortogonales son independientes.

Existen varias maneras de elegir los coeficientes de los contrastes ortogonales para un conjunto de tratamientos. En general, algún elemento en la naturaleza del experimento deberá sugerir las comparaciones que son de interés. Por ejemplo, si hay $a = 3$ tratamientos, donde el tratamiento 1 es de control y donde los niveles del factor en los tratamientos 2 y 3 son de interés para el experimentador, los contrastes ortogonales apropiados podrían ser los siguientes:

Tratamiento	Coeficientes de los contrastes ortogonales	
1 (control)	-2	0
2 (nivel 1)	1	-1
3 (nivel 2)	1	1

Observe que el contraste 1 con $c_i = -2, 1, 1$ compara el efecto promedio del factor con el control, mientras que el contraste 2 con $d_i = 0, -1, 1$ compara los dos niveles del factor de interés.

En general, el método de contrastes (o de contrastes ortogonales) es útil para lo que se llama **comparaciones preplaneadas**. Es decir, los contrastes se especifican antes de llevar a cabo el experimento y de examinar los datos. La razón de esto es que, si las comparaciones se seleccionan después de examinar los

datos, la mayoría de los experimentadores construirían pruebas que corresponderían con las diferencias grandes observadas en las medias. Estas diferencias grandes podrían ser el resultado de la presencia de efectos reales o podrían ser el resultado del error aleatorio. Si los experimentadores se inclinan consistentemente a escoger las diferencias más grandes para hacer las comparaciones, inflarán el error tipo I de la prueba porque es probable que, en un porcentaje inusualmente elevado de las comparaciones seleccionadas, las diferencias observadas serán el resultado del error. Al examen de los datos para seleccionar las comparaciones de interés potencial se le llama con frecuencia **curioso o sondeo de datos**. El método de Scheffé para todas las comparaciones, el cual se comenta en la sección siguiente, permite el curioso o sondeo de datos.

EJEMPLO 3-6

Considere los datos del ejemplo 3-1. Hay cinco medias de los tratamientos y cuatro grados de libertad entre estos tratamientos. Suponga que antes de correr el experimento se especificó la siguiente serie de comparaciones entre las medias de los tratamientos (y sus contrastes asociados):

Hipótesis	Contraste
$H_0: \mu_4 = \mu_5$	$C_1 = -y_4 + y_5$
$H_0: \mu_1 + \mu_3 = \mu_4 + \mu_5$	$C_2 = y_1 + y_3 - y_4 - y_5$
$H_0: \mu_1 = \mu_3$	$C_3 = y_1 - y_3$
$H_0: 4\mu_2 = \mu_1 + \mu_3 + \mu_4 + \mu_5$	$C_4 = -y_1 + 4y_2 - y_3 - y_4 - y_5$

Observe que los coeficientes de los contrastes son ortogonales. Utilizando los datos de la tabla 3-4, se encuentra que los valores numéricos de los contrastes y de las sumas de cuadrados son los siguientes:

$$\begin{aligned}
 C_1 &= -1(108) + 1(54) = -54 & SS_{C_1} &= \frac{(-54)^2}{5(2)} = 291.60 \\
 C_2 &= +1(49) + 1(88) - 1(108) - 1(54) = -25 & SS_{C_2} &= \frac{(-25)^2}{5(4)} = 31.25 \\
 C_3 &= +1(49) - 1(88) = -39 & SS_{C_3} &= \frac{(-39)^2}{5(2)} = 152.10 \\
 C_4 &= -1(49) + 4(77) - 1(88) - 1(108) - 1(54) = 9 & SS_{C_4} &= \frac{(9)^2}{5(20)} = 0.81
 \end{aligned}$$

Tabla 3-11 Análisis de varianza de los datos de la resistencia a la tensión

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Peso porcentual del algodón	475.76	4	118.94	14.76	<0.001
contrastes ortogonales					
$C_1: \mu_4 = \mu_5$	(291.60)	1	291.60	36.18	<0.001
$C_2: \mu_1 + \mu_3 = \mu_4 + \mu_5$	(31.25)	1	31.25	3.88	0.06
$C_3: \mu_1 = \mu_3$	(152.10)	1	152.10	18.87	<0.001
$C_4: 4\mu_2 = \mu_1 + \mu_3 + \mu_4 + \mu_5$	(0.81)	1	0.81	0.10	0.76
Error	161.20	20	8.06		
Total	636.96	24			

Estas sumas de cuadrados de los contrastes hacen la partición completa de la suma de cuadrados de los tratamientos. Las pruebas de estos contrastes ortogonales se incorporan por lo general en el análisis de varianza, como se muestra en la tabla 3-11. Por los valores P se concluye que hay diferencias significativas entre los niveles 4 y 5 y 1 y 3 del peso porcentual del algodón, pero que el *promedio* de los niveles 1 y 3 no difiere del promedio de los niveles 4 y 5 con el nivel $\alpha = 0.05$, y que el nivel 2 no difiere del promedio de los otros cuatro niveles.

3-5.6 Método de Scheffé para comparar todos los contrastes

En muchas situaciones, los experimentadores pueden no conocer de antemano cuáles son los contrastes que quieren comparar, o pueden tener interés en más de $a - 1$ posibles comparaciones. En muchos experimentos de exploración, las comparaciones de interés sólo se descubren después del examen preliminar de los datos. Scheffé [98a] ha propuesto un método para comparar todos y cada uno de los contrastes posibles entre las medias de los tratamientos. En el método de Scheffé, el error tipo I es a lo sumo α para cualquiera de las comparaciones posibles.

Suponga que se ha determinado un conjunto de m contrastes

$$\Gamma_u = c_{1u}\mu_1 + c_{2u}\mu_2 + \cdots + c_{au}\mu_a \quad u = 1, 2, \dots, m \quad (3-31)$$

en las medias de los tratamientos de interés. El contraste correspondiente de los promedios de los tratamientos \bar{y}_i es

$$C_u = c_{1u}\bar{y}_1 + c_{2u}\bar{y}_2 + \cdots + c_{au}\bar{y}_a \quad u = 1, 2, \dots, m \quad (3-32)$$

y el error estándar de este contraste es

$$S_{C_u} = \sqrt{MS_E \sum_{i=1}^a (c_{iu}^2 / n_i)} \quad (3-33)$$

donde n_i es el número de observaciones en el tratamiento i -ésimo. Puede demostrarse que el valor crítico contra el que deberá compararse C_u es

$$S_{\alpha,u} = S_{C_u} \sqrt{(a-1)F_{\alpha,a-1,N-a}} \quad (3-34)$$

Para probar la hipótesis de que el contraste Γ_u difiere de manera significativa de cero, se compara C_u con el valor crítico. Si $|C_u| > S_{\alpha,u}$, se rechaza la hipótesis de que el contraste Γ_u es igual a cero.

El procedimiento de Scheffé puede usarse también para formar intervalos de confianza para todos los contrastes posibles entre las medias de los tratamientos. Los intervalos resultantes, por ejemplo $C_u - S_{\alpha,u} \leq \Gamma_u \leq C_u + S_{\alpha,u}$, son **intervalos de confianza simultáneos** por cuanto la probabilidad de que todos ellos sean verdaderos simultáneamente es al menos $1 - \alpha$.

Para ilustrar el procedimiento, considere los datos del ejemplo 3-1 y suponga que los contrastes de interés son

$$\Gamma_1 = \mu_1 + \mu_3 - \mu_4 - \mu_5$$

y

$$\Gamma_2 = \mu_1 - \mu_4$$

Los valores numéricos de estos contrastes son

$$\begin{aligned} C_1 &= \bar{y}_1 + \bar{y}_3 - \bar{y}_4 - \bar{y}_5 \\ &= 9.80 + 17.60 - 21.60 - 10.80 \\ &= 5.00 \end{aligned}$$

y

$$\begin{aligned}
 C_2 &= \bar{y}_1 - \bar{y}_4 \\
 &= 9.80 - 21.60 \\
 &= -11.80
 \end{aligned}$$

y los errores estándar se encuentran con la ecuación 3-33 como

$$S_{C_1} = \sqrt{MS_E \sum_{i=1}^5 (c_{i1}^2 / n_i)} = \sqrt{8.06(1+1+1+1)/5} = 2.54$$

y

$$S_{C_2} = \sqrt{MS_E \sum_{i=1}^5 (c_{i2}^2 / n_i)} = \sqrt{8.06(1+1)/5} = 1.80$$

Por la ecuación 3-34, los valores críticos de 1% son

$$S_{0.01,1} = S_{C_1} \sqrt{(a-1)F_{0.01,a-1,N-a}} = 2.54\sqrt{4(4.43)} = 10.69$$

y

$$S_{0.01,2} = S_{C_2} \sqrt{(a-1)F_{0.01,a-1,N-a}} = 1.80\sqrt{4(4.43)} = 7.58$$

Puesto que $|C_1| < S_{0.01,1}$, se concluye que el contraste $\Gamma_1 = \mu_1 + \mu_3 - \mu_4 - \mu_5$ es igual a cero; es decir, no existe evidencia sólida para concluir que las medias de los tratamientos 1 y 3 como grupo difieren de las medias de los tratamientos 4 y 5 como grupo. Sin embargo, como $|C_2| > S_{0.01,2}$, se concluye que el contraste $\Gamma_2 = \mu_1 - \mu_4$ no es igual a cero; es decir, las resistencias medias de los tratamientos 1 y 4 difieren significativamente.

En muchas situaciones prácticas, querrán compararse sólo **pares de medias**. Frecuentemente, es posible determinar cuáles son las medias que difieren probando las diferencias entre *todos* los pares de medias de los tratamientos. Por lo tanto, el interés se encuentra en los contrastes de la forma $\Gamma = \mu_i - \mu_j$ para toda $i \neq j$. Aun cuando el método de Scheffé podría aplicarse fácilmente a este problema, no es el procedimiento más sensible para tales comparaciones. Se pasa ahora a la consideración de los métodos diseñados específicamente para las comparaciones por pares entre todas las a medias poblacionales.

3-5.7 Comparación de pares de medias de tratamientos

Suponga que el interés se encuentra en comparar todos los pares de a medias de tratamientos y que las hipótesis nulas que quieren probarse son $H_0: \mu_i = \mu_j$ para toda $i \neq j$. A continuación se presentan cuatro métodos para hacer estas comparaciones.

Prueba de Tukey

Suponga que, después de un análisis de varianza en el que se ha rechazado la hipótesis nula de la igualdad de las medias de los tratamientos, quieren probarse todas las comparaciones de las medias por pares:

$$\begin{aligned}
 H_0: \mu_i &= \mu_j \\
 H_1: \mu_i &\neq \mu_j
 \end{aligned}$$

para toda $i \neq j$. Tukey [111d] propuso un procedimiento para probar hipótesis para las que el nivel de significación global es exactamente α cuando los tamaños de las muestras son iguales y es a lo sumo α cuando

los tamaños de las muestras no son iguales. Este procedimiento puede usarse también para contraer los intervalos de confianza para las diferencias en todos los pares de medias. Para estos intervalos, el nivel de confianza simultáneo es de $100(1 - \alpha)$ por ciento cuando los tamaños de las muestras son iguales y de al menos $100(1 - \alpha)$ por ciento cuando los tamaños de las muestras no son iguales. Se trata de un procedimiento excelente para curiosear sobre los datos cuando el interés se centra en pares de medias.

El procedimiento de Tukey hace uso de la distribución del **estadístico del rango studentizado**

$$q = \frac{\bar{y}_{\max} - \bar{y}_{\min}}{\sqrt{MS_E / n}}$$

donde \bar{y}_{\max} y \bar{y}_{\min} son las medias muestrales mayor y menor, respectivamente, sacadas de un grupo de p medias muestrales. La tabla VIII del apéndice contiene los valores de $q_\alpha(p, f)$, los puntos porcentuales α superiores de q , donde f es el número de grados de libertad asociados con MS_E . Para tamaños de las muestras iguales, la prueba de Tukey declara que dos medias son significativamente diferentes si el valor absoluto de sus diferencias muestrales excede

$$T_\alpha = q_\alpha(a, f) \sqrt{\frac{MS_E}{n}} \quad (3-35)$$

De manera equivalente, podría construirse una serie de intervalos de confianza de $100(1 - \alpha)$ por ciento para todos los pares de medias de la siguiente manera:

$$\bar{y}_i - \bar{y}_j - q_\alpha(a, f) \sqrt{\frac{MS_E}{n}} \leq \mu_i - \mu_j \leq \bar{y}_i - \bar{y}_j + q_\alpha(a, f) \sqrt{\frac{MS_E}{n}}, \quad i \neq j. \quad (3-36)$$

Cuando los tamaños de las muestras no son iguales, las ecuaciones 3-35 y 3-36 quedan como

$$T_\alpha = \frac{q_\alpha(a, f)}{\sqrt{2}} \sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (3-37)$$

y

$$\bar{y}_i - \bar{y}_j - \frac{q_\alpha(a, f)}{\sqrt{2}} \sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \leq \mu_i - \mu_j \leq \bar{y}_i - \bar{y}_j + \frac{q_\alpha(a, f)}{\sqrt{2}} \sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}, \quad i \neq j \quad (3-38)$$

respectivamente. A la versión para tamaños de las muestras diferentes se le llama en ocasiones el **procedimiento Tukey-Kramer**.

