

## INFORME TÉCNICO SOBRE EL ASISTENTE DE MINITAB

Este documento forma parte de un conjunto de informes técnicos que explican la investigación llevada a cabo por los especialistas en estadística de Minitab para desarrollar los métodos y las verificaciones de los datos que se utilizan en el Asistente de Minitab Statistical Software.

# Prueba de desviación estándar para 1 muestra

## Revisión general

La prueba de desviación estándar para 1 muestra se utiliza para estimar la variabilidad del proceso y para comparar la variabilidad con un valor objetivo. Por lo general, la variabilidad se mide usando la varianza o, de manera equivalente, la desviación estándar.

Se han desarrollado muchos métodos estadísticos para evaluar la varianza de una población, cada uno con sus propias fortalezas y limitaciones. El método clásico de chi-cuadrado utilizado para probar la varianza probablemente es el que se emplea con mayor frecuencia, pero es extremadamente sensible al supuesto de normalidad y puede producir resultados extremadamente inexactos cuando los datos son asimétricos o tienen colas pesadas. También se han desarrollado otros métodos, pero estos también tienen inconvenientes. Por ejemplo, algunos métodos son válidos solo para muestras grandes o para datos de una distribución simétrica (consulte el Apéndice A).

En Minitab 15, usamos un método alternativo para muestras grandes que derivamos de una aproximación escalada de chi-cuadrado a la distribución de la varianza de la muestra de Box (1953). Este método, conocido como el método Grados de libertad ajustados (GLAjust), es menos sensible al supuesto de normalidad para muestras suficientemente grandes y se ha demostrado que produce intervalos de confianza más exactos que otros métodos (Hummel, Banga y Hettmansperger, 2005). Sin embargo, más recientemente, se desarrolló un método estadístico revisado de Bonett (2006) que parece proporcionar mejores intervalos de confianza aproximados.

En este trabajo, evaluamos el desempeño del método de Bonett. Además, para la planificación del tamaño de la muestra, investigamos la función de potencia para el procedimiento de prueba equivalente asociado con los intervalos de confianza de Bonett. Con base en nuestros resultados, usamos el método de Bonett para la prueba de desviación estándar para 1 muestra en el Asistente. También examinamos las siguientes verificaciones de los datos que se realizan de manera automática y se muestran en la Tarjeta de informe del Asistente y explicamos cómo afectan los resultados:

- Datos poco comunes
- Validez de la prueba
- Tamaño de la muestra

# Métodos de desviación estándar para 1 muestra

## Comparación del método de Bonett con el método GLAjust

Antes de la publicación del método de Bonett (2006), el procedimiento más robusto para hacer inferencias sobre la varianza de una población solía ser el método GLAjust. Sin embargo, los resultados publicados por Bonett revelan que el método de Bonett proporciona niveles de confianza estables que están cerca del nivel objetivo cuando se toman muestras de tamaño moderado de poblaciones no normales. Por lo tanto, el método de Bonett podría ser preferible para hacer inferencias sobre la desviación estándar o la varianza de una población.

### Objetivo

Queríamos comparar el desempeño del método de Bonett con el método GLAjust cuando se hacen inferencias sobre la varianza de una población individual. Específicamente, queríamos determinar cuál método produce intervalos de confianza más exactos para la varianza (o la desviación estándar) cuando se toman muestras de diferentes tamaños de poblaciones no normales.

Comparamos intervalos de confianza porque el método de Bonett se aplica directamente a los intervalos de confianza. El procedimiento de prueba de hipótesis equivalente asociado con los intervalos de confianza de Bonett puede derivarse. Sin embargo, para comparar directamente nuestros resultados con los publicados en Bonett (2006), examinamos los intervalos de confianza en lugar de las pruebas de hipótesis.

### Método

El método GLAjust y el método de Bonett se definen formalmente en el Apéndice B. Para comparar la exactitud de los intervalos de confianza para cada método, realizamos las siguientes simulaciones. En primer lugar, generamos muestras aleatorias de diferentes tamaños de distribuciones con diferentes propiedades, tales como distribuciones asimétricas y con colas pesadas, simétricas y con colas pesadas y simétricas y con colas livianas. Para cada tamaño de muestra, se extrajeron 10,000 réplicas de muestra de cada distribución, y se calcularon intervalos de confianza bilaterales de 95% para la varianza real de la distribución usando cada uno de los métodos. Luego calculamos la proporción de los 10,000 intervalos que contenían la varianza real, mencionada como la probabilidad de cobertura simulada. Si los intervalos de confianza son exactos, la probabilidad de cobertura simulada debe estar cerca de la probabilidad de cobertura objetivo de 0.95. Además, calculamos las anchuras promedio asociadas con los intervalos de confianza para cada método. Si los intervalos de confianza de los dos métodos tienen

aproximadamente las mismas probabilidades de cobertura simulada, entonces el método que produce intervalos más cortos (en promedio) es más preciso. Para obtener más detalles, consulte el Apéndice C.

## Resultados

El método de Bonett por lo general produce mejores probabilidades de cobertura e intervalos de confianza más exactos que el método GLAjust. Como resultado, las pruebas estadísticas para la varianza basadas en el método de Bonett generan tasas de error Tipo I y Tipo II más bajas. Por esa razón, la prueba de desviación estándar para 1 muestra disponible en el Asistente se basa en el método de Bonett.

Además, nuestros resultados indican que si la distribución tiene colas de moderadas a pesadas, el método de Bonett requiere tamaños de muestra más grandes para alcanzar el nivel de exactitud objetivo:

- Para las distribuciones con colas normales o livianas, un tamaño de muestra de 20 es suficiente.
- Para las distribuciones con colas moderadamente pesadas, el tamaño de la muestra debe ser de al menos 80.
- Para las distribuciones con colas pesadas, el tamaño de la muestra debe ser de al menos 200.

Por lo tanto, para asegurar que los resultados de la prueba de desviación estándar para 1 muestra o de los intervalos de confianza sean válidos para sus datos, el Asistente incluye una verificación de los datos para evaluar simultáneamente el tamaño de la muestra y las colas de la distribución de datos (consulte Validez de la prueba en las verificaciones de los datos, más adelante).

## Desempeño de la potencia teórica

El método de Bonett se aplica directamente a los intervalos de confianza para la varianza (o la desviación estándar). Sin embargo, usando la relación estadística entre las pruebas de hipótesis y los intervalos de confianza, podemos derivar la prueba equivalente que está asociada con los intervalos de confianza aproximados de Bonett. Dado que no está disponible una función exacta de la potencia para esta prueba, necesitábamos derivarla. Además, queríamos evaluar la sensibilidad de la función de potencia teórica al supuesto de normalidad.

