

Método de Bonett

SOBRE EL INTERVALO DE CONFIANZA ROBUSTO DE BONETT PARA UNA RELACIÓN DE DESVIACIONES ESTÁNDAR

Senin J. Banga y Gregory D. Fox
18 de junio de 2013

RESUMEN

Proponemos un procedimiento alternativo para corregir un error sutil en el intervalo de confianza (IC) de Bonett (2006) para la relación de dos desviaciones estándar. El estimador agrupado de la curtosis para el estadístico de la prueba de Layard (1973), en el que se basa el intervalo de Bonett, solo es consistente cuando las varianzas de las poblaciones son iguales. Derivamos un estimador alternativo que es consistente cuando las varianzas de las poblaciones son iguales y cuando no son iguales y utilizamos el nuevo estimador para calcular el IC correcto. Los estudios de simulación revelan que, en general, el nuevo IC es más exacto y más preciso que el IC basado en la prueba W_{50} de Levene/Browne-Forsythe y la prueba L_{50} de Pan (1999). De acuerdo con Pan, observamos que los IC basados en la prueba W_{50} exhiben una pérdida de precisión con muestras pequeñas, produciendo frecuentemente intervalos que tienen una anchura infinita. Los IC que se basan en prueba L_{50} funcionan bien con distribuciones simétricas y casi simétricas, pero no funcionan adecuadamente cuando las poblaciones son asimétricas.

Términos de índice: homogeneidad de varianzas, prueba de Levene, prueba de Brown-Forsythe, prueba de Layard, intervalo de confianza (IC) para la relación de varianzas

1. Introducción

Es bien sabido que la prueba F clásica y los intervalos de confianza (IC) asociados son extremadamente sensibles a las desviaciones de la normalidad; de hecho, tan sensibles que la prueba F clásica no es adecuada para la mayoría de las aplicaciones prácticas. Por esa razón, muchos han propuesto alternativas más robustas. Entre éstas, la prueba conocida como "Prueba W_{50} " suele preferirse porque tiene muy buenas propiedades con respecto al error Tipo I y también es fácil de calcular y de interpretar. (Para los análisis comparativos, consulte Conover et

al. (1981), Balakrishnan y Ma (1990), y Lim y Loh (1996).) La prueba W_{50} se basa en un procedimiento que fue propuesto originalmente por Levene (1960) y mejorado después por Brown y Forsythe (1974). La prueba W_{50} ha sido adoptada ampliamente y se incluye en la mayoría de los paquetes de software más conocidos de herramientas estadísticas, como Minitab Statistical Software, SAS, R y JMP.

Las propiedades de la prueba W_{50} con respecto al error Tipo II son relativamente menos notables que sus propiedades en relación con el error Tipo I. Pan (1999) muestra que, para algunas distribuciones, incluyendo la distribución normal, la potencia de la prueba W_{50} en los problemas de dos muestras tiene un límite superior que posiblemente está muy por debajo de 1. Este límite superior no se ve afectado por la magnitud de la diferencia entre las varianzas de las poblaciones. Esta deficiencia se extiende naturalmente a los IC que se basan en la prueba W_{50} . Pan indica que existe una probabilidad no despreciable de que un IC para el cociente de las varianzas de las poblaciones que esté basado en la prueba W_{50} sea infinito $(0, +\infty)$ y, por lo tanto, poco informativo. La observación de Pan es consistente con los resultados de nuestras propias simulaciones, que se presentan más adelante en este documento.

Pan propone un procedimiento alternativo, denominado L_{50} , para corregir las limitaciones del procedimiento W_{50} . Con base en los resultados de la simulación, Pan concluye que la prueba L_{50} es más potente que la prueba W_{50} y, sin embargo, es igualmente robusta y comparte sus convenientes propiedades asintóticas. Sin embargo, las muestras utilizadas para las simulaciones de Pan se extrajeron de distribuciones simétricas o ligeramente asimétricas con colas que van de pesadas a livianas. El posible impacto de la asimetría en el rendimiento de la prueba L_{50} con muestras pequeñas no se trató específicamente.

Pan también sostiene que el procedimiento L_{50} es tan potente como otros procedimientos notablemente robustos, tales como la prueba modificada de Fligner-Killeen y la prueba adaptativa de Hall-Padmanabhan. Sin embargo, desde el punto de vista práctico, la prueba modificada de Fligner-Killeen y la prueba adaptativa de Hall-Padmanabhan son relativamente menos útiles que las pruebas L_{50} y W_{50} , porque requieren muchos cálculos complicados.

Recientemente, Bonett (2006) propuso un procedimiento alternativo de IC que se basa en la versión para dos muestras de la prueba de Layard (1973) de la homogeneidad de las varianzas. Bonett incluye varios ajustes para mejorar el rendimiento del procedimiento de Layard con muestras pequeñas. Por ejemplo, Bonett propone un estimador agrupado de la curtosis que es asintóticamente equivalente al de Layard, pero que exhibe menos sesgo con muestras pequeñas.

Desafortunadamente, ni el estimador agrupado de la curtosis propuesto originalmente por Layard ni el sustituto propuesto por Bonett son consistentes cuando las varianzas de las poblaciones no son iguales. Por consiguiente, los intervalos que propone Bonett (2006) no son intervalos de confianza adecuados, pero se describen mejor como intervalos de aceptación para la prueba de igualdad de varianzas. Por lo tanto, al restar las probabilidades de cobertura simulada especificadas en Bonett (2006) de la unidad se obtienen las tasas de error Tipo I para la prueba de igualdad de varianzas. La comparación de estas tasas de error Tipo I con las de la prueba original de Layard confirma que los ajustes de Bonett mejoran considerablemente el

rendimiento de la prueba de Layard con muestras pequeñas. Sin embargo, el IC para la relación de las varianzas propuesto por Bonett debe reconsiderarse.

Observe también que Bonett compara los intervalos propuestos con los IC basados en la prueba F aproximada de Shoemaker (2003). Sin embargo, el IC para la relación de las varianzas asociado con la prueba de Shoemaker, que se describe brevemente en la página 106 del informe de Shoemaker, también se basa en el estimador agrupado de la curtosis de Layard. Por lo tanto, los IC calculados en la sección 7 del informe de Shoemaker también se describen mejor como intervalos de aceptación para la prueba de igualdad de varianzas. A pesar de estos errores, con base en los resultados de la simulación de Bonett, se puede concluir que su ajuste mejoró el rendimiento de la prueba de igualdad de varianzas de Layard con muestras pequeñas y que la prueba resultante de igualdad de varianzas funciona mejor que la prueba de Shoemaker.

En el presente trabajo, corregimos el paso en falso de Bonett (2006) extendiendo la forma de dos muestras de la prueba de Layard para probar las hipótesis nulas sobre la relación de las varianzas o desviaciones estándar. Para ello, proponemos un estimador agrupado de la curtosis que es consistente para cualquier relación hipotética dada. Luego invertimos el estadístico de la prueba para obtener el IC para la relación. Por último, realizamos estudios de simulación para evaluar las propiedades de robustez del nuevo IC en diseños de muestras pequeñas. Además, comparamos el rendimiento del nuevo IC para muestras pequeñas con el rendimiento de los IC asociados con la prueba F clásica, la prueba W_{50} y la prueba L_{50} .

2. Prueba de Layard y cierta extensión

Supongamos que $Y_{i1}, \dots, Y_{in_i}, \dots, Y_{k1}, \dots, Y_{kn_k}$ son k muestras independientes, donde cada una de las muestras es independiente y está distribuida idénticamente con media $E(Y_{ij}) = \mu_i$ y varianza $\text{Var}(Y_{ij}) = \sigma_i^2 > 0$. Además, supongamos que las muestras provienen de poblaciones con una curtosis común $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$. Observamos que Layard utiliza el exceso de curtosis $\gamma_e = \gamma - 3$.

Supongamos que \bar{Y}_i y S_i son la media y la desviación estándar de la muestra i , respectivamente. Supongamos también que $\tau^2 = 2 + (1 - 1/\bar{n})\gamma_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$, donde $\bar{n} = \sum n_i/k$. Como se indica en Layard (1973), para muestras grandes, $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$.