EJEMPLO 3-7

Para ilustrar la prueba de Tukey, se usan los datos del experimento del peso porcentual del algodón del ejemplo 3-1. Con $\alpha = 0.05$ y $f = 20$ grados de libertad para el error, en la tabla VIII del apéndice se obtiene $q_{0.05}(5, 20) = 4.23$. Por lo tanto, por la ecuación 3-35,

$$T_{0.05} = q_{0.05}(5, 20) \sqrt{\frac{MS_E}{n}} = 4.23 \sqrt{\frac{8.06}{5}} = 5.37$$

\bar{y}_1	\bar{y}_5	\bar{y}_2	\bar{y}_3	\bar{y}_4
9.8	10.8	15.4	17.6	21.6

Figura 3-12 Resultados de la prueba de Tukey.

Por lo tanto, cualquier par de promedios de los tratamientos que difieran en valor absoluto por más de 5.37 implicaría que el par correspondiente de medias poblacionales son significativamente diferentes. Los cinco promedios de los tratamientos son

$$\bar{y}_1 = 9.8 \quad \bar{y}_2 = 15.4 \quad \bar{y}_3 = 17.6 \quad \bar{y}_4 = 21.6 \quad \bar{y}_5 = 10.8$$

y las diferencias en los promedios son

$$\begin{aligned} \bar{y}_1 - \bar{y}_2 &= 9.8 - 15.4 = -5.6^* \\ \bar{y}_1 - \bar{y}_3 &= 9.8 - 17.6 = -7.8^* \\ \bar{y}_1 - \bar{y}_4 &= 9.8 - 21.6 = -11.8^* \\ \bar{y}_1 - \bar{y}_5 &= 9.8 - 10.8 = -1.0 \\ \bar{y}_2 - \bar{y}_3 &= 15.4 - 17.6 = -2.2 \\ \bar{y}_2 - \bar{y}_4 &= 15.4 - 21.6 = -6.2^* \\ \bar{y}_2 - \bar{y}_5 &= 15.4 - 10.8 = 4.6 \\ \bar{y}_3 - \bar{y}_4 &= 17.6 - 21.6 = -4.0 \\ \bar{y}_3 - \bar{y}_5 &= 17.6 - 10.8 = 6.8^* \\ \bar{y}_4 - \bar{y}_5 &= 21.6 - 10.8 = 10.8^* \end{aligned}$$

Los valores marcados con asterisco indican pares de medias que son significativamente diferentes. Suele ser útil trazar una gráfica, como la de la figura 3-12, donde se subraya a los pares de medias que no difieren significativamente. Esta gráfica da una indicación de que las medias de los tratamientos forman tres grupos: μ_1 y μ_5 , μ_2 y μ_3 , y μ_4 . Sin embargo, la pertenencia a estos grupos no es del todo clara.

.....

Cuando se utiliza cualquiera de los procedimientos para probar las medias por pares, ocasionalmente se encuentra que la prueba F global del análisis de varianza es significativa, pero la comparación de las medias por pares falla para revelar cualquier diferencia significativa. Esta situación ocurre porque la prueba F considera simultáneamente todos los contrastes posibles en los que intervienen las medias de los tratamientos, no sólo las comparaciones por pares. Es decir, en los datos a la mano, quizá no todos los contrastes significativos sean de la forma $\mu_i - \mu_j$.

Algunos paquetes de software de computadora presentan comparaciones por pares con intervalos de confianza. Para el procedimiento de Tukey, estos intervalos se calcularían con la ecuación 3-36 o la 3-38, dependiendo de si los tamaños de las muestras son iguales o no.

La deducción del intervalo de confianza de Tukey de la ecuación 3-36 para tamaños de las muestras iguales es directa. Para el estadístico del rango studentizado q se tiene

$$P\left(\frac{\max(\bar{y}_i - \mu_i) - \min(\bar{y}_i - \mu_i)}{\sqrt{MS_E/n}} \leq q_\alpha(a, f)\right) = 1 - \alpha$$

Si $\max(\bar{y}_i - \mu_i) - \min(\bar{y}_i - \mu_i)$ es menor o igual que $q_\alpha(a, f)\sqrt{MS_E/n}$, debe ser verdadero que $|(\bar{y}_i - \mu_i) - (\bar{y}_j - \mu_j)| \leq q_\alpha(a, f)\sqrt{MS_E/n}$ para cada par de medias. Por lo tanto,

$$P\left(-q_\alpha(a, f)\sqrt{\frac{MS_E}{n}} \leq \bar{y}_i - \bar{y}_j - (\mu_i - \mu_j) \leq q_\alpha(a, f)\sqrt{\frac{MS_E}{n}}\right) = 1 - \alpha$$

Al reordenar esta expresión para aislar $\mu_i - \mu_j$ entre las desigualdades se llegará al conjunto de intervalos de confianza simultáneos de $100(1 - \alpha)$ por ciento dado en la ecuación 3-38.

El método de la diferencia significativa mínima (LSD) de Fisher

En este procedimiento se utiliza el estadístico F para probar $H_0: \mu_i = \mu_j$

$$t_0 = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \quad (3-39)$$

Suponiendo una hipótesis alternativa de dos colas, los pares de medias μ_i y μ_j se declararían significativamente diferentes si $|\bar{y}_i - \bar{y}_j| > t_{\alpha/2, N-a} \sqrt{MS_E (1/n_i + 1/n_j)}$. A la cantidad

$$LSD = t_{\alpha/2, N-a} \sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (3-40)$$

se le llama **diferencia significativa mínima**. Si el diseño es balanceado, $n_1 = n_2 = \dots = n_a = n$, y

$$LSD = t_{\alpha/2, N-a} \sqrt{\frac{2MS_E}{n}} \quad (3-41)$$

Para usar el procedimiento LSD de Fisher, simplemente se compara la diferencia observada entre cada par de promedios con la LSD correspondiente. Si $|\bar{y}_i - \bar{y}_j| > LSD$, se concluye que las medias poblacionales μ_i y μ_j difieren.

EJEMPLO 3-8

Para ilustrar el procedimiento, si se usan los datos del experimento del ejemplo 3-1, la LSD con $\alpha = 0.05$ es

$$LSD = t_{.025, 20} \sqrt{\frac{2MS_E}{n}} = 2.086 \sqrt{\frac{2(8.06)}{5}} = 3.75$$

Por lo tanto, cualquier par de promedios de los tratamientos que difiera del valor absoluto por más de 3.75 implicaría que el par correspondiente de medias poblacionales es significativamente diferente. Las diferencias en los promedios son

$$\begin{aligned} \bar{y}_1 - \bar{y}_2 &= 9.8 - 15.4 = -5.6^* \\ \bar{y}_1 - \bar{y}_3 &= 9.8 - 17.6 = -7.8^* \\ \bar{y}_1 - \bar{y}_4 &= 9.8 - 21.6 = -11.8^* \\ \bar{y}_1 - \bar{y}_5 &= 9.8 - 10.8 = -1.0 \\ \bar{y}_2 - \bar{y}_3 &= 15.4 - 17.6 = -2.2 \\ \bar{y}_2 - \bar{y}_4 &= 15.4 - 21.6 = -6.2^* \\ \bar{y}_2 - \bar{y}_5 &= 15.4 - 10.8 = 4.6^* \\ \bar{y}_3 - \bar{y}_4 &= 17.6 - 21.6 = -4.0^* \\ \bar{y}_3 - \bar{y}_5 &= 17.6 - 10.8 = 6.8^* \\ \bar{y}_4 - \bar{y}_5 &= 21.6 - 10.8 = 10.8^* \end{aligned}$$

\bar{y}_1	\bar{y}_5	\bar{y}_2	\bar{y}_3	\bar{y}_4
9.8	10.8	15.4	17.6	21.6

Figura 3-13 Resultados del procedimiento LSD.

Los valores marcados con asterisco indican pares de medias que son significativamente diferentes. En la figura 3-13 se resumen los resultados. Evidentemente, los únicos pares de medias que no difieren significativamente son 1 y 5 y 2 y 3, y el tratamiento 4 produce una resistencia a la tensión significativamente mayor que los otros tratamientos.

.....

Observe que el riesgo global α puede inflarse de manera considerable al utilizar este método. Específicamente, cuando a se hace más grande, el error tipo I del experimento (el cociente del número de experimentos en los que se comete al menos un error tipo I y el número total de experimentos) se hace grande.

Prueba del rango múltiple de Duncan

Un procedimiento muy utilizado para comparar todos los pares de medias es la **prueba del rango múltiple** desarrollada por Duncan [41]. Para aplicar la prueba del rango múltiple de Duncan cuando los tamaños de las muestras son iguales, los a promedios de los tratamientos se arreglan en orden ascendente, y el error estándar de cada promedio se determina como

$$S_{\bar{y}_i} = \sqrt{\frac{MS_E}{n}} \quad (3-42)$$

Para tamaños de las muestras desiguales, se sustituye n en la ecuación 3-42 con la media armónica n_h del $\{n_i\}$, donde

$$n_h = \frac{a}{\sum_{i=1}^a (1/n_i)} \quad (3-43)$$

Observe que si $n_1 = n_2 = \dots = n_a$, $n_h = n$. En la tabla de Duncan de los rangos significativos (tabla VII del apéndice) se obtienen los valores $r_\alpha(p, f)$ para $p = 2, 3, \dots, a$, donde α es el nivel de significación y f es el número de grados de libertad del error. Estos rangos se convierten en un conjunto de $a - 1$ rangos mínimos de significación (por ejemplo, R_p) para $p = 2, 3, \dots, a$ calculando

$$R_p = r_\alpha(p, f) S_{\bar{y}_i} \quad \text{para } p = 2, 3, \dots, a \quad (3-44)$$

Entonces, se prueban las diferencias observadas entre las medias, empezando con la más grande contra la menor, la cual se compararía con el rango mínimo de significación R_a . Después se calcula la diferencia de la mayor y la segunda menor y se compara con el rango mínimo de significación R_{a-1} . Estas comparaciones se continúan hasta que todas las medias se han comparado con la media mayor. Por último, se calcula la diferencia entre la segunda media mayor y la menor y se compara con el rango mínimo de significación R_{a-1} . Este proceso se continúa hasta que se han considerado las diferencias entre todos los $a(a-1)/2$ pares de medias posibles. Si una diferencia observada es mayor que el rango de significación mínima correspondiente, se concluye que el par de medias en cuestión es significativamente diferente. Para evitar

contradicciones, ninguna de las diferencias entre un par de medias se considera significativa si las dos medias en cuestión se localizan entre otras dos medias que no difieren significativamente.

EJEMPLO 3-9

La prueba del rango múltiple de Duncan puede aplicarse al experimento del ejemplo 3-1. Recuerde que $MS_E = 8.06$, $N = 25$, $n = 5$, y hay 20 grados de libertad del error. Al arreglar los promedios de los tratamientos en orden ascendente, se tiene

$$\begin{aligned}\bar{y}_1 &= 9.8 \\ \bar{y}_5 &= 10.8 \\ \bar{y}_2 &= 15.4 \\ \bar{y}_3 &= 17.6 \\ \bar{y}_4 &= 21.6\end{aligned}$$

El error estándar de cada promedio es $S_{\bar{y}_i} = \sqrt{8.06/5} = 1.27$. En el conjunto de rangos significativos de la tabla VII del apéndice para 20 grados de libertad y $\alpha = 0.05$, se obtiene $r_{0.05}(2, 20) = 2.95$, $r_{0.05}(3, 20) = 3.10$, $r_{0.05}(4, 20) = 3.18$ y $r_{0.05}(5, 20) = 3.25$. Por lo tanto, los rangos de significación mínima son

$$\begin{aligned}R_2 &= r_{0.05}(2, 20)S_{\bar{y}_i} = (2.95)(1.27) = 3.75 \\ R_3 &= r_{0.05}(3, 20)S_{\bar{y}_i} = (3.10)(1.27) = 3.94 \\ R_4 &= r_{0.05}(4, 20)S_{\bar{y}_i} = (3.18)(1.27) = 4.04 \\ R_5 &= r_{0.05}(5, 20)S_{\bar{y}_i} = (3.25)(1.27) = 4.13\end{aligned}$$

Los resultados de las comparaciones serían

$$\begin{aligned}4 \text{ vs. } 1: 21.6 - 9.8 &= 11.8 > 4.13 \ (R_5) \\ 4 \text{ vs. } 5: 21.6 - 10.8 &= 10.8 > 4.04 \ (R_4) \\ 4 \text{ vs. } 2: 21.6 - 15.4 &= 6.2 > 3.94 \ (R_3) \\ 4 \text{ vs. } 3: 21.6 - 17.6 &= 4.0 > 3.75 \ (R_2) \\ 3 \text{ vs. } 1: 17.6 - 9.8 &= 7.8 > 4.04 \ (R_4) \\ 3 \text{ vs. } 5: 17.6 - 10.8 &= 6.8 > 3.95 \ (R_3) \\ 3 \text{ vs. } 2: 17.6 - 15.4 &= 2.2 < 3.75 \ (R_2) \\ 2 \text{ vs. } 1: 15.4 - 9.8 &= 5.6 > 3.94 \ (R_3) \\ 2 \text{ vs. } 5: 15.4 - 10.8 &= 4.6 > 3.75 \ (R_2) \\ 5 \text{ vs. } 1: 10.8 - 9.8 &= 1.0 < 3.75 \ (R_2)\end{aligned}$$

Por el análisis se observa que hay diferencias significativas entre todos los pares de medias con excepción de la 3 y la 2 y la 5 y la 1. En la figura 3-14 se muestra una gráfica en la que esas medias que no son significativamente diferentes aparecen subrayadas. Observe que en este ejemplo la prueba del rango múltiple de Duncan y el método LSD llevan a conclusiones idénticas.