### Objetivo

Queríamos determinar si podíamos usar la función de potencia teórica de la prueba asociada con los intervalos de confianza de Bonett para evaluar los requisitos de potencia y tamaño de la muestra para la prueba de desviación estándar para 1 muestra disponible en el Asistente. Para ello, necesitábamos evaluar si esta función de potencia teórica refleja con exactitud la potencia real de la prueba cuando se analizan datos normales y no normales.

## Método

La función de potencia teórica de la prueba usando el método de Bonett se deriva en el Apéndice C. Realizamos simulaciones para estimar los niveles de potencia real (a los que nos referimos como niveles de potencia simulada) usando el método de Bonett. En primer lugar, generamos muestras aleatorias de diferentes tamaños de las distribuciones descritas en el estudio anterior: distribuciones asimétricas y con colas pesadas, simétricas y con colas pesadas y simétricas y con colas livianas. Para cada distribución, realizamos la prueba en cada una de las 10,000 réplicas de muestra. Para cada tamaño de muestra, calculamos la potencia simulada de la prueba para detectar una determinada diferencia como la fracción de las 10,000 muestras para la que la prueba es significativa. Para comparación, también calculamos el nivel de potencia correspondiente utilizando la función de potencia teórica de la prueba. Si la función de potencia teórica no es demasiado sensible a la normalidad, los niveles de potencia teórica y simulada deben estar cercanos entre sí para los datos normales y no normales. Para obtener más detalles, consulte el Apéndice D.

## Resultados

Nuestras simulaciones revelaron que, cuando la muestra proviene de una distribución con colas normales o livianas, la potencia teórica y la potencia simulada de la prueba usando el método de Bonett son casi iguales. Sin embargo, cuando la muestra proviene de una distribución con colas pesadas, la función de potencia teórica podría ser conservadora y sobrestimar el tamaño de la muestra necesario para lograr una determinada potencia. Por lo tanto, la función de potencia teórica de la prueba asegura que el tamaño de la muestra sea lo suficientemente grande como para detectar una diferencia importante desde el punto de vista práctico en la desviación estándar, independientemente de la distribución. Sin embargo, si los datos provienen de distribuciones con colas pesadas, el tamaño estimado de la muestra podría ser mayor que el tamaño que se requiere en realidad, lo cual podría significar costos más altos de lo necesario en lo que respecta la obtención de las muestras.

# Verificaciones de los datos

## Datos poco comunes

Los datos poco comunes son valores de los datos extremadamente grandes o pequeños, también conocidos como valores atípicos. Los datos poco comunes pueden tener una fuerte influencia en los resultados de los análisis y pueden afectar las probabilidades de encontrar resultados estadísticamente significativos, especialmente cuando la muestra es pequeña. Los datos poco comunes pueden indicar problemas con la recolección de datos o pueden deberse a un comportamiento poco común del proceso que se está estudiando. Por lo tanto, generalmente vale la pena investigar estos puntos de los datos, que deben corregirse cada vez que sea posible.

### Objetivo

Queríamos desarrollar un método para buscar los valores de los datos que son muy grandes o muy pequeños en comparación con la muestra general y que pueden afectar los resultados del análisis.



### Método

Desarrollamos un método para buscar los datos poco comunes con base en el método descrito por Hoaglin, Iglewicz y Tukey (1986) que se utiliza para identificar los valores atípicos en las gráficas de caja.

### Resultados

El Asistente identifica un punto de los datos como poco común si es más de 1.5 veces el rango intercuartil más allá del cuartil superior o inferior de la distribución. Los cuartiles inferior y superior están en los percentiles 25 y 75 de los datos. El rango intercuartil es la diferencia entre los dos cuartiles. Este método funciona de forma adecuada incluso cuando existen múltiples valores atípicos, porque permite detectar cada valor atípico específico.

Al verificar si existen datos poco comunes, el Asistente muestra los siguientes indicadores de estado en la Tarjeta de informe:

Estado	Condición
	No hay puntos poco comunes en los datos.
	Al menos uno de los puntos de los datos es poco común y puede tener una fuerte influencia en los resultados.

## Validez de la prueba

Anteriormente, en la sección Métodos de desviación estándar para 1 muestra, indicamos que el método de Bonett por lo general ofrece mejores resultados que el método GLAjust. Sin embargo, cuando las colas de una distribución son más pesadas, el método de Bonett requiere muestras de mayor tamaño para lograr resultados exactos. Por lo tanto, un método para evaluar la validez de la prueba no debe basarse solamente en el tamaño de la muestra sino también en el peso de las colas de la distribución de origen. Gel et al. (2007) desarrollaron una prueba para determinar si una muestra proviene de una distribución con colas pesadas. Esta prueba, denominada prueba SJ, se basa en la relación de la desviación estándar de la muestra ( $s$ ) y el estimador de cola J (para obtener más detalles, consulte el Apéndice E).

### Objetivo

Para una muestra determinada de datos, necesitábamos crear una regla para evaluar la validez del método de Bonett evaluando el peso de las colas en los datos.



### Método

Realizamos simulaciones para investigar la potencia de la prueba SJ para identificar las distribuciones de colas pesadas. Si la prueba SJ es potente para muestras moderadamente grandes, entonces puede usarse para discriminar entre distribuciones de colas pesadas y colas livianas para nuestro propósito. Para obtener más detalles, consulte el Apéndice F.

### Resultados

Nuestras simulaciones revelaron que cuando las muestras son lo suficientemente grandes, la prueba SJ puede usarse para discriminar entre las distribuciones de colas pesadas y colas livianas. Para muestras de tamaño moderado o grande, los valores  $p$  más pequeños indican colas más pesadas y los valores  $p$  más grandes indican colas más livianas. Sin embargo, como las muestras más grandes tienden a tener valores  $p$  más pequeños que las muestras más pequeñas, también consideramos el tamaño de la muestra para determinar el peso de las colas. Por lo tanto, creamos nuestro conjunto de reglas para que el Asistente clasificara las colas de la distribución para cada muestra con base en el tamaño de la muestra y el valor  $p$  de la prueba SJ. Para ver los rangos específicos de valores  $p$  y tamaños de muestra asociados con las distribuciones de colas livianas, moderadas y pesadas, consulte el Apéndice F.