Para probar la hipótesis nula de igualdad de varianzas, Layard realiza una transformación ortogonal del vector cuyos componentes $Z_i = (n_i - 1)^{1/2} \ln S_i^2 / \tau$ se distribuyen asintóticamente como la distribución normal estándar bajo la hipótesis nula. Posteriormente, utiliza la propiedad de preservación de la distancia de las transformaciones ortogonales para mostrar que el estadístico de prueba S' (especificado abajo) se distribuye asintóticamente como una distribución de chi-cuadrado con $k - 1$ grados de libertad bajo la hipótesis nula de igualdad de varianzas:

$$S' = \sum_{i=1}^k (n_i - 1) \left(\ln S_i^2 - \frac{\sum_{i=1}^k (n_i - 1) \ln S_i^2}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

En general, $Z_i = (n_i - 1)^{1/2} (\ln S_i^2 - \ln \sigma_i^2) / \tau$ se distribuye asintóticamente como la distribución normal estándar. Por lo tanto, las técnicas de Layard pueden aplicarse para derivar el estadístico de prueba T'_k más generalizado:

$$T'_k = \sum_{i=1}^k (n_i - 1) \frac{(\ln S_i^2 - \ln \sigma_i^2)^2}{\tau^2} - \left(\sum_{i=1}^k (n_i - 1) \frac{\ln S_i^2 - \ln \sigma_i^2}{\tau \sqrt{\sum_{i=1}^k (n_i - 1)}} \right)^2$$

T'_k se distribuye asintóticamente como una distribución de chi-cuadrado con $k - 1$ grados de libertad bajo la hipótesis nula y la hipótesis alternativa.

T'_k puede expresarse de una forma que es más similar a la de S' . Al expresar el término cuadrático como una suma doble y realizar algunos cálculos algebraicos, se obtiene lo siguiente:

$$T'_k = \sum_{i=1}^k (n_i - 1) \left(\ln S_i^2 - \ln \sigma_i^2 - \frac{\sum_{i=1}^k (n_i - 1) (\ln S_i^2 - \ln \sigma_i^2)}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

Si todas las varianzas son iguales, entonces $T'_k = S'$. Por lo tanto, S' y T'_k son el mismo estadístico de prueba cuando se evalúa la hipótesis nula de igualdad de varianzas. Sin embargo, T'_k también puede utilizarse de manera más general para probar cualquier hipótesis que se exprese en función de las varianzas. Por ejemplo, T'_k puede usarse para probar cualquier hipótesis nula con la forma $H_0: \sigma_i = \sigma_{0i}$ para cualquier $\sigma_{0i} > 0, i = 1, \dots, k$ dado.

Como $\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$ no se conoce, una prueba basada en S' o T'_k requiere un estimador para la curtosis común de las poblaciones, γ . Por ejemplo, para probar la hipótesis nula de homogeneidad de las varianzas, Layard propone el siguiente estimador agrupado de la curtosis común:

$$\hat{\gamma} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^4}{\left[\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right]^2} \sum_{i=1}^k n_i$$

No obstante, Layard señala que $\hat{\gamma}$ no es necesariamente un estimador consistente de la curtosis común cuando las varianzas no son iguales.

En el caso especial de los diseños de dos muestras, la magnitud de la diferencia entre las desviaciones estándar puede evaluarse probando la hipótesis nula $H_0: \sigma_1/\sigma_2 = \rho_0$ para cierta relación hipotética dada $\rho_0 > 0$. Sin embargo, esta diferencia puede evaluarse de una manera más directa al calcular el IC para la relación de las desviaciones estándar.

Si $\rho_0 = 1$, entonces la hipótesis nula es equivalente a la hipótesis de homogeneidad de las varianzas. Por lo tanto, la prueba puede basarse en $T'_2 = S'$, después de sustituir la versión de dos muestras del estimado de curtosis de Layard por γ en la expresión de $\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$ para obtener $\hat{\tau}^2$.

Sin embargo, si $\rho_0 \neq 1$, entonces la prueba debe basarse en T'_2 y no en S' . Además, si $\rho_0 \neq 1$, entonces el estimador agrupado de la curtosis de Layard no es necesariamente consistente y, por lo tanto, no puede usarse para estimar la curtosis común de las poblaciones. Por lo tanto, se necesita un estimador agrupado de la curtosis alternativo, uno que sea consistente para cualquier relación hipotética $\rho_0 > 0$.

A continuación derivamos dicho estimador. Como es una función de ρ_0 , denotamos el estimador como $\hat{\gamma}_P(\rho_0)$. También definimos el estadístico de prueba $T_2 = \tau^2 T'_2 / \hat{\tau}^2$, donde $\hat{\tau}^2 = 2 + (1 - 1/\bar{n})(\hat{\gamma}_P(\rho_0) - 3)$. Según el teorema de Slutsky, T_2 se distribuye asintóticamente como una distribución de chi-cuadrada con 1 grado de libertad. Por último, invertimos T_2 para obtener los IC para $\rho = \sigma_1/\sigma_2$.

3. IC para la relación de las desviaciones estándar

La sección anterior detalla la necesidad de un estimador alternativo de la curtosis cuando se prueban las hipótesis nulas que se expresan en términos de la relación de las varianzas o desviaciones estándar. El resultado siguiente proporciona ese estimador.

RESULTADO 1

Para cualquier $\rho = \sigma_1/\sigma_2 > 0$ dado, un estimador agrupado de la curtosis consistente para la curtosis común de la población en el modelo de dos muestras puede calcularse como

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2}$$

La demostración de este resultado se encuentra en el Apéndice A.

Tal como se esperaba, $\hat{\gamma}_P(1)$ es idéntico al estimador agrupado de la curtosis de Layard, ya que $\sigma_1/\sigma_2 = 1$ implica que las desviaciones estándar (o varianzas) son iguales.

El estadístico T'_2 , que es la versión de dos muestras del estadístico general T'_k , se calcula como

$$T'_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right) \tau^2}$$

donde $\hat{\rho} = S_1/S_2$, $\rho = \sigma_1/\sigma_2$ y $\tau^2 = 2 + (1 - 1/\bar{n})\hat{\gamma}_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$.

Como se indica en Layard (1973), en muestras grandes, $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$. Bonett (2006) usa una aproximación alternativa, que también se adopta en Shoemaker (2003), $\text{Var}((n_i - 1)^{1/2} \ln S_i^2) \cong \gamma - (n_i - 3)/n_i$. En las muestras grandes, estas aproximaciones son equivalentes. Sin embargo, Shoemaker indica que la última versión es ventajosa cuando su prueba de igualdad de varianzas se usa con muestras pequeñas. Usando este ajuste, el estadístico T_2' puede modificarse de la siguiente manera

$$T_2' = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\gamma - g_1}{n_1 - 1} + \frac{\gamma - g_2}{n_2 - 1}}$$

donde $g_i = (n_i - 3)/n_i$.

Se deduce entonces que el estadístico de prueba $T_2 = \tau^2 T_2' / \hat{\tau}^2$ para probar la hipótesis nula $H_0: \rho = \rho_0$ se calcula como

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{\gamma}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho_0) - g_2}{n_2 - 1}}$$

En esta expresión de T_2 , la raíz cuadrada del denominador puede verse como una estimación en muestras grandes del error estándar de la curtosis agrupada.

Por otro lado, en la expresión de $\hat{\gamma}_P(1) \equiv \hat{\gamma}$, Bonett (2006) usa las medias recortadas de las muestras con la proporción de recorte $1/[2(n_i - 4)^{1/2}]$. En consecuencia, hacemos el mismo ajuste al estimador agrupado de la curtosis:

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - m_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2}$$

donde m_i es la media recortada de la muestra i , con la proporción de recorte $1/[2(n_i - 4)^{1/2}]$. Esta versión del estimador agrupado de la curtosis y la versión anterior son asintóticamente equivalentes, ya que la media recortada m_i es un estimador consistente de la media de la población μ_i . Sin embargo, esta versión alternativa puede mejorar el rendimiento de la prueba basada en T_2 con muestras pequeñas.

El estadístico de prueba T_2 ahora puede invertirse para derivar un IC aproximado para la relación de las varianzas o desviaciones estándar. Pero antes describiremos brevemente el paso en falso en la derivación de los IC de Bonett (2006) para la relación de las desviaciones estándar.

3.1 Intervalos de Bonett

En lugar de invertir T_2 para obtener el IC, Bonett (2006) invierte el siguiente estadístico

$$T = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\hat{\gamma}_P(1) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(1) - g_2}{n_2 - 1}}$$

En consecuencia, el intervalo resultante es simplemente la región de aceptación para la prueba de la igualdad de varianzas. Esto se debe a que el estimador agrupado de la curtosis $\hat{\rho}_p(1)$ solo es consistente cuando las varianzas son iguales, o equivalentemente cuando la relación hipotética es 1. El intervalo resultante se especifica en Bonett (2006) como

$$\exp[\ln(c S_1^2/S_2^2) \pm z_{\alpha/2} se]$$

donde

$$se^2 = \frac{\hat{\gamma}(1) - g_1}{n_1 - 1} + \frac{\hat{\gamma}(1) - g_2}{n_2 - 1}$$

La constante c se incluye como un ajuste en muestras pequeñas para mitigar el efecto de las probabilidades de error de colas desiguales en diseños no balanceados. Esta constante se calcula como

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

La constante desaparece cuando los diseños son balanceados y su efecto se vuelve insignificante con el aumento de los tamaños de las muestras.