.....

\bar{y}_1	\bar{y}_5	\bar{y}_2	\bar{y}_3	\bar{y}_4
9.8	10.8	15.4	17.6	21.6
<u>9.8</u>	<u>10.8</u>	<u>15.4</u>	<u>17.6</u>	21.6

Figura 3-14 Resultados de la prueba de rango múltiple de Duncan.

En la prueba del rango múltiple de Duncan se requiere una diferencia observada más grande para detectar pares significativamente diferentes de medias, cuando el número de medias incluidas en el grupo aumenta. De esta forma, en el ejemplo anterior $R_2 = 3.75$ (dos medias) mientras que $R_3 = 3.94$ (tres medias). Para dos medias, el valor crítico R_2 será exactamente igual al valor LSD de la prueba t . Los valores $r_\alpha(p, f)$ de la tabla VII del apéndice se eligen de tal modo que se obtenga un nivel de protección especificado. Es decir, cuando se comparan dos medias que están p pasos aparte, el **nivel de protección** es $(1 - \alpha)^{p-1}$, donde α es el nivel de significación especificado para dos medias *adyacentes*. Por lo tanto, el índice de error de reportar al menos una diferencia significativa incorrecta entre dos medias es $1 - (1 - \alpha)^{p-1}$, cuando el tamaño del grupo es p . Por ejemplo, si $\alpha = 0.05$, entonces $1 - (1 - 0.05)^1 = 0.05$ es el nivel de significación para comparar el par de medias adyacentes, $1 - (1 - 0.05)^2 = 0.10$ es el nivel de significación para medias que están un paso aparte, y así sucesivamente.

En general, si el nivel de protección es α , las pruebas de las medias tienen un nivel de significación que es mayor o igual que α . Por consiguiente, el procedimiento de Duncan tiene una gran potencia; es decir, es muy eficaz para detectar diferencias entre medias cuando existen diferencias reales. Por esta razón, la prueba del rango múltiple de Duncan es muy popular.

La prueba de Newman-Keuls

Esta prueba fue creada por Newman [90]. Debido a que un interés renovado en la prueba de Newman fue generado por Keuls [64], al procedimiento se le llama la prueba de Newman-Keuls. Operacionalmente, el procedimiento es similar a la prueba del rango múltiple de Duncan, salvo porque las diferencias críticas entre las medias se calculan en una forma un tanto diferente. Específicamente, se calcula una serie de valores críticos

$$K_p = q_\alpha(p, f) S_{\bar{y}_i}, \quad p = 2, 3, \dots, a \quad (3-45)$$

donde $q_\alpha(p, f)$ es el punto porcentual α superior del rango studentizado para grupos de medias de tamaño p y con f grados de libertad del error. Una vez que se calculan los valores K_p con la ecuación 3-45, los pares de medias extremos en los grupos de tamaño p se comparan con K_p exactamente igual que en la prueba del rango múltiple de Duncan.

¿Qué método de comparación por pares debe usarse?

Ciertamente, una pregunta lógica en este punto es qué método de comparación por pares debe usarse. Desafortunadamente, no hay una respuesta precisa para esta pregunta, y los especialistas en estadística están con frecuencia en desacuerdo en cuanto a la utilidad de los diferentes procedimientos. Carmer y Swanson [24] han realizado estudios de simulación Montecarlo con varios procedimientos de comparaciones múltiples, incluyendo algunos que no se han considerado aquí. Estos autores reportan que el método de la diferencia significativa mínima es una prueba muy eficaz para detectar diferencias reales en las medias si se aplica *sólo después* de que la prueba F en el análisis de varianza sea significativa en 5%. Reportan asimismo un buen desempeño en la detección de diferencias reales con la prueba del rango múltiple de Duncan. Esto no es motivo de sorpresa, ya que estos dos métodos son los más poderosos de los que se han comentado aquí. Sin embargo, estos métodos no incluyen el índice de error en el modo del experimento. Debido a que el método de Tukey efectúa un control sobre el índice de error global, muchos experimentadores prefieren su uso.

La prueba de Newman-Keuls es más conservadora que la prueba del rango múltiple de Duncan por cuanto a que el índice de error tipo I es menor. Específicamente, el error tipo I del experimento es α para todas las pruebas que incluyen el mismo número de medias. Por consiguiente, debido a que α es por lo general bajo, la potencia de la prueba de Newman-Keuls casi siempre es menor que la de la prueba del rango múltiple de Duncan. Para demostrar que el procedimiento de Newman-Keuls lleva a una prueba con menor potencia que la prueba del rango múltiple de Duncan, se observa por una comparación de las ta-

blas VII y VIII del apéndice que para $p > 2$ se tiene $q_\alpha(p, f) > r_\alpha(p, f)$. Es decir, es "más difícil" declarar que un par de medias es significativamente diferente al utilizar la prueba de Newman-Keuls que cuando se usa el procedimiento de Duncan. Esto se ilustra a continuación para el caso en que $\alpha = 0.01$, $a = 8$ y $f = 20$:

p	2	3	4	5	6	7	8
$r_{0.01}(p, 20)$	4.02	4.22	4.33	4.40	4.47	4.53	4.58
$q_{0.01}(p, 20)$	4.02	4.64	5.02	5.29	5.51	5.69	5.84

Como se señaló antes, existen otros procedimientos de comparaciones múltiples. Algunos artículos que describen estos métodos son los de Miller [78], O'Neill y Wetherill [91] y Nelson [89]. También se recomienda el libro de Miller [77].

3-5.8 Comparación de medias de tratamientos con un control

En muchos experimentos, uno de los tratamientos es un **control**, y el analista se interesa en comparar cada una de las medias de los $a - 1$ tratamientos restantes con el control. Por lo tanto, sólo es necesario hacer $a - 1$ comparaciones. Un procedimiento para hacer estas comparaciones ha sido desarrollado por Dunnett [42]. Suponga que el tratamiento a es el control y que quieren probarse las hipótesis

$$H_0: \mu_i = \mu_a$$

$$H_1: \mu_i \neq \mu_a$$

para $i = 1, 2, \dots, a - 1$. El procedimiento de Dunnett es una modificación de la prueba t común. Para cada hipótesis se calculan las diferencias observadas en las medias muestrales

$$|\bar{y}_i - \bar{y}_a| \quad i = 1, 2, \dots, a - 1$$

La hipótesis nula $H_0: \mu_i = \mu_a$ se rechaza utilizando un índice α de error tipo I si

$$|\bar{y}_i - \bar{y}_a| > d_\alpha(a - 1, f) \sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_a} \right)} \quad (3-46)$$

donde la constante $d_\alpha(a - 1, f)$ se da en la tabla IX del apéndice. (Pueden hacerse pruebas tanto de una como de dos colas.) Observe que α es el nivel de significación conjunto asociado con las $a - 1$ pruebas.

EJEMPLO 3-10

Para ilustrar la prueba de Dunnett, considere el experimento del ejemplo 3-1, asumiendo que el tratamiento 5 es el control. En este ejemplo, $a = 5$, $a - 1 = 4$, $f = 20$ y $n_i = n = 5$. Con el nivel de 5%, en la tabla IX del apéndice se encuentra que $d_{0.05}(4, 20) = 2.65$. Por lo tanto, la diferencia crítica es

$$d_{0.05}(4, 20) \sqrt{\frac{2MS_E}{n}} = 2.65 \sqrt{\frac{2(8.06)}{5}} = 4.76$$

(Observe que se trata de una simplificación de la ecuación 3-46 que resulta de un diseño balanceado.) Por lo tanto, cualquier media de los tratamientos que difiera del control por más de 4.76 se declararía significativamente diferente. Las diferencias observadas son

$$\begin{aligned} 1 \text{ vs. } 5: \bar{y}_1 - \bar{y}_5 &= 9.8 - 10.8 = -1.0 \\ 2 \text{ vs. } 5: \bar{y}_2 - \bar{y}_5 &= 15.4 - 10.8 = 4.6 \\ 3 \text{ vs. } 5: \bar{y}_3 - \bar{y}_5 &= 17.6 - 10.8 = 6.8 \\ 4 \text{ vs. } 5: \bar{y}_4 - \bar{y}_5 &= 21.6 - 10.8 = 10.8 \end{aligned}$$

Sólo las diferencias $\bar{y}_3 - \bar{y}_5$, $\bar{y}_4 - \bar{y}_5$ indican alguna diferencia significativa cuando se comparan con el control; por lo tanto, se concluye que $\mu_3 \neq \mu_5$ y $\mu_4 \neq \mu_5$.

Cuando se hace la comparación de los tratamientos con un control, una buena idea es usar más observaciones para el tratamiento de control (por ejemplo, n_a) que para los demás tratamientos (por ejemplo, n), suponiendo un número igual de observaciones para los $a - 1$ tratamientos restantes. El cociente n_a/n deberá elegirse de tal modo que sea aproximadamente igual a la raíz cuadrada del número total de tratamientos. Es decir, se elige $n_a/n = \sqrt{a}$.

3-6 MUESTRA DE SALIDA DE COMPUTADORA

Hay una gran cantidad de programas de computadora para apoyar el diseño experimental y la realización de análisis de varianza. En la figura 3-15 se muestra la salida de uno de estos programas, *Design-Expert*, utilizando los datos del experimento con un solo factor del ejemplo 3-1. La suma de cuadrados correspondiente al "Modelo" ("Model") es la $SS_{\text{Tratamientos}}$ usual de un diseño con un solo factor. Esa fuente se identifica adicionalmente como "A". Cuando hay más de un factor en el experimento, la suma de cuadrados ("Sum of Squares") del modelo se descompondrá en varias fuentes (A, B, etc.). Observe que el resumen del análisis de varianza de la parte superior de la salida de computadora contiene las sumas de cuadrados, los grados de libertad ("DF", degrees of freedom), los cuadrados medios ("Mean Square") y el estadístico de prueba F_0 ("F Value") acostumbrados. La columna "Prob > F" es el valor P (de hecho, el límite superior del valor P , ya que a las probabilidades menores que 0.0001 se les asigna el valor por omisión 0.0001).

Además del análisis de varianza básico, el programa presenta información adicional útil. La cantidad "R cuadrada" ("R-Squared") se define como

$$R^2 = \frac{SS_{\text{Modelo}}}{SS_{\text{Total}}} = \frac{475.76}{636.96} = 0.746923$$

y se interpreta en términos generales como la proporción de la variabilidad en los datos "explicada" por el modelo del análisis de varianza. Por lo tanto, en los datos para probar la resistencia de la fibra sintética, el factor "peso porcentual del algodón" explica cerca de 74.69% de la variabilidad en la resistencia a la tensión. Evidentemente, debe tenerse $0 \leq R^2 \leq 1$, siendo más deseables los valores más grandes. En la salida se presentan también otros estadísticos en R^2 . R^2 "ajustada" ("Adj R-Squared") es una variante del estadístico R^2 común que refleja el número de factores presentes en el modelo. Puede ser un estadístico útil en experimentos más complejos en los que intervienen varios factores en el diseño, cuando quiere evaluarse el impacto de aumentar o disminuir el número de términos del modelo. "Desviación estándar" ("Std. Dev.") es la raíz cuadrada del cuadrado medio del error, $\sqrt{8.060} = 2.839$, y "C.V." es el coeficiente de variación, definido como $(\sqrt{MS_E} / \bar{y})100$. El coeficiente de variación mide la variabilidad no explicada o residual de los datos como un porcentaje de la media ("Mean") de la variable de respuesta. "PRESS" son las siglas de *Prediction Error Sum of Squares* (suma de cuadrados del error de predicción) y es una medida de la adecuación con que es posible que el modelo del experimento predecirá las respuestas en un nuevo experimento. Son deseables valores pequeños de PRESS. Alternativamente, puede calcularse una R^2 para predicciones con base en PRESS (más adelante se indicará cómo hacer esto). Esta R^2_{Pred} ("Pred R-Squared") para el problema tratado aquí es 0.6046, el cual no es irrazonable, considerando que el modelo explica cerca de 75% de la variabilidad del experimento en curso. El estadístico "Predicción adecuada" ("Adeq Precision") se calcula dividiendo la diferencia entre la respuesta predicha máxima y la respuesta predicha mínima por la desviación estándar promedio de todas las respuestas predichas. Son deseables valores grandes de esta cantidad, y los valores que exceden cuatro indican por lo general que el modelo tendrá un desempeño razonable en la predicción.