Con base en estos resultados, la Tarjeta de informe del Asistente muestra los siguientes indicadores de estado para evaluar la validez de la prueba de desviación estándar para 1 muestra (método de Bonett) para los datos de su muestra:

Estado	Condición
	<p>No hay evidencia de que su muestra tenga colas pesadas. El tamaño de su muestra es lo suficientemente grande como para verificar esta condición de manera fiable.</p> <p>O</p> <p>Su muestra tiene colas moderadamente pesadas o pesadas. Sin embargo, el tamaño de su muestra es lo suficientemente grande como para compensar, de modo que el valor <math>p</math> debería ser exacto.</p>
	<p>Su muestra tiene colas moderadamente pesadas o pesadas. El tamaño de su muestra no es lo suficientemente grande como para compensar. Sea prudente cuando interprete los resultados.</p> <p>O</p> <p>Su muestra no es lo suficientemente grande como para verificar colas pesadas de manera fiable. Sea prudente cuando interprete los resultados.</p>

## Tamaño de la muestra

Normalmente, una prueba de hipótesis estadística se realiza con el fin de reunir evidencia para rechazar la hipótesis nula de “ninguna diferencia”. Si la muestra es demasiado pequeña, la potencia de la prueba podría no ser adecuada para detectar una diferencia que en realidad existe, lo que produce un error Tipo II. Por lo tanto, resulta crucial asegurarse de que los tamaños de las muestras sean lo suficientemente grandes como para detectar diferencias importantes desde el punto de vista práctico con alta probabilidad.

### Objetivo

Si los datos no proporcionan suficiente evidencia para rechazar la hipótesis nula, necesitamos determinar si los tamaños de las muestras son lo suficientemente grandes como para que la prueba detecte diferencias de interés desde el punto de vista práctico con alta probabilidad. Aunque el objetivo de planificar el tamaño de la muestra es asegurar que los tamaños de las muestras sean lo suficientemente grandes como para detectar diferencias importantes con alta probabilidad, no deben ser tan grandes como para que las diferencias insignificantes se vuelvan estadísticamente significativas con alta probabilidad.

### Método

El análisis de potencia y tamaño de la muestra para la prueba de desviación estándar para 1 muestra se basa en la función de potencia teórica de la prueba. Esta función de potencia proporciona estimaciones adecuadas cuando los datos tienen colas casi normales o colas livianas, pero pueden producir estimaciones conservadoras cuando los datos tienen colas








pesadas (ver los resultados de la simulación resumidos en el punto Desempeño de la función de potencia teórica en la sección Métodos de desviación estándar para 1 muestra).

## Resultados

Cuando los datos no ofrecen suficiente evidencia en contra de la hipótesis nula, el Asistente utiliza la función de potencia de la prueba de aproximación a la normal para calcular las diferencias prácticas que pueden detectarse con un 80% y un 90% de probabilidad para el tamaño de muestra dado. Además, si el usuario especifica una diferencia práctica de interés en particular, el Asistente utiliza la función de potencia de la prueba de aproximación a la normal para calcular los tamaños de las muestras que produzcan un 80% y un 90% de probabilidad de detección de la diferencia.

Para ayudar a interpretar los resultados, la Tarjeta de informe del Asistente para la prueba de desviación estándar para 1 muestra presenta los siguientes indicadores de estado cuando se verifica la potencia y el tamaño de la muestra:

Estado	Condición
	La prueba detecta una diferencia entre la desviación estándar y el valor objetivo, así que la potencia no es un problema. O La potencia es suficiente. La prueba no detectó una diferencia entre la desviación estándar y el valor objetivo, pero la muestra es lo suficientemente grande como para proporcionar al menos un 90% de probabilidad de detectar la diferencia dada.
	La potencia podría ser suficiente. La prueba no detectó una diferencia entre la desviación estándar y el valor objetivo, pero la muestra es lo suficientemente grande como para ofrecer de un 80% a un 90% de probabilidad de detectar la diferencia dada. Se indica el tamaño de la muestra que se necesita para lograr una potencia de 90%.
	La potencia podría no ser suficiente. La prueba no detectó una diferencia entre la desviación estándar y el valor objetivo y la muestra es lo suficientemente grande como para ofrecer de un 60% a un 80% de probabilidad de detectar la diferencia dada. Se indican los tamaños de las muestras que se necesitan para lograr una potencia de 80% y una potencia de 90%.
	La potencia no es suficiente. La prueba no detectó una diferencia entre la desviación estándar y el valor objetivo, y la muestra no es lo suficientemente grande como para proporcionar al menos un 60% de probabilidad de detectar la diferencia dada. Se indican los tamaños de las muestras que se necesitan para lograr una potencia de 80% y una potencia de 90%.
	La prueba no detectó una diferencia entre la desviación estándar y el valor objetivo. Usted no especificó una diferencia práctica que detectar. Dependiendo de sus datos, es posible que el informe indique las diferencias que se podrían detectar con un 80% y un 90% de probabilidad, con base en el tamaño de la muestra y el nivel de significancia.

# Referencias

- Bonett, D.G. (2006). Approximate confidence interval for standard deviation of nonnormal distributions. *Computational Statistics & Data Analysis*, 50, 775-782.
- Box, G.E.P. (1953). Non-normality and tests on variances. *Biometrika*, 40, 318.
- Efron, B. y Tibshirani, R. J. (1993). An introduction to the bootstrap. Boca Raton, FL: Chapman and Hall/CRC.
- Gel, Y. R., Miao, W. y Gastwirth, J. L. (2007). Robust directed tests of normality against heavy-tailed alternatives. *Computational Statistics & Data Analysis*, 51, 2734-2746.
- Hummel, R., Banga, S. y Hettmansperger, T.P. (2005). Better confidence intervals for the variance in a random sample. *Minitab Technical Report*.
- Lee, S.J. y Ping, S. (1996). Testing the variance of symmetric heavy-tailed distributions. *Journal of Statistical Computation and Simulation*, 56, 39-52.

# Apéndice A: Métodos para probar la varianza (o desviación estándar)

En la siguiente tabla se resumen las fortalezas y debilidades asociadas con diversos métodos utilizados para probar la varianza.