La tabla 1 ilustra las consecuencias de interpretar erróneamente los intervalos anteriores como intervalos de confianza. Estos resultados se basan en un pequeño estudio de simulación en el que calculamos las probabilidades de cobertura simulada con base en los intervalos de Bonett (2006). Para los casos de varianzas iguales (columna izquierda), extraemos dos muestras independientes de la distribución normal estándar. Para los casos de varianzas desiguales (columna derecha), escalamos las observaciones de la segunda muestra por un factor constante de 4. Las probabilidades de cobertura estimada se basan en 100,000 réplicas. La cobertura nominal objetivo es 0.95.

Tabla 1 Efecto de las varianzas desiguales de la población en los IC ($\alpha = 0.05$) de Bonett (2006)

n_1, n_2	Probabilidades de cobertura simulada	
	Varianzas iguales	Varianzas desiguales
10, 10	0.963	0.972
50, 50	0.952	0.991
100, 100	0.952	0.994

Si los intervalos se basaran en un estimador agrupado de la curtosis consistente, entonces se podría esperar que las probabilidades de cobertura en los dos casos fueran idénticas. Sin embargo, observe que los intervalos son consistentemente más conservadores cuando las varianzas son desiguales. Además, las probabilidades de cobertura se acercan a 1 a medida que aumentan los tamaños de las muestras. Tenga en cuenta que con los IC aproximados de Shoemaker (2003) se obtienen resultados similares.

3.2 Cálculos del IC

Considere el problema de probar la hipótesis nula $H_0: \rho = \rho_0$ con respecto a la hipótesis alternativa $H_A: \rho \neq \rho_0$, donde $\rho = \sigma_1/\sigma_2$ y $\rho_0 > 0$, con base en el estadístico de prueba T_2 calculado anteriormente. Bajo la hipótesis nula, el estadístico de prueba

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1}}$$

se distribuye asintóticamente como una distribución de chi-cuadrado con 1 grado de libertad. Por lo tanto, la prueba rechaza la hipótesis nula en el nivel de significancia α si y solo si

$$(\ln \hat{\rho}^2 - \ln \rho_0^2)^2 > z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

donde z_α denota el punto percentil $\alpha \times 100$ superior de la distribución normal estándar. Tenga en cuenta que el punto percentil $\alpha \times 100$ superior de la distribución de chi-cuadrado con 1 grado de libertad, $\chi_{1,\alpha}^2$, satisface la siguiente condición: $\chi_{1,\alpha}^2 = z_{\alpha/2}^2$.

Los resultados de la simulación de Bonett (2006) revelan que el ajuste en muestras pequeñas para reducir el efecto de las probabilidades de error de colas desiguales en los diseños no balanceados funcionó adecuadamente. Por lo tanto, hacemos un ajuste similar para la prueba basada en T_2 . Cuando se realiza este ajuste, la prueba rechaza la hipótesis nula si y solo si

$$(\ln \rho_0^2 - \ln(c\hat{\rho}^2))^2 > z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

donde c es la constante de ajuste de Bonett calculada como

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

De modo equivalente, un porcentaje de confianza aproximado de $(1 - \alpha)100$ establecido para $\rho = \sigma_1/\sigma_2$ con base en T_2 se calcula como

$$\left\{ \rho \in (0, \infty): (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho) - g_2}{n_2 - 1} \right) \leq 0 \right\}$$

Tenga en cuenta que c no tiene ningún efecto en los diseños balanceados y solo tiene un efecto insignificante en diseños no balanceados con muestras grandes.

El siguiente resultado proporciona una expresión alternativa de la confianza definida en una forma que resulta conveniente para describir su naturaleza. En esta expresión, el estimador agrupado de la curtosis se reescribe en términos de las curtosis individuales de las muestras al calcularse de la siguiente manera

$$\hat{Y}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1, 2$$

RESULTADO 2

Un porcentaje de confianza aproximado de $(1 - \alpha)100$ establecido para $\rho = \sigma_1/\sigma_2$ con base en T_2 puede expresarse como

$$\hat{\rho}\sqrt{c} \{r \in (0, \infty): H(r^2) \leq 0\}$$

o, de modo equivalente, la confianza establecida para $\rho^2 = \sigma_1^2/\sigma_2^2$ puede expresarse como

$$c\hat{\rho}^2 \{r \in (0, \infty): H(r) \leq 0\}$$

donde

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 se^2(cx), x > 0$$

$$se^2(x) = A \frac{\hat{y}_1 K^2/n_1 + \hat{y}_2 x^2/n_2}{(K + x)^2} - B$$

$$A = \frac{(n_1 + n_2)(n_1 + n_2 - 2)}{(n_1 - 1)(n_2 - 1)}, B = \frac{g_1}{n_1 - 1} + \frac{g_2}{n_2 - 1}, K = \frac{n_1 - 1}{n_2 - 1}$$

Para la comprobación de este resultado, consulte el Apéndice B.

Fácilmente puede verificarse que la función $H(x)$ es continua en la línea real positiva, con $H(0) = H(+\infty) = +\infty$ y $H(1) < 0$. Por lo tanto, según el teorema de los valores intermedios, la función $H(x)$ admite al menos una raíz en el intervalo $(0, 1)$ y al menos una raíz en el intervalo $(0, +\infty)$.

El siguiente resultado describe la confianza como un intervalo o la unión de intervalos separados.

RESULTADO 3

Si la función $H(x)$ tiene exactamente dos raíces, x_L y x_U , entonces $0 < x_L < 1 < x_U$ y la confianza establecida para $\rho^2 = \sigma_1^2/\sigma_2^2$ es el intervalo que se calcula como

$$[c\hat{\rho}^2 x_L, c\hat{\rho}^2 x_U]$$

Se deduce entonces que el IC para $\rho = \sigma_1/\sigma_2$ es el intervalo,

$$[\hat{\rho}\sqrt{cx_L}, \hat{\rho}\sqrt{cx_U}]$$

Por otro lado, si la función $H(x)$ tiene más de dos raíces, entonces la confianza establecida para $\rho^2 = \sigma_1^2/\sigma_2^2$ es la unión de intervalos no superpuestos. Las cotas de cada intervalo son las raíces consecutivas donde la función se abre hacia arriba.

Para la demostración de este resultado, consulte el Apéndice C.

OBSERVACIÓN

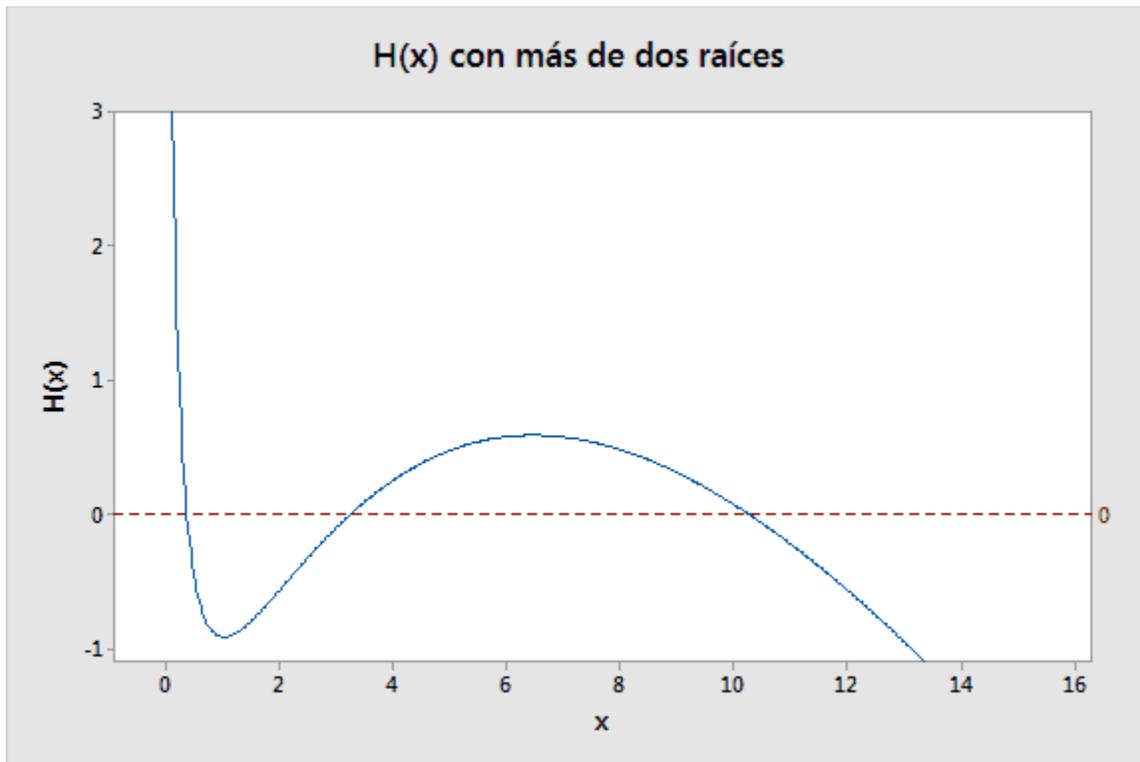
Aunque es matemáticamente posible que la función $H(x)$ admita más de dos raíces, pudimos observar que esto solo ocurre con diseños extremadamente poco comunes y prácticamente sin sentido en los que una o las dos muestras son demasiado pequeñas o están muy desbalanceadas. Conjeturamos que $H(x)$ tiene ya sea dos o cuatro raíces.