Utilice el mouse para posicionarse en una celda y su definición.

Response: Strength in psi
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Model	475.76	4	118.94	14.76	<0.0001	significativo
Residual	475.76	4	118.94	14.76	<0.0001	
Lack of Fit	161.20	20	8.06			
Pure Error	0.000	0				
Cor Total	161.20	20	8.06			
	636.96	24				

El valor F del Modelo de 14.76 implica que el modelo es significativo. Sólo hay una probabilidad de 0.01% de que un "Valor F del Modelo" de esta magnitud pudiera ocurrir debido a ruido.

Los valores de "Prob > F" menores que 0.0500 indican que los términos del modelo son significativos.

En este caso A son términos significativos del modelo.

Los valores mayores que 0.1000 indican que los términos del modelo no son significativos.

Si hay muchos términos del modelo no significativos (sin contar los que se necesitan para apoyar la jerarquización), la reducción del modelo puede mejorarlo.

Std. Dev.	2.84	R-Squared	0.7469
Mean	15.04	Adj R-Squared	0.6963
C.V.	18.88	Pred R-Squared	0.6046
PRESS	251.88	Adeq Precision	9.294

La "R cuadrada predicha" de 0.6046 concuerda razonablemente con la "R cuadrada ajustada" de 0.6963. Una diferencia mayor que 0.20 entre la "R cuadrada predicha" y la "R cuadrada ajustada" indica un posible problema con el modelo y/o los datos.

"Precisión adecuada" mide la relación de la señal a ruido. Es deseable una relación mayor que 4. La relación de 9.294 indica una señal adecuada para usar este modelo para navegar el espacio del diseño.

Treatment Means (Adjusted, if Necessary)

	Estimated Mean	Standard Error
1-15	9.80	1.27
2-20	15.40	1.27
3-25	17.60	1.27
4-30	21.60	1.27
5-35	10.80	1.27

Treatment	Mean Difference	DF	Standard Error	t for H ₀ Coeff=0	Prob > t
1 vs 2	-5.60	1	1.80	-3.12	0.0054
1 vs 3	-7.80	1	1.80	-4.34	0.0003
1 vs 4	-11.80	1	1.80	-6.57	<0.0001
1 vs 5	-1.00	1	1.80	-0.56	0.5838
2 vs 3	-2.20	1	1.80	-1.23	0.2347
2 vs 4	-6.20	1	1.80	-3.45	0.0025
2 vs 5	4.60	1	1.80	2.56	0.0186
3 vs 4	-4.00	1	1.80	-2.23	0.0375
3 vs 5	6.80	1	1.80	3.79	0.0012
4 vs 5	10.80	1	1.80	6.01	<0.0001

Los valores de "Prob > |t|" menores que 0.0500 indican que la diferencia en las medias de los dos tratamientos es significativa.

Los valores de "Prob > |t|" mayores que 0.1000 indican que la diferencia en las medias de los dos tratamientos no es significativa.

Figura 3-15 Salida de computadora de *Design-Expert* para el ejemplo 3-1.

Diagnostic Case Statistics							
Standard Order	Actual Value	Predicted Value	Residual	Leverage	Student Residual	Cook's Distance	Outlier t
1	7.00	9.80	-2.80	0.200	-1.103	0.061	-1.109
2	7.00	9.80	-2.80	0.200	-1.103	0.061	-1.109
3	15.00	9.80	5.20	0.200	2.048	0.210	2.245
4	11.00	9.80	1.20	0.200	0.473	0.011	0.463
5	9.00	9.80	-0.80	0.200	-0.315	0.005	-0.308
6	12.00	15.40	-3.40	0.200	-1.339	0.090	-1.368
7	17.00	15.40	1.60	0.200	0.630	0.020	0.620
8	12.00	15.40	-3.40	0.200	-1.339	0.090	-1.368
9	18.00	15.40	2.60	0.200	1.024	0.052	1.025
10	18.00	15.40	2.60	0.200	1.024	0.052	1.025
11	14.00	17.60	-3.60	0.200	-1.418	0.100	-1.457
12	18.00	17.60	0.40	0.200	0.158	0.001	0.154
13	18.00	17.60	0.40	0.200	0.158	0.001	0.154
14	19.00	17.60	1.40	0.200	0.551	0.015	0.542
15	19.00	17.60	1.40	0.200	0.551	0.015	0.542
16	19.00	21.60	-2.60	0.200	-1.024	0.052	-1.025
17	25.00	21.60	3.40	0.200	1.339	0.090	1.368
18	22.00	21.60	0.40	0.200	0.158	0.001	0.154
19	19.00	21.60	-2.60	0.200	-1.024	0.052	-1.025
20	23.00	21.60	1.40	0.200	0.551	0.015	0.542
21	7.00	10.80	-3.80	0.200	-1.496	0.112	-1.548
22	10.00	10.80	-0.80	0.200	-0.315	0.005	-0.308
23	11.00	10.80	0.20	0.200	0.079	0.000	0.077
24	15.00	10.80	4.20	0.200	1.654	0.137	1.735
25	11.00	10.80	0.20	0.200	0.079	0.000	0.077

Proceder con las gráficas de diagnóstico (el icono siguiente en progresión). Asegurarse de examinar:

- 1) La gráfica de probabilidad normal de los residuales studentizados para verificar la normalidad de los residuales.
- 2) Los residuales studentizados contra los valores predichos para verificar la constante del error.
- 3) Los puntos atípicos t contra el orden de las corridas para buscar puntos atípicos, es decir, valores influyentes o importantes
- 4) La gráfica de Box-Cox para las transformaciones de potencia.

Si todos los estadísticos del modelo y las gráficas de diagnóstico están correctos, finalizar con el icono Model Graphs (Gráficas del Modelo).

Figura 3-15 (Continuación.)

Se hace la estimación de las medias ("Estimated Mean") de los tratamientos y se muestra el error estándar ("Standard Error") (o desviación estándar muestral de la media de cada tratamiento, $\sqrt{MS_E/n}$). Las diferencias entre pares de medias ("Mean Difference") de los tratamientos se investigan utilizando el método LSD de Fisher descrito en la sección 3-5.7.

El programa de computadora también calcula y despliega los residuales, según se definen en la ecuación 3-16. El programa producirá también todas las gráficas de los residuales que se comentaron en la sección 3-4. En la salida se muestran asimismo varios diagnósticos residuales más. Algunos de ellos se revisarán más adelante.

Por último, observe que el programa de computadora incluye también algunas guías para hacer la interpretación. Esta información "aconsejable" es muy común en muchos paquetes de estadística para computadoras personales. Al leer estas guías, recuerde que están escritas en términos muy generales, y quizá no se ajusten exactamente a los requerimientos de redacción del reporte de un experimentador particular. Esta salida aconsejable puede ser eliminada por el usuario.

3-7 DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA

En cualquier problema de diseño experimental, una decisión crítica es la elección del tamaño de la muestra; es decir, determinar el número de réplicas que deben correrse. En general, si el experimentador tiene interés en detectar efectos pequeños, se necesitan más réplicas que cuando el experimentador se interesa en detectar efectos grandes. En esta sección se analizan varios enfoques para determinar el tamaño de la muestra. Aun cuando la revisión se centra en un diseño con un solo factor, la mayoría de los métodos pueden usarse en situaciones experimentales más complejas.

3-7.1 Curvas de operación característica

Recuerde que una **curva de operación característica** es una gráfica de la probabilidad del error tipo II de una prueba estadística para un tamaño de la muestra particular contra un parámetro que refleja la medida en que la hipótesis nula es falsa. El experimentador puede usar estas curvas como guía en la selección del número de réplicas para que el diseño sea sensible a diferencias potenciales importantes en los tratamientos.

Se considera la probabilidad del error tipo II del modelo con efectos fijos para el caso en que se usa el mismo tamaño de las muestras en cada tratamiento, por ejemplo

$$\begin{aligned}\beta &= 1 - P\{\text{Rechazar } H_0 \mid H_0 \text{ es falsa}\} \\ &= 1 - P\{F_0 > F_{\alpha, a-1, N-a} \mid H_0 \text{ es falsa}\}\end{aligned}\quad (3-47)$$

Para evaluar el enunciado de probabilidad de la ecuación 3-47, es necesario conocer cuál es la distribución del estadístico de prueba F_0 si la hipótesis nula es falsa. Puede demostrarse que, si H_0 es falsa, el estadístico $F_0 = MS_{\text{Tratamientos}}/MS_E$ se distribuye como una variable aleatoria F **no central** con $a - 1$ y $N - a$ grados de libertad y parámetro de no centralidad δ . Si $\delta = 0$, la distribución F no central se convierte en la distribución F (central) común.

Las curvas de operación característica que se presentan en la parte V del apéndice se usan para evaluar el enunciado de probabilidad de la ecuación 3-47. En estas curvas se grafica la probabilidad del error tipo II (β) contra un parámetro Φ , donde

$$\Phi^2 = \frac{n \sum_{i=1}^a \tau_i^2}{a\sigma^2} \quad (3-48)$$

La cantidad Φ^2 está relacionada con el parámetro de no centralidad δ . Se cuenta con curvas para $\alpha = 0.05$ y $\alpha = 0.01$ y un rango de grados de libertad para el numerador y el denominador.

Al usar las curvas de operación característica, el experimentador debe especificar el parámetro Φ . Con frecuencia es difícil hacer esto en la práctica. Una manera de determinar Φ es elegir los valores reales de las medias de los tratamientos para los que querría rechazarse la hipótesis nula con una alta probabilidad. Por lo tanto, si $\mu_1, \mu_2, \dots, \mu_a$ son las medias de los tratamientos especificadas, la τ_i de la ecuación 3-48 se encuentra como $\tau_i = \mu_i - \bar{\mu}$, donde $\bar{\mu} = (1/a) \sum_{i=1}^a \mu_i$ es el promedio de las medias de los tratamientos individuales. Se requiere asimismo una estimación de σ^2 . En ocasiones se cuenta con este valor por experiencia previa, un experimento anterior o una prueba preliminar (como se sugirió en el capítulo 1), o por una estimación discrecional. Cuando no se tiene la seguridad acerca del valor de σ^2 , los tamaños de las muestras podrían determinarse para un rango de valores posibles de σ^2 , a fin de estudiar el efecto de este parámetro sobre el tamaño de la muestra requerido, antes de hacer la elección final.

EJEMPLO 3-11

Considere el experimento de la resistencia a la tensión descrito en el ejemplo 3-1. Suponga que el experimentador está interesado en rechazar la hipótesis nula con una probabilidad de al menos 0.90 si las medias de los cinco tratamientos son

$$\mu_1=11 \quad \mu_2=12 \quad \mu_3=15 \quad \mu_4=18 \quad \text{y} \quad \mu_5=19$$

Planea utilizar $\alpha = 0.01$. En este caso, puesto que $\sum_{i=1}^5 \mu_i = 75$, se tiene $\bar{\mu} = (1/5)75 = 15$ y

$$\tau_1 = \mu_1 - \bar{\mu} = 11 - 15 = -4$$

$$\tau_2 = \mu_2 - \bar{\mu} = 12 - 15 = -3$$

$$\tau_3 = \mu_3 - \bar{\mu} = 15 - 15 = 0$$

$$\tau_4 = \mu_4 - \bar{\mu} = 18 - 15 = 3$$

$$\tau_5 = \mu_5 - \bar{\mu} = 19 - 15 = 4$$

Por lo tanto, $\sum_{i=1}^5 \tau_i^2 = 50$. Suponga que el experimentador piensa que la desviación estándar de la resistencia a la tensión con cualquier nivel particular del peso porcentual del algodón no será mayor que $\sigma = 3$ psi. Entonces, al utilizar la ecuación 3-48, se tiene

$$\Phi^2 = \frac{n \sum_{i=1}^5 \tau_i^2}{a\sigma^2} = \frac{n(50)}{5(3)^2} = 1.11n$$

Se usa la curva de operación característica para $a - 1 = 5 - 1 = 4$ con $N - a = a(n - 1) = 5(n - 1)$ grados de libertad del error y $\alpha = 0.01$ (ver la parte V del apéndice). Como primera conjetura para el tamaño de la muestra requerido, se prueba con $n = 4$ réplicas. Esto produce $\Phi^2 = 1.11(4) = 4.44$, $\Phi = 2.11$ y $5(3) = 15$ grados de libertad del error. Por consiguiente, en la parte V se encuentra que $\beta = 0.30$. Por lo tanto, la potencia de la prueba es aproximadamente $1 - \beta = 1 - 0.30 = 0.70$, que es menor que el 0.90 requerido, por lo que se concluye que $n = 4$ réplicas no son suficientes. Procediendo de manera similar, puede construirse la siguiente tabla:

n	Φ^2	Φ	$a(n - 1)$	β	Potencia ($1 - \beta$)
4	4.44	2.11	15	0.30	0.70
5	5.55	2.36	20	0.15	0.85
6	6.66	2.58	25	0.04	0.96

Por lo tanto, deben realizarse al menos $n = 6$ réplicas para obtener una prueba con la potencia requerida.