Método	Comentario
Procedimiento clásico de chi-cuadrado	Extremadamente sensible al supuesto de normalidad. Incluso las pequeñas desviaciones de la normalidad pueden producir resultados inexactos, sin importar qué tan grande sea la muestra. De hecho, cuando los datos se desvían de la normalidad, si se aumenta el tamaño de la muestra, disminuye la exactitud del procedimiento.
Método de muestras grandes basado en la distribución asintótica normal de la transformación logarítmica de la varianza de la muestra	Por lo general, es mejor que el método clásico de chi-cuadrado, pero requiere tamaños de muestra más grandes para ser insensible al supuesto de normalidad.
Método de muestras grandes basado en la expansión de Edgeworth para pruebas unilaterales (de cola superior) Consulte Lee y Ping (1996).	Produce tasas aceptables de error Tipo I, pero requiere que los datos provengan de una distribución simétrica.
Método de muestras grandes basado en una aproximación de la distribución de la varianza de la muestra por una distribución escalada de chi-cuadrado. Este método se conoce como el método Grados de libertad ajustados (GLAjust). Consulte Hummel, Banga y Hettmansperger (2005).	Ofrece mejor probabilidad de cobertura que el método basado en la distribución asintótica normal de la transformación logarítmica de la varianza de la muestra y el método ABC no paramétrico de aproximación bootstrap para los intervalos de confianza (Efron y Tibshirani, 1993). Se utiliza para la prueba de 1 varianza en Minitab 15.
Distribución asintótica ajustada de Bonett de la transformación logarítmica de la varianza de la muestra Consulte Bonett (2006).	Ofrece una probabilidad de cobertura adecuada para los intervalos de confianza, incluso para muestras moderadamente grandes. Sin embargo, requiere muestras mucho más grandes cuando los datos provienen de distribuciones con colas pesadas. Se utiliza para la prueba de 1 varianza y la prueba de desviación estándar para 1 muestra del Asistente en Minitab 16.

# Apéndice B: Definición del método de Bonett y del método GLAjust

Supongamos que  $x_1, \dots, x_n$  es una muestra aleatoria observada con un tamaño de  $n$  proveniente de una población con cuarto momento finito. Supongamos que  $\bar{x}$  y  $s$  son la media y la desviación estándar observadas de la muestra, respectivamente. Además, supongamos que  $\gamma$  y  $\gamma_e$  son la curtosis y el exceso de curtosis de la población, respectivamente, de modo que  $\gamma_e = \gamma - 3$ . Por consiguiente, para una población normal,  $\gamma = 3$  y  $\gamma_e = 0$ . Supongamos también que  $\sigma^2$  es la varianza desconocida de la población. En las secciones siguientes, presentamos dos métodos para hacer una inferencia sobre  $\sigma^2$ , el método Grados de libertad ajustados (GLAjust) y el método de Bonett.

## Fórmula B1: Método GLAjust

El método GLAjust se basa en una aproximación de la distribución de la varianza de la muestra por una distribución escalada de chi-cuadrado (consulte Box, 1953). Más específicamente, los dos primeros momentos de la varianza de la muestra se hacen coincidir con los momentos de una distribución escalada de chi-cuadrado para determinar la escala desconocida y los grados de libertad. Este enfoque produce el siguiente intervalo de confianza bilateral aproximado de  $(1 - \alpha)100$  por ciento para la varianza:

$$\left[ \frac{rs^2}{\chi_{r,\alpha/2}^2}, \frac{rs^2}{\chi_{r,1-\alpha/2}^2} \right]$$

donde

$$r = \frac{2n}{\hat{\gamma}_e + 2n/(n-1)}$$
$$\hat{\gamma}_e = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Esta estimación del exceso de curtosis es idéntica a la que se utiliza para los comandos de Estadísticas básicas en Minitab.

## Fórmula B2: Método de Bonett

El método de Bonett se basa en el enfoque clásico ampliamente conocido, que utiliza el teorema del límite central y el método  $\delta$  de Cramer para obtener una distribución asintótica de la transformación logarítmica de la varianza de la muestra. La transformación logarítmica se utiliza para acelerar la convergencia a la normalidad. Usando este enfoque, el intervalo de confianza bilateral aproximado de  $(1 - \alpha)100$  por ciento se define como:

$$[s^2 \exp(-z_{\alpha/2} se), s^2 \exp(z_{\alpha/2} se)]$$

donde  $z_{\alpha}$  es el percentil superior de la distribución normal estándar y  $se$  es una estimación asintótica del error estándar de la varianza de la muestra con transformación logarítmica, calculada como:

$$se = \sqrt{\frac{\hat{\gamma} - (n-3)/n}{n-1}} = \sqrt{\frac{\hat{\gamma}_e + 2 + 3/n}{n-1}}$$

Antes, Hummel et al. (2005) realizaron estudios de simulación que demostraron que el método GLAjust es superior a este enfoque clásico. Sin embargo, Bonett realiza dos ajustes al enfoque clásico para superar sus limitaciones. El primer ajuste implica la estimación de la curtosis. Para estimar la curtosis, Bonett utiliza la siguiente fórmula:

$$\hat{\gamma}_e = \frac{n}{(n-1)^2} \sum_{i=1}^n \left( \frac{x_i - m}{s} \right)^4 - 3$$

donde  $m$  es una media recortada con la proporción del recorte igual a  $1/2\sqrt{n-4}$ . Esta estimación de la curtosis tiende a mejorar la exactitud de los niveles de confianza para las distribuciones (simétricas o asimétricas) de colas pesadas.

Para el segundo ajuste, Bonett determina de manera empírica un multiplicador constante para la varianza de la muestra y el error estándar. Este multiplicador constante iguala de un modo aproximado las probabilidades de cola cuando la muestra es pequeña y se calcula como:

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

Estos ajustes producen el intervalo de confianza bilateral aproximado de  $(1 - \alpha)100$  por ciento de Bonett para la varianza:

$$[cs^2 \exp(-c z_{\alpha/2} se), cs^2 \exp(c z_{\alpha/2} se)]$$

# Apéndice C: Comparación entre el método de Bonett y el método GLAjust

## Simulación C1: Comparación de los intervalos de confianza

Queríamos comparar la exactitud de los intervalos de confianza para la varianza que se calculan utilizando el método GLAjust y el método de Bonett. Generamos muestras aleatorias de diferentes tamaños ( $n = 20, 30, 40, 50, 60, 80, 100, 150, 200, 250, 300$ ) de varias distribuciones y calculamos los intervalos de confianza usando cada método. Las distribuciones incluían:

- Distribución normal estándar ( $N(0,1)$ )
- Distribuciones simétricas y de colas livianas, incluyendo la distribución uniforme ( $U(0,1)$ ) y la distribución Beta con ambos parámetros establecidos en 3 ( $B(3,3)$ )
- Distribuciones simétricas y de colas pesadas, incluyendo distribuciones  $t$  con 5 y 10 grados de libertad ( $t(5), t(10)$ ) y la distribución de Laplace con ubicación 0 y escala 1 ( $Lpl$ )
- Distribuciones asimétricas y de colas pesadas, incluyendo la distribución exponencial con escala 1 ( $Exp$ ) y distribuciones de chi-cuadrado con 3, 5 y 10 grados de libertad ( $Chi(3)$ ,  $Chi(5)$ ,  $Chi(10)$ )
- Distribución asimétrica hacia la izquierda y de colas pesadas; específicamente, la distribución Beta con los parámetros establecidos en 8 y 1, respectivamente ( $B(8,1)$ )

Además, para evaluar el efecto directo de los valores atípicos, generamos muestras de distribuciones normales contaminadas definidas como

$$CN(p, \sigma) = pN(0,1) + (1 - p)N(0, \sigma)$$

donde  $p$  es el parámetro de mezcla y  $1 - p$  es la proporción de contaminación (que equivale a la proporción de valores atípicos). Seleccionamos dos poblaciones normales contaminadas para el estudio:  $CN(0.9,3)$ , donde el 10% de la población son valores atípicos, y  $CN(0.8,3)$ , donde el 20% de la población son valores atípicos. Estas dos distribuciones son simétricas y tienen colas largas debido a los valores atípicos.