El siguiente ejemplo se basa en datos que fueron manipulados para obligar a la función $H(x)$ a tener más de dos raíces. Los datos se resumen de la siguiente manera: $n_1 = 169$, $n_2 = 7$, $S_1 = 301.855$, $S_2 = 4606.170$, $\hat{\gamma}_1 = 1.877$, $\hat{\gamma}_2 = 6.761$, $c = 0.728$, $A = 30.381$, $B = 0.101$ y $K = 28.000$.

Para $\alpha = 0.05$, la función $H(x)$ se calcula como

$$H(x) = (\ln x)^2 - 1.960^2 \left(30.381 \frac{1.877 \times 28^2 / 169 + 6.761 \times (.728x)^2 / 7}{(28.000 + 0.728x)^2} - 0.101 \right)$$

La función $H(x)$ en este caso tiene cuatro raíces. La gráfica de la función se muestra abajo. Tenga en cuenta que la cuarta raíz no se ve en la gráfica porque es demasiado grande. Sin embargo, sabemos que la cuarta raíz existe porque $H(+\infty) = +\infty$.



Las cuatro raíces se calculan numéricamente como $x_1 = 0.389$, $x_2 = 3.282$, $x_3 = 10.194$ y $x_4 = 39685.0$. La relación estimada de las desviaciones estándar es $\hat{\rho} = S_1/S_2 = 0.066$. La confianza establecida para $\rho^2 = \sigma_1^2/\sigma_2^2$ puede expresarse como

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] \cup [c \hat{\rho}^2 x_3, c \hat{\rho}^2 x_4] = [0.001, 0.010] \cup [0.032, 124.072]$$

La confianza establecida para la relación de las desviaciones estándar, ρ , se obtiene sacando la raíz cuadrada de las cotas de los intervalos.

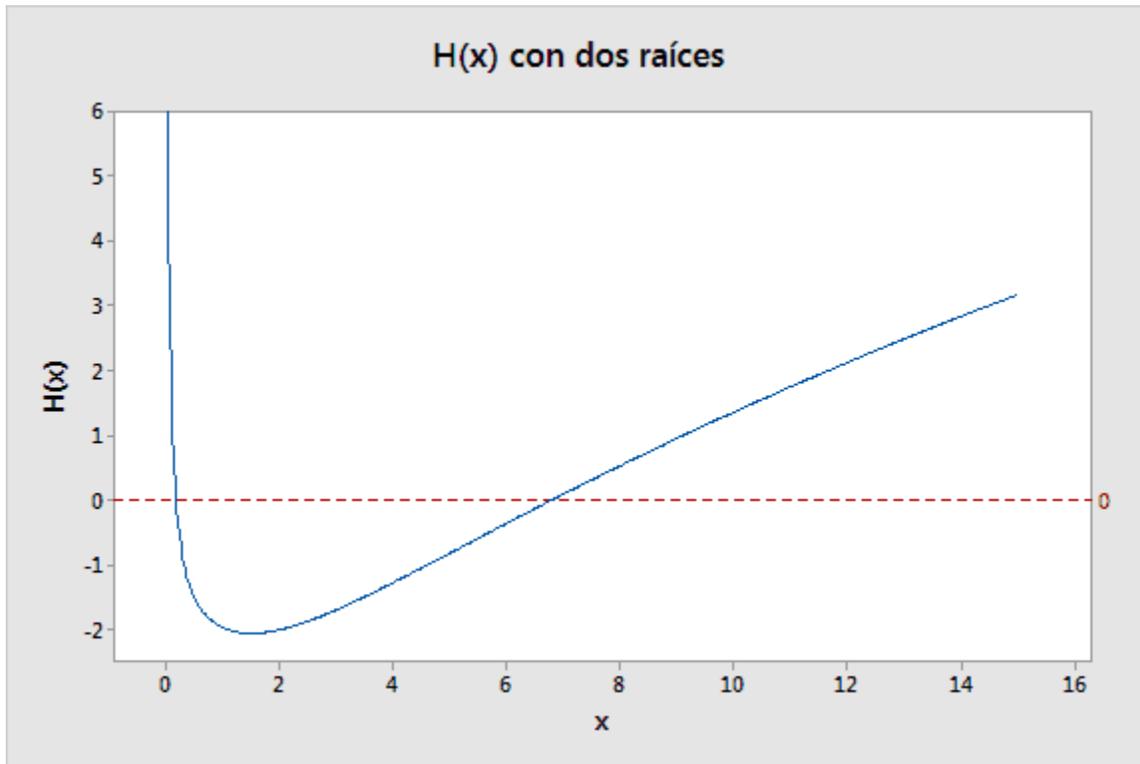
Cuando las muestras no son demasiado pequeñas ($n_i \geq 10$) y la disparidad de sus tamaños no es grande, la función $H(x)$ por lo general admite dos raíces. Una de las raíces está por debajo de la unidad y la otra raíz está por encima de la unidad, como se describe en el resultado 2. El siguiente es un ejemplo que se basa en datos generados de manera aleatoria. Los datos pueden

resumirse de la siguiente manera: $n_1 = 10$, $n_2 = 12$, $S_1 = 1.150$, $S_2 = 1.043$, $\hat{\gamma}_1 = 2.704$, $\hat{\gamma}_2 = 3.671$, $c = 1.041$, $A = 4.444$, $B = 0.146$ y $K = 0.818$.

Para $\alpha = 0.05$, la función $H(x)$ se calcula en este caso como

$$H(x) = (\ln x)^2 - 1.960^2 \left(4.444 \frac{2.704 \times 0.818^2 / 10 + 3.671 \times (1.041x)^2 / 12}{(0.818 + 1.041x)^2} - 0.146 \right)$$

La función $H(x)$ tiene dos raíces como se muestra a continuación:



Las dos raíces se calculan numéricamente como $x_1 = 0.200$ y $x_2 = 6.824$. La relación estimada de las desviaciones estándar es $\hat{\rho} = S_1/S_2 = 1.102$. La confianza establecida para $\rho^2 = \sigma_1^2/\sigma_2^2$ es el intervalo calculado como

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] = [0.253, 8.634]$$

El IC para la relación de las desviaciones estándar, ρ , se obtiene sacando la raíz cuadrada de las cotas del intervalo anterior.

Ahora describiremos dos algoritmos para hallar los límites de confianza.

El primer algoritmo consiste en utilizar un procedimiento que busca raíces numéricas para hallar las raíces de la función $H(x)$. La raíz que corresponde a un límite de confianza inferior para la relación de las varianzas está dentro del intervalo $(0, 1)$. Si denotamos esta raíz con x_L , entonces, según el resultado 3, el límite de confianza inferior para la relación de las varianzas se calcula como $c \hat{\rho}^2 x_L$, y el límite de confianza inferior para la relación de la desviación estándar se obtiene como $\hat{\rho} \sqrt{c x_L}$. De manera similar, el límite de confianza superior para la relación de las varianzas es $c \hat{\rho}^2 x_U$ y el límite de confianza superior para la relación de las desviaciones estándar

es $\hat{\rho}\sqrt{cx_U}$, donde $x_U > 1$ es la otra raíz de $H(x)$. Un enfoque sencillo para hallar el límite de confianza superior es aprovechar el hecho de que el límite inferior para $1/\rho^2$ es el límite superior para ρ^2 . En primer lugar, los roles de la primera muestra y la segunda muestra se intercambian en la expresión de la función $H(x)$ como si se estuviera calculando el límite de confianza para la relación $1/\rho^2 = \sigma_2^2/\sigma_1^2$. En segundo lugar, el algoritmo para hallar el límite inferior se aplica a la nueva función $H(x)$. Finalmente, el límite resultante se invierte para obtener el límite de confianza superior deseado.