El único problema con este enfoque para usar las curvas de operación característica es que por lo general es difícil seleccionar el conjunto de las medias de los tratamientos en el que se basará la decisión del tamaño de la muestra. Un enfoque alternativo es seleccionar un tamaño de la muestra tal que si la diferencia entre las medias de dos tratamientos cualesquiera excede un valor especificado, la hipótesis nula deberá rechazarse. Si la diferencia entre las medias de dos tratamientos cualesquiera es tan grande como D , puede demostrarse que el valor mínimo de Φ^2 es

$$\Phi^2 = \frac{nD^2}{2a\sigma^2} \quad (3-49)$$

Puesto que éste es un valor mínimo de Φ^2 , el tamaño de la muestra correspondiente que se obtiene de la curva de operación característica es un valor conservador; es decir, proporciona una potencia al menos tan grande como la que especificó el experimentador.

Para ilustrar este enfoque, suponga que en el experimento de la resistencia a la tensión del ejemplo 3-1, el experimentador quisiera rechazar la hipótesis nula con una probabilidad de al menos 0.90 si las medias de dos tratamientos cualesquiera difieren hasta en 10 psi. Entonces, suponiendo que $\sigma = 3$ psi, se encuentra que el valor mínimo de Φ^2 es

$$\Phi^2 = \frac{n(10)^2}{2(5)(3^2)} = 1.11n$$

y, por el análisis del ejemplo 3-11, se concluye que se necesitan $n = 6$ réplicas para obtener la sensibilidad deseada cuando $\alpha = 0.01$.

3-7.2 Especificación de un incremento de la desviación estándar

Este enfoque es útil en ocasiones para elegir el tamaño de la muestra. Si las medias de los tratamientos no difieren, la desviación estándar de una observación elegida al azar es σ . Sin embargo, si las medias de los tratamientos son diferentes, la desviación estándar de una observación elegida al azar es

$$\sqrt{\sigma^2 + \left(\sum_{i=1}^a \tau_i^2 / a \right)}$$

Si se escoge un porcentaje P para el incremento de la desviación estándar de una observación, más allá del cual quiera rechazarse la hipótesis de que las medias de todos los tratamientos son iguales, esto es equivalente a escoger

$$\frac{\sqrt{\sigma^2 + \left(\sum_{i=1}^a \tau_i^2 / a \right)}}{\sigma} = 1 + 0.01P \quad (P = \text{por ciento})$$

o

$$\frac{\sqrt{\left(\sum_{i=1}^a \tau_i^2 / a \right)}}{\sigma} = \sqrt{(1 + 0.01P)^2 - 1}$$

de donde

$$\Phi = \frac{\sqrt{\sum_{i=1}^a \tau_i^2 / a}}{\sigma / \sqrt{n}} = \sqrt{(1 + 0.01P)^2 - 1}(\sqrt{n}) \quad (3-50)$$

Por lo tanto, para un valor especificado de P , Φ puede calcularse con la ecuación 3-50 y después usar las curvas de operación característica de la parte V del apéndice para determinar el tamaño de la muestra requerido.

Por ejemplo, en el experimento de la resistencia a la tensión del ejemplo 3-1, suponga que se desea detectar un incremento de la desviación estándar de 20% con una probabilidad de al menos 0.90 y $\alpha = 0.05$. Entonces

$$\Phi = \sqrt{(1.2)^2 - 1}(\sqrt{n}) = 0.66\sqrt{n}$$

La referencia a las curvas de operación característica indica que se necesita $n = 9$ para obtener la sensibilidad deseada.

3-7.3 Método para estimar el intervalo de confianza

En este enfoque se supone que el experimentador quiere expresar los resultados finales en términos de intervalos de confianza y que está dispuesto a especificar por anticipado cuál es el ancho que desea para estos intervalos de confianza. Por ejemplo, suponga que en el experimento de la resistencia a la tensión del ejemplo 3-1 se quiere que un intervalo de confianza de 95% para la diferencia en la resistencia a la tensión media de dos pesos porcentuales del algodón cualesquiera sea ± 5 psi y una estimación previa de σ es 3. Entonces, al utilizar la ecuación 3-13, se encuentra que la precisión del intervalo de confianza es

$$\pm t_{\alpha/2, N-a} \sqrt{\frac{2MS_E}{n}}$$

Suponga que se prueba con $n = 5$ réplicas. Entonces, al usar $\sigma^2 = 3^2 = 9$ como una estimación de MS_E , la precisión del intervalo de confianza es

$$\pm 2.086 \sqrt{\frac{2(9)}{5}} = \pm 3.96$$

que es más preciso que el requerimiento. Al probar con $n = 4$ se obtiene

$$\pm 2.132 \sqrt{\frac{2(9)}{4}} = \pm 4.52$$

Al probar con $n = 3$ se obtiene

$$\pm 2.228 \sqrt{\frac{2(9)}{3}} = \pm 5.46$$

Evidentemente, $n = 4$ es el tamaño de la muestra menor que llevará a la precisión deseada.

El nivel de significación consignado en el ejemplo anterior se aplica a un solo intervalo de confianza. Sin embargo, puede usarse el mismo enfoque general si el experimentador desea especificar de antemano un *conjunto* de intervalos de confianza acerca del cual se hace un **enunciado de confianza simultáneo o conjunto** (ver los comentarios acerca de los intervalos de confianza simultáneos de la sección 3-3.3). Además, los intervalos de confianza podrían construirse con respecto a contrastes más generales en las medias de los tratamientos, que la comparación por pares ilustrada antes.

3-8 IDENTIFICACIÓN DE EFECTOS DE DISPERSIÓN

Nos hemos enfocado aquí en el uso del análisis de varianza y de otros métodos relacionados para determinar los niveles del factor que resultan en diferencias entre las medias de los tratamientos o los niveles del factor. Se acostumbra referirse a estos efectos como **efectos de localización**. Cuando ocurrió la desigual-

Tabla 3-12 Datos del experimento de fundición

Algoritmo para controlar la proporción	Observaciones					
	1	2	3	4	5	6
1	4.93(0.05)	4.86(0.04)	4.75(0.05)	4.95(0.06)	4.79(0.03)	4.88(0.05)
2	4.85(0.04)	4.91(0.02)	4.79(0.03)	4.85(0.05)	4.75(0.03)	4.85(0.02)
3	4.83(0.09)	4.88(0.13)	4.90(0.11)	4.75(0.15)	4.82(0.08)	4.90(0.12)
4	4.89(0.03)	4.77(0.04)	4.94(0.05)	4.86(0.05)	4.79(0.03)	4.76(0.02)

dad de la varianza con los diferentes niveles del factor, se utilizaron transformaciones para estabilizar la varianza y mejorar así las inferencias hechas sobre los efectos de localización. Sin embargo, en algunos problemas el interés se centra en descubrir si los diferentes niveles del factor afectan la **variabilidad**; es decir, el interés está en descubrir **efectos de dispersión** potenciales. Esto ocurrirá siempre que la desviación estándar, la varianza o cualquier otra medida de la variabilidad se use como variable de respuesta.

Para ilustrar estos conceptos, considere los datos de la tabla 3-12, los cuales se obtuvieron de un experimento diseñado en una fundición de aluminio. El aluminio se produce combinando alúmina con otros ingredientes en una celda de reacción y aplicando calor al hacer pasar una corriente eléctrica a través de la celda. La alúmina se agrega de manera continua a la celda para mantener la proporción apropiada de la misma con respecto a los otros ingredientes. En este experimento se investigaron cuatro algoritmos para controlar la proporción. Las variables de respuesta estudiadas se relacionaron con el voltaje de la celda. Específicamente, un sensor registra el voltaje de la celda varias veces cada segundo, produciendo miles de mediciones del voltaje durante cada corrida del experimento. Los ingenieros del proceso decidieron usar como variables de respuesta el voltaje promedio y la desviación estándar del voltaje de la celda (indicado entre paréntesis) en la corrida experimental. El voltaje promedio es importante porque afecta la temperatura de la celda, y la desviación estándar del voltaje (llamada “ruido del crisol” por los ingenieros del proceso) es importante porque afecta la eficiencia global de la celda.

Se llevó a cabo un análisis de varianza para determinar si los diferentes algoritmos para controlar la proporción afectan el voltaje promedio de la celda. Éste reveló que el algoritmo para controlar la proporción no tuvo ningún **efecto de localización**; es decir, al cambiar los algoritmos para controlar la proporción no hubo ningún cambio en el voltaje promedio de la celda. (Referirse al problema 3-28.)

Para investigar los efectos de dispersión, lo mejor suele ser utilizar

$$\log(s) \quad \text{o} \quad \log(s^2)$$

como variable de respuesta, ya que la transformación logarítmica es eficaz para estabilizar la variabilidad en la distribución de la desviación estándar muestral. Puesto que todas las desviaciones estándar del voltaje del crisol son menores que la unidad, se usará

$$y = -\ln(s)$$

como la variable de respuesta. En la tabla 3-13 se presenta el análisis de varianza para esta respuesta, el logaritmo natural del “ruido del crisol”. Observe que la elección de un algoritmo para controlar la proporción afecta el ruido del crisol; es decir, el algoritmo para controlar la proporción tiene un **efecto de dis-**

Tabla 3-13 Análisis de varianza del logaritmo natural del ruido del crisol

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Algoritmo para controlar la proporción	6.166	3	2.055	21.96	<0.001
Error	1.872	20	0.094		
Total	8.038	23			

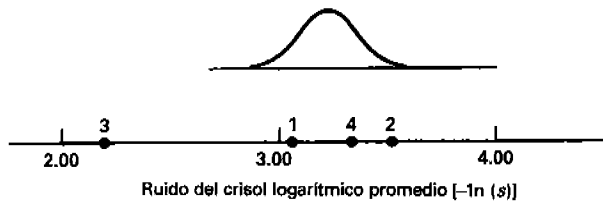


Figura 3-16 Ruido del crisol logarítmico promedio $[-\ln(s)]$ de cuatro algoritmos para controlar la proporción en relación con una distribución t escalada con factor de escalamiento $\sqrt{MS_E/n} = \sqrt{0.094/6} = 0.125$.

persión. Las pruebas estándares de la adecuación del modelo, incluyendo las gráficas de probabilidad normal de los residuales, indican que no hay problemas con la validez del experimento. (Referirse al problema 3-29.)

En la figura 3-16 se grafica el logaritmo promedio del ruido del crisol de cada algoritmo para controlar la proporción y se presenta también una distribución t escalada que se usa como **distribución de referencia** para discriminar entre los algoritmos de la proporción. Esta gráfica revela con toda claridad que el algoritmo 3 para controlar la proporción produce más ruido del crisol o una desviación estándar del voltaje de la celda mayor que los otros algoritmos. No parece haber gran diferencia entre los algoritmos 1, 2 y 4.

3-9 EL ENFOQUE DE REGRESIÓN PARA EL ANÁLISIS DE VARIANZA

Se ha ofrecido un desarrollo intuitivo o heurístico del análisis de varianza. Sin embargo, es posible presentar un desarrollo más formal. El método será de utilidad más adelante para entender los fundamentos del análisis estadístico de diseños más complejos. Llamada la **prueba general de significación de la regresión**, el procedimiento consiste en esencia en encontrar la reducción en la suma de cuadrados total para ajustar el modelo con todos los parámetros incluidos y la reducción en la suma de cuadrados cuando el modelo se restringe a la hipótesis nula. La diferencia entre estas dos sumas de cuadrados es la suma de cuadrados de los tratamientos con la que puede realizarse la prueba de la hipótesis nula. El procedimiento requiere los estimadores de mínimos cuadrados de los parámetros en el modelo del análisis de varianza. Se dieron ya (en la sección 3-3.3) las estimaciones de estos parámetros; sin embargo, ahora se presenta un desarrollo formal.