Para cada tamaño de muestra, se extrajeron 10,000 réplicas de muestra de cada distribución, y se calcularon intervalos de confianza bilaterales de 95% usando cada uno de los métodos. El generador de muestras aleatorias se sembró para que ambos métodos se aplicaran a las mismas muestras. Con base en estos intervalos de confianza, posteriormente calculamos las probabilidades de cobertura simulada ( $CovP$ ) y las anchuras promedio de los intervalos ( $AveW$ ) para cada método. Si los intervalos de confianza de los dos métodos tienen aproximadamente

las mismas probabilidades de cobertura simulada, entonces el método que produce intervalos más cortos (en promedio) es más preciso. Como usamos un nivel de confianza objetivo de 95%, el error de la simulación fue  $\sqrt{0.95(0.05)/10,000} = 0.2\%$ .

Los resultados de la simulación se muestran a continuación, en las tablas 1 y 2.

**Tabla 1** Probabilidades de cobertura simulada de los intervalos de confianza bilaterales de 95% para la varianza calculadas usando el método GLAjust y el método de Bonett. Estas muestras se generaon a partir de distribuciones simétricas con colas livianas, normales, casi normales o pesadas.

Distribución		Distribuciones simétricas con colas livianas, normales o casi normales				Distribuciones simétricas con colas pesadas			
		U(0,1)	B(3,3)	N(0,1)	t(10)	Lpl	CN(0.8, 3)	CN (0.9, 3)	T(5)
<b>Asimetría</b>		<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Curtosis (<math>\gamma_e</math>)</b>		<b>-1.200</b>	<b>-0.667</b>	<b>0</b>	<b>1.000</b>	<b>3.000</b>	<b>4.544</b>	<b>5.333</b>	<b>6.000</b>

**n = 10**

<b>GLAjust</b>	<b>CovP</b>	0.910	0.909	0.903	0.883	0.853	0.793	0.815	0.858
	<b>AveW</b>	0.154	0.087	3.276	5.160	13.924	21.658	14.913	11.742
<b>Bonett</b>	<b>CovP</b>	0.972	0.967	0.962	0.952	0.919	0.891	0.920	0.935
	<b>AveW</b>	0.242	0.115	3.710	5.134	10.566	15.335	10.367	8.578

**n = 20**

<b>GLAjust</b>	<b>CovP</b>	0.937	0.937	0.923	0.909	0.881	0.819	0.817	0.868
	<b>AveW</b>	0.080	0.045	1.572	2.463	5.781	9.265	6.539	5.151
<b>Bonett</b>	<b>CovP</b>	0.953	0.954	0.946	0.934	0.909	0.856	0.864	0.904
	<b>AveW</b>	0.100	0.051	1.683	2.422	4.932	7.282	4.945	4.026

**n = 30**

<b>GLAjust</b>	<b>CovP</b>	0.946	0.942	0.933	0.917	0.894	0.851	0.823	0.882
	<b>AveW</b>	0.061	0.034	1.170	1.764	4.117	6.330	4.557	3.667
<b>Bonett</b>	<b>CovP</b>	0.951	0.950	0.947	0.933	0.909	0.869	0.852	0.907
	<b>AveW</b>	0.070	0.037	1.221	1.750	3.654	5.383	3.736	2.997

**n = 40**

Distribución		Distribuciones simétricas con colas livianas, normales o casi normales				Distribuciones simétricas con colas pesadas			
		U(0,1)	B(3,3)	N(0,1)	t(10)	Lpl	CN(0.8, 3)	CN (0.9, 3)	T(5)
<b>Asimetría</b>		<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Curtosis (<math>\gamma_e</math>)</b>		<b>-1.200</b>	<b>-0.667</b>	<b>0</b>	<b>1.000</b>	<b>3.000</b>	<b>4.544</b>	<b>5.333</b>	<b>6.000</b>
<b>GLAjust</b>	<b>CovP</b>	0.953	0.947	0.932	0.922	0.904	0.867	0.833	0.890
	<b>AveW</b>	0.051	0.028	0.971	1.489	3.246	5.131	3.654	3.024
<b>Bonett</b>	<b>CovP</b>	0.954	0.951	0.941	0.936	0.914	0.879	0.856	0.907
	<b>AveW</b>	0.057	0.030	1.002	1.469	2.994	4.519	3.128	2.542
<b>n = 50</b>									
<b>GLAjust</b>	<b>CovP</b>	0.951	0.945	0.937	0.925	0.911	0.878	0.838	0.893
	<b>AveW</b>	0.045	0.025	0.849	1.291	2.789	4.357	3.091	2.603
<b>Bonett</b>	<b>CovP</b>	0.951	0.947	0.944	0.938	0.918	0.888	0.855	0.908
	<b>AveW</b>	0.049	0.026	0.870	1.280	2.613	3.939	2.729	2.240
<b>n = 60</b>									
<b>GLAjust</b>	<b>CovP</b>	0.949	0.943	0.938	0.926	0.913	0.890	0.853	0.899
	<b>AveW</b>	0.040	0.022	0.766	1.155	2.490	3.857	2.768	2.283
<b>Bonett</b>	<b>CovP</b>	0.949	0.947	0.943	0.935	0.918	0.896	0.868	0.910
	<b>AveW</b>	0.043	0.023	0.781	1.147	2.354	3.552	2.498	2.023
<b>n = 70</b>									
<b>GLAjust</b>	<b>CovP</b>	0.948	0.945	0.940	0.930	0.913	0.890	0.858	0.896
	<b>AveW</b>	0.037	0.020	0.701	1.056	2.283	3.458	2.475	2.049
<b>Bonett</b>	<b>CovP</b>	0.947	0.946	0.944	0.938	0.918	0.894	0.868	0.905
	<b>AveW</b>	0.039	0.021	0.713	1.049	2.174	3.227	2.272	1.828
<b>n = 80</b>									
<b>GLAjust</b>	<b>CovP</b>	0.947	0.949	0.938	0.929	0.918	0.905	0.869	0.902
	<b>AveW</b>	0.034	0.019	0.652	0.988	2.089	3.205	2.300	1.906
<b>Bonett</b>	<b>CovP</b>	0.946	0.950	0.942	0.935	0.923	0.907	0.877	0.911
	<b>AveW</b>	0.036	0.019	0.662	0.982	2.005	3.014	2.133	1.716