Un enfoque alternativo consiste en calcular recursivamente el límite de confianza inferior para la relación de las varianzas utilizando la relación de recurrencia calculada como

$$\rho_0^2 = 1$$

$$\rho_{i+1}^2 = \exp \left[\ln(c \hat{\rho}^2) - z_{\alpha/2} \sqrt{\frac{\hat{Y}_P(\rho_i) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_i) - g_2}{n_2 - 1}} \right], i = 0, 1, 2, \dots$$

El límite de confianza inferior para la relación de las varianzas es ρ_{j+1}^2 , de modo que $|\rho_{j+1}^2 - \rho_j^2| < \varepsilon$, donde $j > 0$ y ε se escoge de manera que sea pequeño (por ejemplo $\varepsilon = 10^{-6}$). Para hallar el límite de confianza superior, simplemente reemplazamos $-z_{\alpha/2}$ por $+z_{\alpha/2}$ en la expresión anterior.

Evidentemente, los dos algoritmos para calcular los límites de confianza son equivalentes, porque el procedimiento recursivo es en esencia un procedimiento iterativo para resolver la ecuación $H(\rho^2/(c\hat{\rho}^2)) = 0$ para ρ^2 . El algoritmo recursivo es más fácil de implementar y, por lo tanto, es una alternativa útil cuando no está disponible un procedimiento para hallar raíces.

4. Estudios de simulación y resultados

En este trabajo, derivamos un procedimiento para extender la prueba de Layard para la igualdad de dos varianzas con el fin de probar la relación de varianzas. A este procedimiento nos referimos como la prueba extendida de Layard para la relación (ELTR, por sus siglas en inglés). En esta sección, investigamos las propiedades en muestras pequeñas de los IC basados en el procedimiento ELTR. Seguimos el enfoque general adoptado por Bonett (2006).

Comparamos los IC que se basan en el procedimiento ELTR con IC que se basan en la prueba L_{50} (Pan, 1999) y la prueba W_{50} (la prueba de Levene/Brown-Forsythe). Para el estudio 1, incluimos también, con fines comparativos, los IC que se basan en la prueba F clásica. Es bien sabido que, cuando los datos se distribuyen normalmente, la prueba F clásica es óptima. Tenga en cuenta que los cálculos de los IC basados en las pruebas W_{50} y L_{50} se encuentran en Pan (1999). Los cálculos de los IC basados en la prueba F pueden encontrarse en muchos libros de texto de introducción a la Estadística. También están disponibles en Bonett (2006).

Realizamos tres estudios de simulación, cada uno con 100,000 réplicas de muestra. Cada réplica se compone de dos muestras independientes con un tamaño de moderado a pequeño. Cada

muestra se extrajo de una distribución original con propiedades conocidas, incluyendo simetría, asimetría, colas pesadas y colas livianas. El error estándar asociado con cada simulación es aproximadamente 0.0009, 0.0007 y 0.0003 para los niveles de confianza nominales de 90%, 95% y 99%, respectivamente.

Para evaluar el desempeño de cada procedimiento, registramos la probabilidad de cobertura alcanzada y la anchura media de los intervalos simulados para la relación de las varianzas. Algunos de los intervalos asociados con la prueba W_{50} tenían anchuras infinitas (una posibilidad expuesta por Pan (1999)). En tales casos, registramos tanto la anchura media de los intervalos finitos como el porcentaje de intervalos con anchura infinita. Todas las simulaciones se realizaron utilizando la Versión 8 del paquete de software Mathematica.

Estudio 1: Comparación de las probabilidades de cobertura para datos normales

En el primer estudio, generamos muestras aleatorias de diferentes tamaños de la distribución normal. Los resultados se presentan en la tabla 2.

Tabla 2 Comparación de probabilidades de cobertura y anchuras promedio de los intervalos

$1 - \alpha$	n_1, n_2	Medida	Procedimiento			
			F	ELTR	L_{50}	W_{50}
0.90	10, 10	Cobertura	0.898	0.918	0.913	0.921
		<i>Anchura</i>	<i>3.72</i>	<i>5.06</i>	<i>4.72</i>	<i>8.03</i> <i>(0.01%)</i>
	30, 10	Cobertura	0.900	0.909	0.897	0.911
		<i>Anchura</i>	<i>2.42</i>	<i>3.01</i>	<i>3.58</i>	<i>3.17</i>
	25, 25	Cobertura	0.902	0.907	0.914	0.916
		<i>Anchura</i>	<i>1.61</i>	<i>1.73</i>	<i>1.85</i>	<i>1.938</i>
	50, 50	Cobertura	0.900	0.901	0.906	0.907
		<i>Anchura</i>	<i>1.03</i>	<i>1.06</i>	<i>1.13</i>	<i>1.15</i>
0.95	10, 10	Cobertura	0.949	0.963	0.958	0.964
		<i>Anchura</i>	<i>4.90</i>	<i>7.72</i>	<i>6.52</i>	<i>497.24</i> <i>(0.20%)</i>
	30, 10	Cobertura	0.950	0.957	0.945	0.959
		<i>Anchura</i>	<i>2.98</i>	<i>4.91</i>	<i>4.67</i>	<i>4.07</i>
	25, 25	Cobertura	0.951	0.955	0.958	0.961

$1 - \alpha$	n_1, n_2	Medida	Procedimiento			
			F	ELTR	L_{50}	W_{50}
		Anchura	1.99	2.24	2.31	2.49
	50, 50	Cobertura	0.951	0.952	0.953	0.954
		Anchura	1.25	1.31	1.38	1.41
0.99	10, 10	Cobertura	0.989	0.993	0.992	0.994
		Anchura	8.29	17.76	12.52	$> 10^4$ (8.8%)
	30, 10	Cobertura	0.990	0.992	0.986	0.994
		Anchura	4.26	15.76	8.26	6.77
	25, 25	Cobertura	0.990	0.992	0.992	0.993
		Anchura	2.86	3.66	3.43	4.03
	50, 50	Cobertura	0.990	0.991	0.991	0.991
		Anchura	1.71	1.89	1.92	2.02

Las filas sombreadas muestran las probabilidades de cobertura alcanzada (Cobertura) para cada procedimiento en cada nivel de confianza ($1 - \alpha$) y cada combinación de tamaños de muestra (n_1, n_2). La media de las anchuras de los intervalos (Anchura) se muestra debajo de cada probabilidad de cobertura. Si una condición tenía intervalos infinitos, registramos tanto la media de los intervalos finitos como el porcentaje de intervalos que eran infinitos.

Como se esperaba, los resultados muestran que los IC asociados con el procedimiento F son los más exactos y los más precisos. Las probabilidades de cobertura alcanzadas con el procedimiento F están más cerca de la cobertura objetivo que las asociadas con los otros procedimientos. Además, las anchuras promedio de los intervalos asociados con el procedimiento F son más pequeñas que las asociadas con los otros procedimientos. Sin embargo, la tabla también revela que los IC que se construyen utilizando los procedimientos ELTR y L_{50} son casi tan exactos y precisos como los basados en el procedimiento F.

Los intervalos basados en la prueba W_{50} también son bastante exactos. No obstante, los intervalos de W_{50} pueden ser muy amplios e incluso pueden tener anchura infinita, dependiendo del tamaño de las muestras. Tenga en cuenta que, cuando ambas muestras tienen solamente 10 observaciones, al menos el 0.01% de los intervalos producidos por el procedimiento W_{50} son infinitamente amplios. Asimismo, el porcentaje de intervalos infinitos aumenta cuando aumenta la cobertura objetivo. En la mayoría de las condiciones, las anchuras medias de los intervalos de ELTR y L_{50} son más pequeñas que las anchuras medias de los intervalos de W_{50} .

Estudio 2: Comparación de las probabilidades de cobertura para datos no normales

El segundo estudio está diseñado para evaluar y comparar el desempeño de los procedimientos ELTR, L_{50} y W_{50} cuando las distribuciones originales no son normales. Incluimos también una distribución normal contaminada con el fin de evaluar el impacto de los valores atípicos en el desempeño de los procedimientos. Denotamos esta distribución contaminada como CN(0.1, 3) para indicar que, aunque el 90% de las observaciones se extrajo de la distribución normal estándar, el 10% restante se extrajo de una población normal con una media de 0 y una desviación estándar de 3. Los resultados se presentan en la tabla 3.