3-9.1 Estimación de mínimos cuadrados de los parámetros del modelo

Se desarrollan ahora los estimadores de los parámetros en el modelo con un solo factor

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

utilizando el método de mínimos cuadrados. Para encontrar los estimadores de mínimos cuadrados de μ y τ_i , primero se forma la suma de cuadrados de los errores

$$L = \sum_{i=1}^a \sum_{j=1}^n \varepsilon_{ij}^2 = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \mu - \tau_i)^2 \quad (3-51)$$

y se eligen después los valores de μ y τ_i , por ejemplo $\hat{\mu}$ y $\hat{\tau}_i$, que minimicen L . Los valores adecuados serían las soluciones de las $a + 1$ ecuaciones simultáneas

$$\begin{aligned} \frac{\partial L}{\partial \mu} \bigg|_{\hat{\mu}, \hat{\tau}_i} &= 0 \\ \frac{\partial L}{\partial \tau_i} \bigg|_{\hat{\mu}, \hat{\tau}_i} &= 0 \quad i = 1, 2, \dots, a \end{aligned}$$

Al derivar la ecuación 3-51 con respecto a μ y τ_i y al igualar con cero se obtiene

$$-2 \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \hat{\mu} - \hat{\tau}_i) = 0$$

y

$$-2 \sum_{j=1}^n (y_{ij} - \hat{\mu} - \hat{\tau}_i) = 0 \quad i = 1, 2, \dots, a$$

de la que, después de simplificar, se obtiene

$$\begin{aligned} N\hat{\mu} + n\hat{\tau}_1 + n\hat{\tau}_2 + \dots + n\hat{\tau}_a &= y_{..} \\ n\hat{\mu} + n\hat{\tau}_1 &= y_{1.} \\ n\hat{\mu} + n\hat{\tau}_2 &= y_{2.} \\ &\vdots \\ n\hat{\mu} + n\hat{\tau}_a &= y_{a.} \end{aligned} \quad (3-52)$$

A las $a + 1$ ecuaciones (ecuación 3-52) con $a + 1$ incógnitas se les llama las **ecuaciones normales de mínimos cuadrados**. Observe que si se suman la últimas a ecuaciones normales, se obtiene la primera ecuación normal. Por lo tanto, las ecuaciones normales no son linealmente independientes, y no existe una solución única para $\mu, \tau_1, \dots, \tau_a$. Esta dificultad puede superarse mediante varios métodos. Puesto que los efectos de los tratamientos se han definido como desviaciones de la media global, parece razonable aplicar la **restricción**

$$\sum_{i=1}^a \hat{\tau}_i = 0 \quad (3-53)$$

Utilizando esta restricción, se obtiene como solución de las ecuaciones normales

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..} \\ \hat{\tau}_i &= \bar{y}_{i.} - \bar{y}_{..} \quad i = 1, 2, \dots, a \end{aligned} \quad (3-54)$$

Evidentemente, esta solución no es única y depende de la restricción (ecuación 3-53) que se ha elegido. Al principio esto puede parecer desafortunado porque dos experimentadores diferentes podrían analizar los mismos datos y obtener resultados diferentes si aplican restricciones diferentes. Sin embargo, ciertas **funciones** del parámetro del modelo *son* estimadas de manera única, independientemente de la restricción. Algunos ejemplos son $\tau_i - \tau_j$, que se estimaría con $\hat{\tau}_i - \hat{\tau}_j = \bar{y}_{i.} - \bar{y}_{j.}$, y la media del tratamiento i -ésimo $\mu_i = \mu + \tau_i$, que se estimaría con $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i.}$.

Puesto que el interés se encuentra generalmente en las diferencias entre los efectos de los tratamientos y no en sus valores reales, no produce preocupación alguna que τ_i no pueda estimarse de manera úni-

ca. En general, cualquier función de los parámetros del modelo que sea una combinación lineal del miembro del lado izquierdo de las ecuaciones normales (ecuaciones 3-52) puede estimarse de manera única. A las funciones que se estiman de manera única independientemente de la restricción que se use se les llama **funciones estimables**. Para más información, ver el material suplementario del texto de este capítulo. Nos encontramos listos para usar estas estimaciones de los parámetros en un desarrollo general del análisis de varianza.

3-9.2 Prueba general de significación de la regresión

Una parte fundamental de este procedimiento es escribir las ecuaciones normales del modelo. Estas ecuaciones siempre podrán obtenerse formando la función de mínimos cuadrados y derivándola con respecto a cada parámetro desconocido, como se hizo en la sección 3-9.1. Sin embargo, se cuenta con un método más sencillo. Las reglas siguientes permiten escribir directamente las ecuaciones normales del modelo de *cualquier* diseño experimental:

REGLA 1. Hay una ecuación normal para cada parámetro del modelo que va a estimarse.

REGLA 2. El miembro derecho de cualquier ecuación normal es sólo la suma de todas las observaciones que contienen el parámetro asociado con esa ecuación normal particular.

Para ilustrar esta regla, considere el modelo con un solo factor. La primera ecuación normal corresponde al parámetro μ ; por lo tanto, el miembro derecho es $y_{..}$, ya que *todas* las observaciones incluyen a μ .

REGLA 3. El miembro izquierdo de cualquier ecuación normal es la suma de todos los parámetros del modelo, donde cada parámetro está multiplicado por el número de veces que aparece en el total del miembro derecho. Los parámetros se escriben con un acento circunflejo (^) para indicar que son *estimadores* y no los verdaderos valores de los parámetros.

Por ejemplo, considere la primera ecuación normal en un experimento con un solo factor. De acuerdo con las reglas anteriores, ésta sería

$$N\hat{\mu} + n\hat{\tau}_1 + n\hat{\tau}_2 + \cdots + n\hat{\tau}_a = y_{..}$$

porque μ aparece en las N observaciones, τ_1 sólo aparece en las n observaciones hechas bajo el primer tratamiento, τ_2 aparece sólo en las n observaciones tomadas bajo el segundo tratamiento, etc. Por la ecuación 3-52 se verifica que la ecuación presentada arriba es correcta. La segunda ecuación normal correspondería a τ_1 y es

$$n\hat{\mu} + n\hat{\tau}_1 = y_{1.}$$

porque sólo las observaciones del primer tratamiento contienen a τ_1 (esto da $y_{1.}$ como miembro derecho), μ y τ_1 aparecen exactamente n veces en $y_{1.}$, y todas las demás τ_i aparecen cero veces. En general, el miembro izquierdo de cualquier ecuación normal es el valor esperado del miembro derecho.

Ahora bien, considere encontrar la reducción en la suma de cuadrados ajustando un modelo particular a los datos. Al ajustar un modelo a los datos se “explica” parte de la variabilidad; es decir, la variabilidad no explicada se reduce en cierta cantidad. La reducción en la variabilidad no explicada es siempre la suma de las estimaciones de los parámetros, cada una de ellas multiplicada por el segundo miembro de la

ecuación normal que corresponde al parámetro específico. Por ejemplo, en un experimento con un solo factor, la reducción debida al ajuste del modelo completo $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ es

$$\begin{aligned} R(\mu, \tau) &= \hat{\mu}y_{..} + \hat{\tau}_1 y_{1.} + \hat{\tau}_2 y_{2.} + \cdots + \hat{\tau}_a y_{a.} \\ &= \hat{\mu}y_{..} + \sum_{i=1}^a \hat{\tau}_i y_{i.} \end{aligned} \quad (3-55)$$

La notación $R(\mu, \tau)$ significa la reducción en la suma de cuadrados a partir del ajuste del modelo que contiene a μ y $\{\tau_i\}$. A $R(\mu, \tau)$ se le llama en ocasiones la suma de cuadrados “de regresión” del modelo completo $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$. El número de grados de libertad asociado con una reducción en la suma de cuadrados, tal como $R(\mu, \tau)$, siempre es igual al número de ecuaciones normales linealmente independientes. El resto de la variabilidad no explicada por el modelo se encuentra con

$$SS_E = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - R(\mu, \tau) \quad (3-56)$$

Esta cantidad se usa en el denominador del estadístico de prueba de $H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$.

A continuación se ilustra la prueba general de significación de la regresión para un experimento con un solo factor y se demuestra que produce el análisis de varianza de un solo factor común. El modelo es $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, y las ecuaciones normales se encuentran con las reglas anteriores como

$$\begin{array}{rcl} N\hat{\mu} + n\hat{\tau}_1 + n\hat{\tau}_2 + \cdots + n\hat{\tau}_a & = & y_{..} \\ n\hat{\mu} + n\hat{\tau}_1 & = & y_{1.} \\ n\hat{\mu} & + & n\hat{\tau}_2 = y_{2.} \\ & & \vdots \\ n\hat{\mu} & & + n\hat{\tau}_a = y_{a.} \end{array}$$

Compare estas ecuaciones normales con las que se obtuvieron en la ecuación 3-52.

Al aplicar la restricción $\sum_{i=1}^a \hat{\tau}_i = 0$, los estimadores de μ y τ_i son

$$\hat{\mu} = \bar{y}_{..} \quad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} \quad i = 1, 2, \dots, a$$

La reducción en la suma de cuadrados debida al ajuste de este modelo completo se encuentra con la ecuación 3-55 como

$$\begin{aligned} R(\mu, \tau) &= \hat{\mu}y_{..} + \sum_{i=1}^a \hat{\tau}_i y_{i.} \\ &= (\bar{y}_{..})y_{..} + \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})y_{i.} \\ &= \frac{y_{..}^2}{N} + \sum_{i=1}^a \bar{y}_{i.} y_{i.} - \bar{y}_{..} \sum_{i=1}^a y_{i.} \\ &= \sum_{i=1}^a \frac{y_{i.}^2}{n} \end{aligned}$$

que tiene a grados de libertad porque hay a ecuaciones normales linealmente independientes. La suma de cuadrados del error es, por la ecuación 3-56,

$$\begin{aligned} SS_E &= \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - R(\mu, \tau) \\ &= \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \sum_{i=1}^a \frac{y_i^2}{n} \end{aligned}$$

y tiene $N - a$ grados de libertad.

Para encontrar la suma de cuadrados que resulta de los efectos de los tratamientos (el $\{\tau_i\}$), se considera que el modelo se restringe a la hipótesis nula; es decir, $\tau_i = 0$ para toda i . El modelo reducido es $y_{ij} = \mu + \varepsilon_{ij}$. Hay una sola ecuación normal para este modelo:

$$N\hat{\mu} = y_{..}$$

y el estimador de μ es $\hat{\mu} = \bar{y}_{..}$. Por lo tanto, la reducción en la suma de cuadrados que resulta de ajustar el modelo reducido que sólo contiene a μ es

$$R(\mu) = (\bar{y}_{..})(y_{..}) = \frac{y_{..}^2}{N}$$

Puesto que hay una sola ecuación normal para este modelo reducido, $R(\mu)$ tiene un grado de libertad. La suma de cuadrados debida al $\{\tau_i\}$, dado que μ ya está incluida en el modelo, es la diferencia entre $R(\mu, \tau)$ y $R(\mu)$, que es

$$\begin{aligned} R(\tau|\mu) &= R(\mu, \tau) - R(\mu) \\ &= R(\text{Modelo Completo}) - R(\text{Modelo Reducido}) \\ &= \frac{1}{n} \sum_{i=1}^a y_i^2 - \frac{y_{..}^2}{N} \end{aligned}$$

con $a - 1$ grados de libertad, que por la ecuación 3-9 se identifica como $SS_{\text{Tratamientos}}$. Estableciendo el supuesto de normalidad usual, el estadístico apropiado para probar $H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$ es

$$F_0 = \frac{R(\tau|\mu) / (a-1)}{\left[\sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - R(\mu, \tau) \right] / (N-a)}$$

que se distribuye como $F_{a-1, N-a}$ bajo la hipótesis nula. Se trata, desde luego, del estadístico de prueba para el análisis de varianza de un solo factor.

3-10 MÉTODOS NO PARAMÉTRICOS EN EL ANÁLISIS DE VARIANZA

3-10.1 La prueba de Kruskal-Wallis

En situaciones en las que el supuesto de normalidad no está justificado, el experimentador quizá quiera usar un procedimiento alternativo del análisis de varianza con la prueba F que no dependa de este su-

puesto. Kruskal y Wallis [68] han desarrollado este procedimiento. La prueba de Kruskal-Wallis se usa para probar la hipótesis nula de que los a tratamientos son idénticos contra la hipótesis alternativa de que algunos de los tratamientos generan observaciones que son mayores que otras. Debido a que el procedimiento está diseñado para ser sensible al probar las diferencias en las medias, en ocasiones es conveniente considerar la prueba de Kruskal-Wallis como una prueba de la igualdad de las medias de los tratamientos. La prueba de Kruskal-Wallis es una **alternativa no paramétrica** del análisis de varianza usual.