Distribución	Distribuciones simétricas con colas livianas, normales o casi normales				Distribuciones simétricas con colas pesadas			
	U(0,1)	B(3,3)	N(0,1)	t(10)	Lpl	CN(0.8, 3)	CN(0.9, 3)	T(5)
<b>Asimetría</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Curtosis (<math>\gamma_e</math>)</b>	<b>-1.200</b>	<b>-0.667</b>	<b>0</b>	<b>1.000</b>	<b>3.000</b>	<b>4.544</b>	<b>5.333</b>	<b>6.000</b>

$n = 90$

<b>GLAjust</b>	<b>CovP</b>	0.946	0.947	0.948	0.929	0.918	0.908	0.869	0.901
	<b>AveW</b>	0.032	0.018	0.611	0.921	1.951	2.982	2.124	1.874
<b>Bonett</b>	<b>CovP</b>	0.945	0.948	0.952	0.936	0.920	0.910	0.874	0.909
	<b>AveW</b>	0.034	0.018	0.618	0.916	1.882	2.822	1.984	1.646

$n = 100$

<b>GLAjust</b>	<b>CovP</b>	0.947	0.951	0.945	0.933	0.920	0.910	0.885	0.912
	<b>AveW</b>	0.030	0.017	0.576	0.873	1.830	2.801	2.017	1.658
<b>Bonett</b>	<b>CovP</b>	0.946	0.953	0.948	0.937	0.923	0.912	0.891	0.916
	<b>AveW</b>	0.032	0.017	0.583	0.869	1.772	2.666	1.899	1.522

$n = 150$

<b>GLAjust</b>	<b>CovP</b>	0.949	0.951	0.947	0.936	0.932	0.925	0.896	0.912
	<b>AveW</b>	0.024	0.014	0.464	0.700	1.470	2.228	1.602	1.325
<b>Bonett</b>	<b>CovP</b>	0.948	0.952	0.949	0.939	0.933	0.924	0.898	0.915
	<b>AveW</b>	0.025	0.014	0.467	0.698	1.438	2.156	1.539	1.251

$n = 200$

<b>GLAjust</b>	<b>CovP</b>	0.943	0.949	0.948	0.938	0.927	0.930	0.914	0.918
	<b>AveW</b>	0.021	0.012	0.400	0.605	1.265	1.906	1.373	1.178
<b>Bonett</b>	<b>CovP</b>	0.942	0.951	0.949	0.940	0.928	0.930	0.915	0.920
	<b>AveW</b>	0.021	0.012	0.402	0.603	1.245	1.860	1.333	1.106

$n = 250$

<b>GLAjust</b>	<b>CovP</b>	0.952	0.952	0.949	0.942	0.938	0.929	0.909	0.915
	<b>AveW</b>	0.019	0.010	0.355	0.538	1.120	1.690	1.219	1.037

Distribución		Distribuciones simétricas con colas livianas, normales o casi normales				Distribuciones simétricas con colas pesadas			
		U(0,1)	B(3,3)	N(0,1)	t(10)	Lpl	CN(0.8, 3)	CN(0.9, 3)	T(5)
<b>Asimetría</b>		<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Curtosis (<math>\gamma_e</math>)</b>		<b>-1.200</b>	<b>-0.667</b>	<b>0</b>	<b>1.000</b>	<b>3.000</b>	<b>4.544</b>	<b>5.333</b>	<b>6.000</b>
<b>Bonett</b>	<b>CovP</b>	0.951	0.952	0.949	0.944	0.941	0.929	0.909	0.916
	<b>AveW</b>	0.019	0.010	0.357	0.537	1.106	1.657	1.190	0.986
<b>n = 300</b>									
<b>GLAjust</b>	<b>CovP</b>	0.950	0.948	0.951	0.940	0.938	0.936	0.920	0.914
	<b>AveW</b>	0.017	0.009	0.324	0.490	1.019	1.544	1.115	0.933
<b>Bonett</b>	<b>CovP</b>	0.950	0.947	0.951	0.942	0.937	0.929	0.920	0.916
	<b>AveW</b>	0.017	0.010	0.325	0.489	1.009	1.657	1.093	0.897

**Tabla 2** Probabilidades de cobertura simulada de los intervalos de confianza bilaterales de 95% para la varianza calculadas usando el método GLAjust y el método de Bonett. Estas muestras se generaron a partir de distribuciones asimétricas con colas casi normales, moderadamente pesadas o pesadas.

Distribución		Distribuciones asimétricas con colas casi normales o moderadamente pesadas			Distribuciones asimétricas con colas pesadas	
		Chi(10)	B(8,1)	Chi(5)	Chi(3)	Exp
<b>Asimetría</b>		<b>0.894</b>	<b>-1.423</b>	<b>1.265</b>	<b>1.633</b>	<b>2</b>
<b>Curtosis (<math>\gamma_e</math>)</b>		<b>1.200</b>	<b>2.284</b>	<b>2.400</b>	<b>4.000</b>	<b>6</b>
<b>n = 10</b>						
<b>GLAjust</b>	<b>CovP</b>	0.869	0.815	0.836	0.797	0.758
	<b>AveW</b>	93.383	0.065	61.994	47.821	10.711
<b>Bonett</b>	<b>CovP</b>	0.950	0.917	0.938	0.911	0.882
	<b>AveW</b>	91.006	0.058	53.830	38.137	7.498
<b>n = 20</b>						
<b>GLAjust</b>	<b>CovP</b>	0.889	0.862	0.862	0.833	0.811
	<b>AveW</b>	41.497	0.026	25.479	20.099	4.293