Tabla 3 Comparación de probabilidades de cobertura y anchuras promedio de los intervalos en algunos modelos no normales cuando el nivel de confianza nominal es $1 - \alpha = 0.95$

Distribución n [γ] n_1, n_2	ELTR	L_{50}	W_{50}	Distribución n [γ] n_1, n_2	ELTR	L_{50}	W_{50}
Uniforme				$\chi^2(5)$			
[1.8]				[5.4]			
10, 10	0.971	0.971	0.966	10, 10	0.956	0.938	0.956
	<i>5.27</i>	<i>4.87</i>	<i>42.08</i> (0.1%)		<i>11.61</i>	<i>8.78</i>	$> 10^4$ (2.6%)
10, 30	0.964	0.961	0.957	10, 30	0.959	0.923	0.956
	<i>2.51</i>	<i>2.4</i>	<i>2.89</i>		<i>6.25</i>	<i>4.14</i>	<i>190.645</i> (0.3%)
25, 25	0.967	0.972	0.968	25, 25	0.956	0.944	0.954
	<i>1.43</i>	<i>1.79</i>	<i>1.88</i>		<i>3.66</i>	<i>2.92</i>	<i>3.26</i>
50, 50	0.959	0.962	0.959	50, 50	0.959	0.946	0.952
	<i>0.83</i>	<i>1.06</i>	<i>1.08</i>		<i>2.07</i>	<i>1.7</i>	<i>1.77</i>
Beta (3, 3)				Exponencia I			
[2.5]				[9]			
10, 10	0.968	0.966	0.966	10, 10	0.947	0.916	0.950
	<i>6.26</i>	<i>5.59</i>	<i>254.62</i> (0.1%)		<i>20.99</i>	<i>14.47</i>	$> 10^4$ (9.1%)
10, 30	0.960	0.954	0.960	10, 30	0.954	0.896	0.953

Distribució n [γ] n_1, n_2	ELTR	L_{50}	W_{50}	Distribució n [γ] n_1, n_2	ELTR	L_{50}	W_{50}
	3.14	2.76	3.71		10.46	6.19	$> 10^4$ (4.1%)
25, 25	0.959	0.966	0.965	25, 25	0.956	0.931	0.951
	1.81	2.06	2.18		6.09	4.13	5.48 (0.008%)
50, 50	0.957	0.959	0.958	50, 50	0.962	0.942	0.952
	1.06	1.23	1.26		3.18	2.24	2.38
Laplace				$\chi^2(1)$			
[6]				[15]			
10, 10	0.946	0.935	0.961	10, 10	0.928	0.889	0.947
	13.47	10.45	$> 10^4$ (3.0%)		55.09	37.4	$> 10^5$ (25.1%)
10, 30	0.947	0.919	0.957	10, 30	0.943	0.882	0.956
	6.78	4.82	$> 10^4$ (0.4%)		18.71	11.14	$> 10^6$ (25.7%)
25, 25	0.945	0.940	0.952	25, 25	0.952	0.925	0.954
	4.00	3.372	3.86		10.97	6.84	$> 10^4$ (0.4%)
50, 50	0.952	0.949	0.955	50, 50	0.958	0.936	0.951
	2.19	1.91	1.99		5.08	3.31	3.75 (0.001%)
t(5)				Lognormal			
[9]				[113.9]			
10, 10	0.957	0.946	0.965	10, 10	0.923	0.876	0.955
	11.07	8.81	$> 10^3$ (2.0%)		59.22	46.15	$> 10^5$ (23.0%)
10, 30	0.957	0.930	0.959	10, 30	0.949	0.866	0.958
	6.06	4.24	$> 10^3$ (0.7%)		29.13	17.67	$> 10^6$ (31.6%)
25, 25	0.954	0.948	0.960	25, 25	0.947	0.917	0.965

Distribución n [γ] n_1, n_2	ELTR	L_{50}	W_{50}	Distribución n [γ] n_1, n_2	ELTR	L_{50}	W_{50}
	3.54	2.93	4.86 (0.01%)		16.21	8.73	$> 10^4$ (2.4%)
50, 50	0.954	0.947	0.954	50, 50	0.955	0.928	0.960
	2.10	1.71	1.77 (0.003%)		8.62	4.11	164.38 (0.2%)
Normales (absolutos)				CN(0.1, 3)			
[3.9]				[8.3]			
10, 10	0.956	0.942	0.954	10, 10	0.977	0.965	0.979
	10.41	7.89	$> 10^4$ (1.5%)		12.64	9.52	$> 10^4$ (4.9%)
10, 30	0.959	0.930	0.954	10, 30	0.981	0.952	0.979
	5.18	3.64	13.00 (0.02%)		7.82	4.71	944.68 (1.1%)
25, 25	0.959	0.952	0.959	25, 25	0.982	0.972	0.981
	3.01	2.62	2.88		4.63	3.22	3.71
50, 50	0.960	0.951	0.954	50, 50	0.983	0.972	0.978
	1.69	1.54	1.59		2.64	1.83	1.91

Las filas sombreadas muestran las probabilidades de cobertura alcanzada para cada procedimiento, distribución original y combinación de tamaños de muestra. La media de las anchuras de los intervalos se muestra debajo de cada probabilidad de cobertura. Si una condición tenía intervalos infinitos, registramos tanto la media de los intervalos finitos como el porcentaje de intervalos que eran infinitos. La curtosis (γ) de cada distribución original se muestra entre corchetes.

Para las distribuciones simétricas de colas livianas, los resultados indican que los tres métodos producen de manera similar probabilidades de cobertura conservadoras. Sin embargo, los intervalos de ELTR y L_{50} son más precisos para muestras pequeñas que los intervalos de W_{50} . Por ejemplo, cuando las muestras provienen de una distribución Beta con parámetros de (3, 3), las probabilidades de cobertura alcanzada para los intervalos de ELTR y L_{50} son al menos tan exactas como las de los intervalos de W_{50} , pero los intervalos de W_{50} son consistentemente más amplios.

Los intervalos de ELTR y W_{50} también son un poco conservadores para las distribuciones simétricas de colas pesadas, mientras que los intervalos de L_{50} son liberales. Los intervalos de

L_{50} son incluso más liberales cuando los diseños no son balanceados. Por ejemplo, cuando se extraen muestras con tamaños de 10 y 30 de la distribución de Laplace, la probabilidad de cobertura alcanzada para los intervalos de L_{50} es 0.919. Además, cuando se extraen muestras con los mismos tamaños de una distribución t con 5 grados de libertad, la probabilidad de cobertura alcanzada para los intervalos de L_{50} es 0.930.

Los intervalos de L_{50} también son bastante liberales cuando se extraen muestras pequeñas de distribuciones de colas pesadas muy asimétricas. Por ejemplo, cuando las muestras provienen de una distribución lognormal, la cobertura alcanzada puede ser tan baja como 0.866. Para estas distribuciones, el método W_{50} es el menos liberal de los tres métodos. Sin embargo, son demasiados los intervalos de W_{50} que tienen anchura infinita. Por ejemplo, cuando se extraen muestras de la distribución de chi-cuadrado con 1 grado de libertad ($\chi^2(1)$), más del 25% de los intervalos de W_{50} puede tener anchura infinita. Los intervalos de ELTR son relativamente menos exactos, pero considerablemente más estrechos y, por lo tanto, más informativos que los intervalos de W_{50} .

Por último, observamos que los tres procedimientos son afectados negativamente por los valores atípicos. El método L_{50} es el menos afectado, lo que podría esperarse porque el método L_{50} se derivó para reducir el efecto de los valores atípicos en la prueba W_{50} (Pan, 1999). Cuando se extraen muestras de la distribución normal contaminada, NC (0.1, 3), el mínimo de las probabilidades de cobertura alcanzada para los procedimientos ELTR y W_{50} es 0.977. Resultados adicionales de la simulación (no se muestran) indican que estos intervalos mejoran, aunque lentamente, con el aumento de los tamaños de las muestras.

Estudio 3: Sensibilidad al supuesto de igualdad de curtosis

Nuestro estudio final investiga la sensibilidad del procedimiento ELTR al supuesto de igualdad de curtosis bajo el cual se deriva. Examinamos el desempeño del procedimiento ELTR cuando las curtosis de las poblaciones originales no son iguales, es decir, cuando $\gamma_1 \neq \gamma_2$. También incluimos los procedimientos L_{50} y W_{50} , porque se derivan bajo el supuesto de que las poblaciones son similares. Este supuesto de similitud no está determinado cuando las curtosis de las poblaciones originales no son iguales. Los resultados se presentan en la tabla 4.