Para realizar la prueba de Kruskal-Wallis, primero se hace la clasificación en rangos de las y_{ij} observaciones en orden ascendente y cada observación se reemplaza con su rango, por ejemplo R_{ij} , asignándole a la observación menor el rango 1. En el caso de empates (observaciones que tienen el mismo valor), se asigna el rango promedio a cada una de las observaciones empatadas. Sea R_i la suma de los rangos del tratamiento i -ésimo. El estadístico de prueba es

$$H = \frac{1}{S^2} \left[\sum_{i=1}^a \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] \quad (3-57)$$

donde n_i es el número de observaciones del tratamiento i -ésimo, N es el número total de observaciones y

$$S^2 = \frac{1}{N-1} \left[\sum_{i=1}^a \sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4} \right] \quad (3-58)$$

Observe que S^2 es sólo la varianza de los rangos. Si no hay empates, $S^2 = N(N+1)/12$, y el estadístico de prueba se simplifica a

$$H = \frac{12}{N(N+1)} \sum_{i=1}^a \frac{R_i^2}{n_i} - 3(N+1) \quad (3-59)$$

Cuando el número de empates es moderado, habrá pequeñas diferencias entre las ecuaciones 3-57 y 3-59, y puede usarse la forma más simple (ecuación 3-59). Si las n_i son razonablemente grandes, por ejemplo $n_i \geq 5$, H se distribuye aproximadamente como χ^2_{a-1} bajo la hipótesis nula. Por lo tanto, si

$$H > \chi^2_{\alpha, a-1}$$

la hipótesis nula se rechaza. También podría usarse el enfoque del valor P .

EJEMPLO 3-12

En la tabla 3-14 se muestran los datos del ejemplo 3-1 y sus rangos correspondientes. Puesto que hay un número bastante grande de empates, la ecuación 3-57 se usa como el estadístico de prueba. Por la ecuación 3-58 se encuentra

$$\begin{aligned} S^2 &= \frac{1}{N-1} \left[\sum_{i=1}^a \sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4} \right] \\ &= \frac{1}{24} \left[5497.79 - \frac{25(26)^2}{4} \right] \\ &= 53.03 \end{aligned}$$

Tabla 3-14 Datos y rangos para el experimento de la resistencia a la tensión del ejemplo 3-1

Peso porcentual del algodón									
15		20		25		30		35	
y_{1j}	R_{1j}	y_{2j}	R_{2j}	y_{3j}	R_{3j}	y_{4j}	R_{4j}	y_{5j}	R_{5j}
7	2.0	12	9.5	14	11.0	19	20.5	7	2.0
7	2.0	17	14.0	18	16.5	25	25.0	10	5.0
15	12.5	12	9.5	18	16.5	22	23.0	11	7.0
11	7.0	18	16.5	19	20.5	19	20.5	15	12.5
9	4.0	18	16.5	19	20.5	23	24.0	11	7.0
$R_{.i}$	27.5		66.0		85.0		113.0		33.5

y el estadístico de prueba es

$$\begin{aligned}
 H &= \frac{1}{S^2} \left[\sum_{i=1}^a \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] \\
 &= \frac{1}{53.03} \left[5245.0 - \frac{25(26)^2}{4} \right] \\
 &= 19.25
 \end{aligned}$$

Puesto que $H > \chi_{0.01,4}^2$, $4 = 13.28$, se rechazaría la hipótesis nula y se concluiría que los tratamientos difieren. (El valor P para $H = 19.25$ es $P = 0.0002$.) Se trata de la misma conclusión obtenida por el análisis de varianza usual con la prueba F .

3-10.2 Comentarios generales sobre la transformación de rangos

Al procedimiento utilizado en la sección anterior de reemplazar las observaciones con sus rangos se le llama la **transformación de rangos**. Es una técnica muy poderosa y útil. Si se aplicara la prueba F común a los rangos en lugar de a los datos originales, se obtendría

$$F_0 = \frac{H/(a-1)}{(N-1-H)/(N-a)} \quad (3-60)$$

como el estadístico de prueba (ver Conover [20], p. 337). Observe que cuando el estadístico H de Kruskal-Wallis se incrementa o decrementa, F_0 también se incrementa o decrementa, por lo que la prueba de Kruskal-Wallis es equivalente a aplicar el análisis de varianza común a los rangos.

La transformación de rangos tiene una amplia aplicabilidad en los problemas de diseño experimental para los que no existe ninguna alternativa no paramétrica para el análisis de varianza. Esto incluye muchos de los diseños de capítulos subsecuentes de este libro. Si los datos están en rangos y se aplica la prueba F común, el resultado es un procedimiento aproximado que tiene buenas propiedades estadísticas (ver Conover e Iman [30a, b]). Cuando existe preocupación acerca del supuesto de normalidad o por el efecto de puntos atípicos o valores “absurdos”, se recomienda que el análisis de varianza común se realice tanto en los datos originales como en los rangos. Cuando ambos procedimientos producen resultados similares, probablemente los supuestos del análisis de varianza se satisfacen razonablemente, y el análisis estándar es satisfactorio. Cuando los dos procedimientos difieren, deberá darse preferencia a la transformación de rangos, ya que es menos posible que sea distorsionada por una condición de no normalidad o la presencia de observaciones inusuales. En tales casos, tal vez el experimentador quiera investigar el uso de transfor-

maciones para la falta de normalidad y examinar los datos y el procedimiento experimental a fin de determinar si hay puntos atípicos y por qué han ocurrido.

3-11 PROBLEMAS

- 3-1. Se estudia la resistencia a la tensión del cemento portland. Pueden usarse económicamente cuatro diferentes técnicas de mezclado. Se han colectado los siguientes datos:

Técnica de mezclado		Resistencia a la tensión (lb/pulg ²)			
1	3129	3000	2865	2890	
2	3200	3300	2975	3150	
3	2800	2900	2985	3050	
4	2600	2700	2600	2765	

- Probar la hipótesis de que las técnicas de mezclado afectan la resistencia del cemento. Utilizar $\alpha = 0.05$.
 - Construir una representación gráfica como se describió en la sección 3-5.3 para comparar las resistencias a la tensión promedio de las cuatro técnicas de mezclado. ¿A qué conclusiones se llega?
 - Usar el método LSD de Fisher con $\alpha = 0.05$ para hacer comparaciones entre pares de medias.
 - Construir una gráfica de probabilidad normal de los residuales. ¿Qué conclusiones se sacarían acerca de la validez del supuesto de normalidad?
 - Graficar los residuales contra la resistencia a la tensión predicha. Comentar la gráfica.
 - Hacer un diagrama de dispersión de los resultados como ayuda para la interpretación de los resultados de este experimento.
- 3-2.
 - Resolver de nuevo el inciso *b* del problema 3-1 utilizando la prueba del rango múltiple de Duncan con $\alpha = 0.05$. ¿Hay alguna diferencia en las conclusiones?
 - Resolver de nuevo el inciso *b* del problema 3-1 utilizando la prueba de Tukey con $\alpha = 0.05$. ¿Se llega a las mismas conclusiones con la prueba de Tukey que las obtenidas con el procedimiento gráfico y/o con la prueba del rango múltiple de Duncan?
 - Explicar la diferencia entre los procedimientos de Duncan y de Tukey.
- 3-3. Considere nuevamente el problema 3-1. Encontrar un intervalo de confianza de 95% para la resistencia a la tensión media del cemento portland que produce cada una de las cuatro técnicas de mezclado. Encontrar también un intervalo de confianza de 95% para la diferencia en las medias de las técnicas 1 y 3. ¿Sirve esto de ayuda para interpretar los resultados del experimento?
- 3-4. Se llevó a cabo un experimento a fin de determinar si cuatro temperaturas de cocción específicas afectan la densidad de cierto tipo de ladrillo. El experimento produjo los siguientes datos:

Temperatura		Densidad			
100	21.8	21.9	21.7	21.6	21.7
125	21.7	21.4	21.5	21.4	
150	21.9	21.8	21.8	21.6	21.5
175	21.9	21.7	21.8	21.4	

- ¿La temperatura de cocción afecta la densidad de los ladrillos? Utilizar $\alpha = 0.05$.
- ¿Es apropiado comparar las medias utilizando la prueba del rango múltiple de Duncan (por ejemplo) en este experimento?
- Analizar los residuales de este experimento. ¿Se satisfacen los supuestos del análisis de varianza?
- Construir una representación gráfica de los tratamientos como se describió en la sección 3-5.3. ¿Esta gráfica resume adecuadamente los resultados del análisis de varianza del inciso *a*?

- 3-5. Resolver de nuevo el inciso *d* del problema 3-4 utilizando el método LSD de Fisher. ¿A qué conclusiones se llega? Explicar en detalle cómo se modificó la técnica para tomar en cuenta los tamaños de las muestras desiguales.
- 3-6. Un fabricante de televisores está interesado en el efecto de cuatro tipos diferentes de recubrimientos para cinescopios de color sobre la conductividad de un cinescopio. Se obtienen los siguientes datos de la conductividad:

Tipo de recubrimiento	Conductividad			
1	143	141	150	146
2	152	149	137	143
3	134	136	132	127
4	129	127	132	129

- a) ¿Hay alguna diferencia en la conductividad debida al tipo de recubrimiento? Utilizar $\alpha = 0.05$.
- b) Estimar la media global y los efectos de los tratamientos.
- c) Calcular la estimación de un intervalo de confianza de 95% para la media del tipo de recubrimiento 4. Calcular la estimación de un intervalo de confianza de 99% para la diferencia media entre los tipos de recubrimiento 1 y 4.
- d) Probar todos los pares de medias utilizando el método LSD de Fisher con $\alpha = 0.05$.
- e) Usar el método gráfico comentado en la sección 3-5.3 para comparar las medias. ¿Cuál es el tipo de recubrimiento que produce la conductividad más alta?
- f) Suponiendo que el recubrimiento tipo 4 es el que se está usando actualmente, ¿qué se recomendaría al fabricante? Quiere minimizarse la conductividad.
- 3-7. Considere nuevamente el experimento del problema 3-6. Analizar los residuales y sacar conclusiones acerca de la adecuación del modelo.
- 3-8. En un artículo de *ACI Materials Journal* (vol. 84, pp. 213-216) se describen varios experimentos para investigar el varillado del concreto para eliminar el aire atrapado. Se usó un cilindro de 3 × 6 pulgadas; y el número de veces que esta barra se utilizó es la variable del diseño. La resistencia a la compresión resultante de la muestra de concreto es la respuesta. Los datos se muestran en la tabla siguiente:

Nivel de varillado	Resistencia a la compresión		
10	1530	1530	1440
15	1610	1650	1500
20	1560	1730	1530
25	1500	1490	1510

- a) ¿Hay alguna diferencia en la resistencia a la compresión debida al nivel de varillado? Utilizar $\alpha = 0.05$.
- b) Encontrar el valor *P* para el estadístico *F* del inciso *a*.
- c) Analizar los residuales de este experimento. ¿Qué conclusiones pueden sacarse acerca de los supuestos fundamentales del modelo?
- d) Construir una representación gráfica para comparar las medias de los tratamientos, como se describió en la sección 3-5.3.
- 3-9. En un artículo de *Environment International* (vol. 18, no. 4) se describe un experimento en el que se investigó la cantidad de radón liberado en las duchas. Se usó agua enriquecida con radón en el experimento, y se probaron seis diámetros diferentes de los orificios de las regaderas. Los datos del experimento se muestran en la siguiente tabla:

Diámetro de los orificios	Radón liberado (%)			
0.37	80	83	83	85
0.51	75	75	79	79
0.71	74	73	76	77
1.02	67	72	74	74
1.40	62	62	67	69
1.99	60	61	64	66

- a) ¿El tamaño de los orificios afecta el porcentaje promedio del radón liberado? Utilizar $\alpha = 0.05$.
- b) Encontrar el valor P para el estadístico F del inciso a.
- c) Analizar los residuales de este experimento.
- d) Encontrar un intervalo de confianza de 95% para el porcentaje promedio de radón liberado cuando el diámetro de los orificios es 1.40.
- e) Construir una representación gráfica para comparar las medias de los tratamientos, como se describió en la sección 3-5.3. ¿Qué conclusiones pueden sacarse?
- 3-10. Se determinó el tiempo de respuesta en milisegundos para tres diferentes tipos de circuitos que podrían usarse en un mecanismo de desconexión automática. Los resultados se muestran en la siguiente tabla:

Tipo de circuito	Tiempo de respuesta				
1	9	12	10	8	15
2	20	21	23	17	30
3	6	5	8	16	7

- a) Probar la hipótesis de que los tres tipos de circuitos tienen el mismo tiempo de respuesta. Utilizar $\alpha = 0.01$.
- b) Usar la prueba de Tukey para comparar pares de medias de los tratamientos. Utilizar $\alpha = 0.01$.
- c) Usar el procedimiento gráfico de la sección 3-5.3 para comparar las medias de los tratamientos. ¿Qué conclusiones pueden sacarse? ¿Cómo se comparan con las conclusiones del inciso b?
- d) Construir un conjunto de contrastes ortogonales, suponiendo que al principio del experimento se sospechaba que el tiempo de respuesta del circuito tipo 2 era diferente del de los otros dos.
- e) Si el lector fuera el ingeniero de diseño y quisiera minimizar el tiempo de respuesta, ¿qué tipo de circuito seleccionaría?
- f) Analizar los residuales de este experimento. ¿Se satisfacen los supuestos del análisis de varianza básico?
- 3-11. Se estudia la vida efectiva de los fluidos aislantes en una carga acelerada de 35 kV. Se han obtenido datos de una prueba para cuatro tipos de fluidos. Los resultados fueron los siguientes:

Tipo de fluido	Vida (en horas) con 35 kV de carga					
1	17.6	18.9	16.3	17.4	20.1	21.6
2	16.9	15.3	18.6	17.1	19.5	20.3
3	21.4	23.6	19.4	18.5	20.5	22.3
4	19.3	21.1	16.9	17.5	18.3	19.8

- a) ¿Hay algún indicio de que los fluidos difieran? Utilizar $\alpha = 0.05$.
 b) ¿Cuál fluido seleccionaría el lector, dado que el objetivo es conseguir la vida efectiva más larga?
 c) Analizar los residuales de este experimento. ¿Se satisfacen los supuestos del análisis de varianza básico?
- 3-12. Se estudian cuatro diferentes tipos de diseños de un circuito digital de computadora para comparar la cantidad de ruido presente. Se obtienen los siguientes datos:

Diseño del circuito		Ruido observado			
1	19	20	19	30	8
2	80	61	73	56	80
3	47	26	25	35	50
4	95	46	83	78	97

- a) ¿La cantidad de ruido presente es la misma para los cuatro diseños? Utilizar $\alpha = 0.05$.
 b) Analizar los residuales de este experimento. ¿Se satisfacen los supuestos del análisis de varianza?
 c) ¿Qué diseño del circuito se seleccionaría para usarlo? El ruido bajo es mejor.
- 3-13. Se pide a cuatro químicos que determinen el porcentaje de alcohol metílico en cierto compuesto químico. Cada químico hace tres determinaciones, y los resultados son los siguientes:

Químico	Porcentaje de alcohol metílico		
1	84.99	84.04	84.38
2	85.15	85.13	84.88
3	84.72	84.48	85.16
4	84.20	84.10	84.55

- a) ¿Los químicos difieren significativamente? Utilizar $\alpha = 0.05$.
 b) Analizar los residuales de este experimento.
 c) Si el químico 2 es un empleado nuevo, construir un conjunto razonable de contrastes ortogonales que podría haberse usado al principio del experimento.
- 3-14. Se someten a estudio tres marcas de baterías. Se sospecha que las vidas (en semanas) de las tres marcas son diferentes. Se prueban cinco baterías de cada marca con los resultados siguientes:

Semanas de vida		
Marca 1	Marca 2	Marca 3
100	76	108
96	80	100
92	75	96
96	84	98
92	82	100

- a) ¿Las vidas de estas tres marcas son diferentes?
 b) Analizar los residuales de este experimento.
 c) Construir la estimación de un intervalo de confianza de 95% para la vida media de la batería marca 2. Construir la estimación del intervalo de confianza de 99% para la diferencia media entre las vidas de las baterías marcas 2 y 3.

- d) ¿Qué marca seleccionaría el lector para usarla? Si el fabricante reemplazara sin cargo cualquier batería que dure menos de 85 semanas, ¿qué porcentaje esperaría reemplazar la compañía?
- 3-15. Se están investigando cuatro catalizadores que pueden afectar la concentración de un componente en una mezcla líquida de tres componentes. Se obtienen las siguientes concentraciones:

Catalizador			
1	2	3	4
58.2	56.3	50.1	52.9
57.2	54.5	54.2	49.9
58.4	57.0	55.4	50.0
55.8	55.3		51.7
54.9			

- a) ¿Los cuatro catalizadores tienen el mismo efecto sobre la concentración?
- b) Analizar los residuales de este experimento.
- c) Construir la estimación de un intervalo de confianza de 99% para la respuesta media del catalizador 1.
- 3-16. Se llevó a cabo un experimento para investigar la eficacia de cinco materiales aislantes. Se probaron cuatro muestras de cada material con un nivel elevado de voltaje para acelerar el tiempo de falla. Los tiempos de falla (en minutos) se muestran abajo:

Material	Tiempo de falla (minutos)			
1	110	157	194	178
2	1	2	4	18
3	880	1256	5276	4355
4	495	7040	5307	10,050
5	7	5	29	2

- a) ¿Los cinco materiales tienen el mismo efecto sobre el tiempo de falla?
- b) Graficar los residuales contra la respuesta predicha. Construir una gráfica de probabilidad normal de los residuales. ¿Qué información transmiten estas gráficas?
- c) Con base en la respuesta del inciso b, realizar otro análisis de los datos del tiempo de falla y sacar las conclusiones apropiadas.
- 3-17. Un fabricante de semiconductores ha desarrollado tres métodos diferentes para reducir el conteo de partículas en las obleas. Los tres métodos se prueban en cinco obleas y se obtiene el conteo de partículas después del tratamiento. Los datos se muestran abajo:

Método	Conteo				
1	31	10	21	4	1
2	62	40	24	30	35
3	53	27	120	97	68

- a) ¿Todos los métodos tienen el mismo efecto sobre el conteo promedio de partículas?
- b) Graficar los residuales contra la respuesta predicha. Construir una gráfica de probabilidad normal de los residuales. ¿Hay motivo de preocupación potencial acerca de la validez de los supuestos?
- c) Con base en la respuesta del inciso b, realizar otro análisis de los datos del conteo de partículas y sacar las conclusiones apropiadas.

- 3-18. Considere la prueba de la igualdad de las medias de dos poblaciones normales, donde las varianzas son desconocidas pero se suponen iguales. El procedimiento de prueba apropiado es la prueba t agrupada o combinada. Demostrar que la prueba t combinada es equivalente al análisis de varianza de un solo factor.
- 3-19. Demostrar que la varianza de la combinación lineal $\sum_{i=1}^n c_i y_i$ es $\sigma^2 \sum_{i=1}^n n_i c_i^2$.
- 3-20. En un experimento con efectos fijos, suponga que hay n observaciones para cada uno de cuatro tratamientos. Sean Q_1^2 , Q_2^2 , Q_3^2 los componentes con un solo grado de libertad de los contrastes ortogonales. Demostrar que $SS_{\text{Tratamientos}} = Q_1^2 + Q_2^2 + Q_3^2$.
- 3-21. Utilizar la prueba de Bartlett para determinar si el supuesto de la igualdad de las varianzas se satisface en el problema 3-14. Utilizar $\alpha = 0.05$. ¿Se llegó a la misma conclusión respecto de la igualdad de las varianzas con el examen de las gráficas de los residuales?
- 3-22. Utilizar la prueba de Levene modificada para determinar si el supuesto de las varianzas iguales se satisface en el problema 3-14. Utilizar $\alpha = 0.05$. ¿Se llegó a la misma conclusión respecto de la igualdad de las varianzas con el examen de las gráficas de los residuales?
- 3-23. Referirse al problema 3-10. Si quiere detectarse una diferencia máxima en los tiempos de respuesta promedio de 10 milisegundos con una probabilidad de al menos 0.90, ¿qué tamaño de la muestra deberá usarse? ¿Cómo se obtendría una estimación preliminar de σ^2 ?
- 3-24. Referirse al problema 3-14.
- Si quiere detectarse una diferencia máxima en la vida de las baterías de 10 horas con una probabilidad de al menos 0.90, ¿qué tamaño de la muestra deberá usarse? Comentar cómo se obtendría una estimación preliminar de σ^2 para responder esta pregunta.
 - Si la diferencia entre las marcas es lo suficientemente grande para que la desviación estándar de una observación se incremente en 25%, ¿qué tamaño de la muestra deberá usarse si quiere detectarse esto con una probabilidad de al menos 0.90?
- 3-25. Considere el experimento del problema 3-14. Si quiere construirse un intervalo de confianza de 95% para la diferencia en las vidas medias de dos baterías que tenga una precisión de ± 2 semanas, ¿cuántas baterías de cada marca deben probarse?
- 3-26. Suponga que cuatro poblaciones normales tienen medias $\mu_1 = 50$, $\mu_2 = 60$, $\mu_3 = 50$ y $\mu_4 = 60$. ¿Cuántas observaciones deberán hacerse en cada población para que la probabilidad de rechazar la hipótesis nula de la igualdad de las medias poblacionales sea al menos 0.90? Suponer que $\alpha = 0.05$ y que una estimación razonable de la varianza de error es $\sigma^2 = 25$.
- 3-27. Referirse al problema 3-26.
- ¿En qué forma cambiaría la respuesta si una estimación razonable de la varianza del error experimental fuera $\sigma^2 = 36$?
 - ¿En qué forma cambiaría la respuesta si una estimación razonable de la varianza del error experimental fuera $\sigma^2 = 49$?
 - ¿Puede sacarse alguna conclusión acerca de la sensibilidad de la respuesta dada en esta situación particular acerca de cómo afecta la estimación de σ la decisión referente al tamaño de la muestra?
 - ¿Puede hacerse alguna recomendación acerca de cómo debería usarse este enfoque general para elegir n en la práctica?
- 3-28. Referirse al experimento de la fundición de aluminio descrito en la sección 3-8. Verificar que los métodos para controlar la proporción de alúmina no afectan el voltaje promedio de la celda. Construir una gráfica de probabilidad normal de los residuales. Graficar los residuales contra los valores predichos. ¿Existe algún indicio de que se violan algunos de los supuestos fundamentales?
- 3-29. Referirse al experimento de la fundición de aluminio de la sección 3-8. Verificar el análisis de varianza del ruido del crisol que se resume en la tabla 3-13. Examinar las gráficas de los residuales usuales y comentar la validez del experimento.
- 3-30. Se investigaron cuatro diferentes velocidades de alimentación en un experimento con una máquina CNC que produce una pieza que se usa en la unidad de potencia auxiliar de un avión. El ingeniero de manufactura a cargo del experimento sabe que una dimensión crítica de la pieza de interés puede ser afectada por la velocidad de alimentación. Sin embargo, la experiencia previa indica que es probable que sólo estén presentes

efectos de dispersión. Es decir, al cambiarse la velocidad de alimentación no se afecta la dimensión *promedio*, pero podría afectarse la variabilidad dimensional. El ingeniero realiza cinco corridas de producción con cada velocidad de alimentación y obtiene la desviación estándar de la dimensión crítica (en 10^{-3} mm). Los datos se muestran abajo. Suponer que todas las corridas se hicieron en orden aleatorio.

Velocidad de alimentación (pulgadas/minuto)	Corrida de producción				
	1	2	3	4	5
10	0.09	0.10	0.13	0.08	0.07
12	0.06	0.09	0.12	0.07	0.12
14	0.11	0.08	0.08	0.05	0.06
16	0.19	0.13	0.15	0.20	0.11

- a) ¿La velocidad de alimentación tiene algún efecto sobre la desviación estándar de esta dimensión crítica?
 - b) Usar los residuales de este experimento para investigar la adecuación del modelo. ¿Hay algún problema con la validez experimental?
- 3-31. Considere los datos del problema 3-10.
- a) Escribir las ecuaciones normales de mínimos cuadrados para este problema y resolverlas para $\hat{\mu}$ y $\hat{\tau}_i$, utilizando la restricción usual ($\sum_{i=1}^3 \hat{\tau}_i = 0$). Estimar $\tau_1 - \tau_2$.
 - b) Resolver las ecuaciones del inciso a utilizando la restricción $\hat{\tau}_3 = 0$. ¿Los estimadores $\hat{\tau}_i$ y $\hat{\mu}$ son los mismos que se encontraron en el inciso a? ¿Por qué? Estimar ahora $\tau_1 - \tau_2$ y comparar la respuesta con la del inciso a. ¿Qué afirmación puede hacerse respecto de estimar los contrastes en las τ_i ?
 - c) Estimar $\mu + \tau_1$, $2\tau_1 - \tau_2 - \tau_3$, y $\mu + \tau_1 + \tau_2$ utilizando las dos soluciones de las ecuaciones normales. Comparar los resultados obtenidos en cada caso.
- 3-32. Aplicar la prueba general de significación de la regresión en el experimento del ejemplo 3-1. Demostrar que el procedimiento produce los mismos resultados que el análisis de varianza usual.
- 3-33. Usar la prueba de Kruskal-Wallis en el experimento del problema 3-11. Comparar las conclusiones obtenidas con las del análisis de varianza usual.
- 3-34. Usar la prueba de Kruskal-Wallis en el experimento del problema 3-12. ¿Los resultados son comparables con los encontrados por el análisis de varianza usual?
- 3-35. Considere el experimento del ejemplo 3-1. Suponga que la observación mayor de la resistencia a la tensión se registró incorrectamente como 50. ¿Qué efecto tiene esto sobre el análisis de varianza usual? ¿Qué efecto tiene sobre la prueba de Kruskal-Wallis?