Distribución		Distribuciones asimétricas con colas casi normales o moderadamente pesadas			Distribuciones asimétricas con colas pesadas	
		Chi(10)	B(8,1)	Chi(5)	Chi(3)	Exp
<b>Asimetría</b>		<b>0.894</b>	<b>-1.423</b>	<b>1.265</b>	<b>1.633</b>	<b>2</b>
<b>Curtosis (<math>\gamma_e</math>)</b>		<b>1.200</b>	<b>2.284</b>	<b>2.400</b>	<b>4.000</b>	<b>6</b>
<b>Bonett</b>	<b>CovP</b>	0.932	0.912	0.913	0.893	0.877
	<b>AveW</b>	41.600	0.026	24.094	17.232	3.370
<b>n = 30</b>						
<b>GLAjust</b>	<b>CovP</b>	0.901	0.881	0.880	0.864	0.838
	<b>AveW</b>	30.021	0.018	18.182	13.630	2.844
<b>Bonett</b>	<b>CovP</b>	0.931	0.920	0.914	0.906	0.885
	<b>AveW</b>	30.462	0.019	17.858	12.634	2.441
<b>n = 40</b>						
<b>GLAjust</b>	<b>CovP</b>	0.909	0.882	0.885	0.867	0.862
	<b>AveW</b>	24.459	0.015	14.577	10.649	2.193
<b>Bonett</b>	<b>CovP</b>	0.930	0.915	0.913	0.904	0.898
	<b>AveW</b>	24.952	0.015	14.504	1.991	1.991
<b>n = 50</b>						
<b>GLAjust</b>	<b>CovP</b>	0.912	0.900	0.892	0.871	0.868
	<b>AveW</b>	21.373	0.013	12.694	9.115	1.861
<b>Bonett</b>	<b>CovP</b>	0.930	0.927	0.916	0.903	0.901
	<b>AveW</b>	21.814	0.013	12.741	8.897	1.735
<b>n = 60</b>						
<b>GLAjust</b>	<b>CovP</b>	0.915	0.908	0.901	0.890	0.875
	<b>AveW</b>	18.928	0.011	11.338	8.211	1.645
<b>Bonett</b>	<b>CovP</b>	0.930	0.933	0.923	0.917	0.900
	<b>AveW</b>	19.369	0.012	11.456	8.093	1.554
<b>n = 70</b>						
<b>GLAjust</b>	<b>CovP</b>	0.915	0.910	0.904	0.898	0.881
	<b>AveW</b>	17.513	0.010	10.307	7.461	1.488

Distribución		Distribuciones asimétricas con colas casi normales o moderadamente pesadas			Distribuciones asimétricas con colas pesadas	
		Chi(10)	B(8,1)	Chi(5)	Chi(3)	Exp
<b>Asimetría</b>		<b>0.894</b>	<b>-1.423</b>	<b>1.265</b>	<b>1.633</b>	<b>2</b>
<b>Curtosis (<math>\gamma_e</math>)</b>		<b>1.200</b>	<b>2.284</b>	<b>2.400</b>	<b>4.000</b>	<b>6</b>
<b>Bonett</b>	<b>CovP</b>	0.932	0.932	0.922	0.919	0.906
	<b>AveW</b>	17.906	0.011	10.464	7.408	1.429
<b>n = 80</b>						
<b>GLAjust</b>	<b>CovP</b>	0.920	0.916	0.911	0.904	0.890
	<b>AveW</b>	16.157	0.009	9.604	6.892	1.349
<b>Bonett</b>	<b>CovP</b>	0.935	0.936	0.929	0.924	0.915
	<b>AveW</b>	16.537	0.010	9.765	6.882	1.314
<b>n = 90</b>						
<b>GLAjust</b>	<b>CovP</b>	0.924	0.918	0.911	0.897	0.894
	<b>AveW</b>	15.250	0.009	9.007	6.323	1.255
<b>Bonett</b>	<b>CovP</b>	0.938	0.936	0.929	0.918	0.913
	<b>AveW</b>	15.609	0.009	9.175	6.366	1.230
<b>n = 100</b>						
<b>GLAjust</b>	<b>CovP</b>	0.926	0.919	0.915	0.908	0.895
	<b>AveW</b>	14.332	0.008	8.451	6.016	1.171
<b>Bonett</b>	<b>CovP</b>	0.935	0.936	0.931	0.924	0.916
	<b>AveW</b>	14.664	0.009	8.625	6.063	1.158
<b>n = 150</b>						
<b>GLAjust</b>	<b>CovP</b>	0.933	0.925	0.923	0.913	0.911
	<b>AveW</b>	11.606	0.007	6.781	4.792	0.933
<b>Bonett</b>	<b>CovP</b>	0.943	0.941	0.936	0.929	0.928
	<b>AveW</b>	11.846	0.007	6.942	4.875	0.937
<b>n = 200</b>						
<b>GLAjust</b>	<b>CovP</b>	0.935	0.934	0.926	0.916	0.915
	<b>AveW</b>	9.973	0.006	5.849	4.127	0.799

Distribución		Distribuciones asimétricas con colas casi normales o moderadamente pesadas			Distribuciones asimétricas con colas pesadas	
		Chi(10)	B(8,1)	Chi(5)	Chi(3)	Exp
<b>Asimetría</b>		<b>0.894</b>	<b>-1.423</b>	<b>1.265</b>	<b>1.633</b>	<b>2</b>
<b>Curtosis (<math>\gamma_e</math>)</b>		<b>1.200</b>	<b>2.284</b>	<b>2.400</b>	<b>4.000</b>	<b>6</b>
<b>Bonett</b>	<b>CovP</b>	0.942	0.948	0.936	0.930	0.931
	<b>AveW</b>	10.185	0.006	5.991	4.212	0.808
<b>n = 250</b>						
<b>GLAjust</b>	<b>CovP</b>	0.938	0.939	0.934	0.926	0.922
	<b>AveW</b>	8.899	0.005	5.231	3.652	0.705
<b>Bonett</b>	<b>CovP</b>	0.946	0.951	0.944	0.936	0.931
	<b>AveW</b>	9.078	0.005	5.355	3.735	0.716
<b>n = 300</b>						
<b>GLAjust</b>	<b>CovP</b>	0.942	0.938	0.934	0.931	0.922
	<b>AveW</b>	8.156	0.005	4.749	3.344	0.640
<b>Bonett</b>	<b>CovP</b>	0.947	0.948	0.943	0.941	0.933
	<b>AveW</b>	8.314	0.005	4.862	3.419	0.651

Nuestros resultados son muy consistentes con los publicados por Bonett (2006). Como se muestra en las tablas 1 y 2, los intervalos de confianza calculados usando el método de Bonett son superiores a los intervalos de confianza calculados usando el método GLAjust, porque producen probabilidades de cobertura más cercanas al nivel objetivo de 0.95 e intervalos de confianza más estrechos, en promedio. Si los intervalos de confianza de los dos métodos tienen aproximadamente las mismas probabilidades de cobertura simulada, entonces el método que produce intervalos más cortos (en promedio) es más preciso. Esto significa que la prueba estadística para la varianza que se basa en el método de Bonett funciona mejor y produce tasas más bajas de error Tipo I y Tipo II. Cuando los tamaños de las muestras son grandes, los dos métodos producen resultados casi idénticos, pero para tamaños de muestra de pequeños a moderados, el método de Bonett es superior.