Tabla 4 Sensibilidad del procedimiento ELTR al supuesto de igualdad de curtosis cuando el nivel de confianza nominal es $1 - \alpha = 0.95$

Dist. 1, Dist. 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}	Dist. 1, Dist. 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}
Beta (3, 3), Normal [2.5, 3]				Normal, CN (0.9, 3) [3, 8.3]			

Dist. 1, Dist. 2 $[\gamma_1, \gamma_2]$ n_1, n_2	ELTR	L_{50}	W_{50}	Dist. 1, Dist. 2 $[\gamma_1, \gamma_2]$ n_1, n_2	ELTR	L_{50}	W_{50}
10, 10	0.964	0.961	0.964	10, 10	0.955	0.948	0.951
	0.27	0.23	204.50 (0.20%)		6.88	5.16	$> 10^4$ (4.89%)
30, 10	0.946	0.939	0.946	30, 10	0.941	0.910	0.942
	0.16	0.17	0.15		5.26	3.77	3.20
10, 30	0.966	0.956	0.967	10, 30	0.961	0.950	0.958
	0.14	0.11	0.17		4.26	2.40	630.42 (1.10%)
50, 50	0.951	0.950	0.949	50, 50	0.936	0.910	0.907
	0.04	0.05	0.05		1.27	1.11	1.19
Normal, Laplace [3, 6]				Normales (no negativos), $\chi^2(5)$ [3.9, 5.4]			
10, 10	0.941	0.935	0.947	10, 10	0.956	0.940	0.954
	6.67	5.17	$> 10^6$ (2.90%)		0.42	0.32	304.41 (2.60%)
30, 10	0.912	0.888	0.914	30, 10	0.954	0.918	0.949
	5.06	3.85	3.21		0.33	0.22	0.20
10, 30	0.963	0.943	0.955	10, 30	0.962	0.934	0.958
	3.33	2.25	$> 10^3$ (0.40%)		0.23	0.15	3.28 (0.30%)
50, 50	0.935	0.894	0.889	50, 50	0.955	0.941	0.945
	0.98	1.04	1.12		0.07	0.06	0.07
Normal, Normales (no negativos) [3, 3.9]				$\chi^2(5)$, Exponencial [5.4, 9]			
10, 10	0.956	0.948	0.957	10, 10	0.938	0.914	0.940
	28.16	20.65	$> 10^4$ (1.50%)		211.17	137.88	$> 10^6$ (9.10%)

Dist. 1, Dist. 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}	Dist. 1, Dist. 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}
30, 10	0.946	0.924	0.947	30, 10	0.928	0.875	0.929
	20.59	14.83	12.78		194.70	93.02	83.02
10, 30	0.961	0.946	0.962	10, 30	0.968	0.930	0.954
	14.06	9.37	49.11 (0.02%)		102.35	55.29	> 10 ⁵ (3.90%)
50, 50	0.953	0.950	0.952	50, 50	0.950	0.920	0.923
	4.32	4.16	4.33		29.64	23.37	25.54

Las filas sombreadas muestran las probabilidades de cobertura alcanzada para cada procedimiento, combinación de distribuciones originales (Dist. 1, Dist. 2) y combinación de tamaños de muestra. La media de las anchuras de los intervalos se muestra debajo de cada probabilidad de cobertura. Si una condición tiene intervalos infinitos, registramos tanto la media de los intervalos finitos como el porcentaje de intervalos que eran infinitos. La curtosis de cada distribución original (γ_1, γ_2) se muestra entre corchetes.

En general, el desempeño del procedimiento ELTR no parece verse afectado negativamente por las curtosis desiguales cuando las muestras son lo suficientemente grandes. Sin embargo, cuando los diseños no son balanceados y la muestra más pequeña se obtiene de la distribución con colas más pesadas, las probabilidades de cobertura alcanzada son liberales. Las probabilidades de cobertura alcanzada son mejores cuando la muestra más grande se extrae de la distribución con colas más pesadas.

Cuando los tamaños de las muestras son lo suficientemente grandes, los intervalos de L_{50} y W_{50} por lo general también parecen ser robustos ante la desigualdad de las distribuciones provocada por las curtosis desiguales. Sin embargo, tenga en cuenta que cuando las muestras se extraen de una distribución normal y una distribución de Laplace o de una distribución normal y una distribución normal contaminada, las probabilidades de cobertura para los intervalos de L_{50} y W_{50} no son estables, incluso para muestras tan grandes como 50.

Los intervalos de L_{50} son generalmente más liberales que los intervalos de ELTR y W_{50} . En tres casos, las probabilidades de cobertura alcanzada para los intervalos de L_{50} son menores que 0.90. En cambio, solo una de las probabilidades de cobertura alcanzada para los intervalos de W_{50} está por debajo de 0.90. La menor probabilidad de cobertura alcanzada para los intervalos de ELTR es 0.912.

El estudio anterior (tabla 3) muestra que los tres procedimientos producen intervalos que son notablemente más conservadores cuando ambas muestras provienen de la distribución normal contaminada, CN(0.1, 3). El presente estudio revela que los tres procedimientos funcionan mucho mejor cuando solo una de las muestras se extrae de CN (0.1, 3). Sin embargo, tenga en

cuenta que el funcionamiento de los intervalos de L_{50} y W_{50} parece desmejorar considerablemente cuando el tamaño de la muestra aumenta a 50.

5. Ejemplo

En esta sección, aplicamos los cuatro procedimientos—F, ELTR, L_{50} y W_{50} —a un conjunto de datos tomado de Pan (1999). Ott (1993, página 352) describe los datos de la siguiente manera:

Un químico que trabaja en una mina de hierro sospecha que la varianza en la cantidad (peso, en onzas) de óxido de hierro por libra de mineral tiende a aumentar a mitad que aumenta la cantidad media de óxido de hierro por libra. Para comprobar esta teoría, diez muestras de 1 libra de mineral de hierro se seleccionan en dos lugares diferentes: ubicación 1, que presenta un contenido medio de óxido de hierro mucho mayor que el de la otra, la ubicación 2. A continuación se muestran las cantidades de óxido de hierro contenidas en las muestras de mineral:

Ubicación 1	8.1	7.4	9.3	7.5	7.1	8.7	9.1	7.9	8.4	8.8
Ubicación 2	3.9	4.4	4.7	3.6	4.1	3.9	4.6	3.5	4.0	4.2

Los IC de 95% para $\sigma_2/\sigma_1 = 1/\rho$, calculados utilizando los cuatro métodos diferentes, se indican en la siguiente tabla:

Procedimiento	IC 95%
F	(0.262, 1.055)
ELTR	(0.277, 0.924)
L_{50} (Pan)	(0.295, 0.938)
W_{50} (Levene/Brown-Forsythe)	(0.237, 0.908)

6. Conclusión

Nuestras simulaciones revelan que, en general, los IC basados en el procedimiento ELTR son tan exactos como los IC derivados de las pruebas L_{50} y W_{50} . Sin embargo, los intervalos de ELTR y los intervalos de L_{50} son más precisos que los intervalos de W_{50} para la mayoría de las distribuciones. Los intervalos de W_{50} tienden a ser más exactos que los intervalos de ELTR y los intervalos de L_{50} cuando se extraen muestras pequeñas de distribuciones de colas pesadas muy asimétricas. Sin embargo, esta ventaja por lo general es compensada por una notable pérdida de precisión. Los intervalos de W_{50} resultantes suelen ser demasiado amplios y es probable que tengan anchura infinita.

Por su diseño, los intervalos de L_{50} mejoran la precisión de los intervalos de W_{50} . No obstante, en el caso de las poblaciones asimétricas, los intervalos de L_{50} son tan cortos que producen probabilidades de cobertura excesivamente liberales. Por el contrario, los intervalos de ELTR son

más estables en general. Los intervalos de ELTR no suelen ser demasiado largos ni demasiado cortos, así que las probabilidades de cobertura por lo general no son demasiado conservadoras ni demasiado liberales. Por lo tanto, el procedimiento ELTR parece ser el mejor procedimiento para la mayoría de los efectos prácticos.

Los intervalos de ELTR son un poco más difíciles de calcular que los intervalos basados en la prueba L_{50} o la prueba W_{50} . Sin embargo, en general, la mayor precisión (comparada con la de los intervalos de W_{50}) y la mayor exactitud (comparada con la de los intervalos de L_{50}) compensan de sobra el esfuerzo de cálculo extra. El procedimiento ELTR se ha implementado como parte del análisis Varianza de dos muestras en la versión 17 de Minitab Statistical Software, donde se menciona como procedimiento de Bonett.

Para investigaciones futuras, se podría considerar investigar las propiedades de la prueba de Layard con muestras pequeñas en diseños de múltiples muestras cuando el estimador agrupado de la curtosis de Layard se reemplaza por la versión más robusta de Bonett, calculado como

$$\hat{\gamma}_A = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{\left[\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^2 \right]^2} \sum_{i=1}^k n_i$$

donde m_i es la media recortada de la muestra i , con la proporción de recorte $1/[2(n_i - 4)^{1/2}]$, y $i = 1, \dots, k$.

Además, podría ser beneficioso usar la aproximación de Shoemaker de la varianza asintótica de la varianza de la muestra con transformación logarítmica.

Por último, observamos que los intervalos propuestos por Bonett (2006), aunque no son adecuados como intervalos de confianza, sí son notablemente exactos y precisos para la mayoría de las distribuciones cuando se interpretan como regiones de aceptación para la prueba de la igualdad de dos varianzas. Estas regiones de aceptación son muy adecuadas para usarlas como la base de un procedimiento gráfico para comparar múltiples varianzas. Hochberg, Weiss y Hart (1982) propusieron un procedimiento similar para probar la igualdad de las medias. Ese procedimiento se utilizó como parte del análisis Prueba de varianzas iguales en la versión 17 de Minitab Statistical Software, donde se menciona como el procedimiento Comparaciones múltiples.

7. Apéndices

Apéndice A: Comprobación del resultado 1

Usando las notaciones y los supuestos de la sección 2, sea $X_j = \rho Y_{2j}$ para un $\rho = \sigma_1/\sigma_2$ dado. Entonces

$$\text{Var}(X_j) = \rho^2 \text{Var}(Y_{2j}) = \rho^2 \sigma_2^2 = \sigma_1^2 = \text{Var}(Y_{1j})$$

y

$$E\left(X_j - \mu_{X_j}\right)^4 / \sigma_{X_j}^4 = \rho^4 E(Y_{2j} - \mu_2) / (\rho^4 \sigma_2^4) = E(Y_{2j} - \mu_2) / \sigma_2^4 = \gamma$$

Puesto que $E(Y_{1j} - \mu_1)^4 / \sigma_1^4 = \gamma$ según los supuestos, se deduce que las poblaciones originales de las dos muestras Y_{1j} y $X_j = \rho Y_{2j}$ tienen la misma varianza σ_1^2 y la misma curtosis γ . Según Layard (1973), un estimador agrupado de la curtosis consistente para γ con base en las dos muestras Y_{1j} y X_j se obtiene como

$$\hat{\gamma}' = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \sum_{j=1}^{n_2} (X_j - \bar{X})^4}{[(n_1 - 1)S_1^2 + (n_2 - 1)S_X^2]^2} = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = \hat{\gamma}_P(\rho)$$

según se requiera.

Apéndice B: Comprobación del resultado 2

Ya establecimos que un porcentaje de confianza aproximado de $(1 - \alpha)100$ establecido para $\rho = \sigma_1/\sigma_2$ con base en T_2 se calcula como

$$\left\{ \rho \in (0, \infty): (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \leq 0 \right\}$$

El estimador agrupado de la curtosis puede expresarse en términos de los estimadores de curtosis correspondientes a las muestras individuales, y se calcula como

$$\hat{\gamma}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1, 2$$

Más específicamente, si suponemos que $t = \rho/\hat{\rho}$, entonces

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - m_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = (n_1 + n_2) \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2}$$

donde $K = (n_1 - 1)/(n_2 - 1)$.

En consecuencia, el término de error estándar al cuadrado puede expresarse como

$$\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2} - B$$

donde

$$A = \frac{(n_1 + n_2)(n_1 + n_2 - 2)}{(n_1 - 1)(n_2 - 1)}, B = \frac{k_1}{n_1 - 1} + \frac{k_2}{n_2 - 1}$$

Por lo tanto, si suponemos que $r^2 = \rho^2 / (c\hat{\rho}^2)$, entonces puede verse fácilmente que

$$\begin{aligned} (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \\ = (\ln r^2)^2 - z_{\alpha/2}^2 \left(A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 c^2 r^4/n_2}{(K + c r^2)^2} - B \right) \end{aligned}$$

Se deduce que un porcentaje de confianza aproximado de $(1 - \alpha)100$ establecido para $\rho = \sigma_1/\sigma_2$ con base en T_2 puede calcularse como

$$\hat{\rho}\sqrt{c} \{r \in (0, \infty): H(r^2) \leq 0\}$$

o, de modo equivalente, la confianza establecida para $\rho^2 = \sigma_1^2/\sigma_2^2$ puede expresarse como

$$c\hat{\rho}^2 \{r \in (0, \infty): H(r) \leq 0\}$$

donde

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 se^2(cx), x > 0$$

y

$$se^2(x) = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 x^2/n_2}{(K + x)^2} - B$$

Apéndice C: Comprobación del resultado 3

Fácilmente puede verificarse que la $H(x)$ es continua en la línea real positiva, con $H(0) = H(+\infty) = +\infty$ y $H(1) < 0$. Según el teorema de los valores intermedios, la función $H(x)$ admite al menos una raíz en el intervalo $(0, 1)$ y al menos una raíz en el intervalo $(0, +\infty)$. Por consiguiente, si la función $H(x)$ tiene exactamente dos raíces, entonces una raíz está por debajo de 1 y la otra está por encima de 1. Puesto que esta función se abre hacia arriba, la desigualdad $H(r) \leq 0$ se cumple si r se encuentra entre las raíces. Estas raíces definen las cotas del IC para $\rho^2/(c\hat{\rho}^2)$. Por lo tanto, si suponemos que $x_L < 1 < x_U$ es las dos raíces, entonces, según el resultado 2, el límite de confianza inferior para la relación de las varianzas, ρ^2 , se calcula como $c\hat{\rho}^2 x_L$, y el límite de confianza inferior para la relación de la desviación estándar se obtiene como $\hat{\rho}\sqrt{cx_L}$. De manera similar, es el límite de confianza superior para la relación de las varianzas es $c\hat{\rho}^2 x_U$ y el límite de confianza superior para la relación de las desviaciones estándar es $\hat{\rho}\sqrt{cx_U}$.

Por otro lado, si la función $H(x)$ tiene más de dos raíces, entonces la desigualdad $H(r) \leq 0$ se cumple si r se encuentra entre raíces consecutivas donde la función se abre hacia arriba. Entonces, la confianza establecida es la unión de los intervalos que no se superponen.

8. Referencias

Balakrishnan, N. y Ma, C. W. (1990). A Comparative Study of Various Tests for the Equality of Two Population Variances. *Journal of Statistical Computation and Simulation*, 35, 41–89.

- Bonett D. G. (2006). Robust Confidence Interval for a Ratio of Standard Deviations. *Applied Psychological Measurements*, 30, 432–439.
- Boos, D. D. y Brownie, C. (1989). Bootstrap Methods for Testing Homogeneity of Variances. *Technometrics*, 31, 69–82.
- Brown, M. B. y Forsythe A. B. (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association*, 69, 364–367.
- Conover, W. J., Johnson, M. E. y Johnson, M. M. (1981). A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics*, 23, 351–361.
- Fligner, M. A. y Killeen, T. J. (1976). Distribution-Free Two-Sample Tests for Scale. *Journal of the American Statistical Association*, 71, 210–213.
- Hall, P. y Padmanabhan, A. R. (1997). Adaptive Inference for the Two-Sample Scale Problem. *Technometrics*, 39, 412–422.
- Hochberg, Y., Weiss, G. y Hart S., (1982). On Graphical Procedures for Multiple Comparisons. *Journal of the American Statistical Association*, 77, 767–772.
- Layard, M. W. J. (1973). Robust Large-Sample Tests for Homogeneity of Variances. *Journal of the American Statistical Association*, 68, 195–198.
- Levene, H. (1960). "Robust Tests for Equality of Variances," in I. Olkin, ed., *Contributions to Probability and Statistics*, Palo Alto, CA: Stanford University Press, 278–292.
- Lim, T.-S. y Loh, W.-Y. (1996). A Comparison of Tests of Equality of Variances. *Computational Statistics and Data Analysis*, 22, 287–301.
- Ott, L. (1993). *An Introduction to Statistical Methods and Data Analysis*, Belmont, CA: Duxbury Press.
- Pan, G. (1999). On a Levene Type Test for Equality of Two Variances. *Journal of Statistical Computation and Simulation*, 63, 59–71.
- Shoemaker, L. H. (2003). Fixing the F Test for Equal Variances. *The American Statistician*, 57, 105–114.
- Wolfram, S. (1999). *The Mathematica Book*, 4th ed. Wolfram Media/Cambridge University Press.

© 2020 Minitab, LLC. All rights reserved. Minitab®, Minitab Workspace™, Companion by Minitab®, Salford Predictive Modeler®, SPM®, and the Minitab® logo are all registered trademarks of Minitab, LLC, in the United States and other countries. Additional trademarks of Minitab, LLC can be found at www.minitab.com. All other marks referenced remain the property of their respective owners.