Aunque el método de Bonett generalmente funciona mejor que el método GLAjust, consistentemente produce probabilidades de cobertura por debajo de la cobertura objetivo de 0.95 para distribuciones de colas pesadas (simétricas o asimétricas) incluso para muestras muy grandes ( $n > 100$ ). Esto se ilustra a continuación en la figura 1, que grafica las probabilidades de cobertura simulada del método de Bonett en función del exceso de curtosis real de la población para tamaños de muestra pequeños, moderados y grandes.

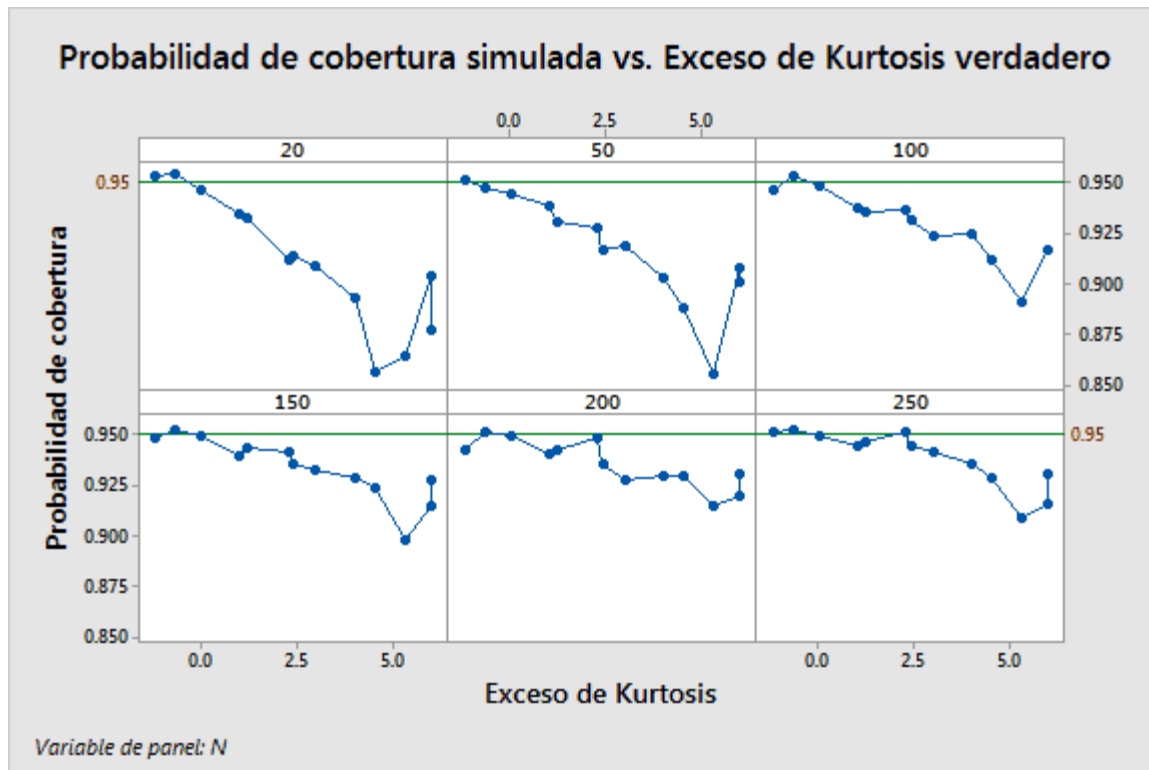


Figura 1 Probabilidades de cobertura simulada para los intervalos de confianza de 95% de Bonett graficadas en función del exceso de curtosis de cada distribución con diferentes tamaños de muestra.

En la figura 1 también puede observarse que mientras mayor sea la curtosis, mayor será el tamaño de la muestra que se necesitará para que las probabilidades de cobertura simulada se acerquen al nivel objetivo. Como se indicó anteriormente, las probabilidades de cobertura simulada para el método de Bonett son bajas para las distribuciones de colas pesadas. Sin embargo, para distribuciones de colas más livianas, tales como las distribuciones uniforme y Beta(3,3), las probabilidades de cobertura simulada son estables y alcanzan el objetivo para tamaños de muestra tan pequeños como 20. Por lo tanto, basamos nuestro criterio para determinar la validez del método de Bonett en el tamaño de la muestra y en el peso de las colas de la distribución de la cual se extrajo la muestra.

Como un primer paso para crear este criterio, clasificamos las distribuciones en tres categorías según el peso de las colas:

- **Distribuciones de colas livianas o de colas normales (tipo L):** Son distribuciones para las cuales los intervalos de confianza de Bonett producen probabilidades de cobertura estables cercanas al nivel de cobertura objetivo. Para estas distribuciones, los tamaños de muestra tan bajos como 20 producen resultados exactos. Como ejemplos se pueden mencionar la distribución uniforme, la distribución Beta(3,3), la distribución normal, la distribución t con 10 grados de libertad y la distribución de chi-cuadrado con 10 grados de libertad.

- **Distribuciones con colas moderadamente pesadas (tipo M):** Para estas distribuciones, el método de Bonett requiere un tamaño de muestra mínimo de 80 para que las probabilidades de cobertura simulada estén cerca de la cobertura objetivo. Ejemplos de estas distribuciones son la distribución de chi-cuadrado con 5 grados de libertad y la distribución Beta(8,1).
- **Distribuciones de colas pesadas (tipo H):** Son distribuciones para las cuales los intervalos de confianza de Bonett producen probabilidades de cobertura que están muy por debajo de la cobertura objetivo, a menos que los tamaños de muestra sean extremadamente grandes ( $n \geq 200$ ). Los ejemplos incluyen la distribución t con 5 grados de libertad, la distribución de Laplace, la distribución de chi-cuadrado con 3 grados de libertad, la distribución exponencial y las dos distribuciones normales contaminadas, CN(0.9,3) y CN(0.8,3).

Por lo tanto, una regla general para evaluar la validez del método de Bonett requiere que creamos un procedimiento para identificar de cuál de los 3 tipos de distribución provienen los datos de la muestra. Desarrollamos este procedimiento como parte de la Validez de la prueba en las verificaciones de los datos. Para obtener más detalles, consulte el Apéndice E.

# Apéndice D: Potencia teórica

Derivamos la función de potencia teórica de la prueba asociada con el método de Bonett y realizamos simulaciones para comparar la potencia teórica y la potencia simulada de la prueba. Si las curvas de potencia teórica y simulada están cercanas entre sí, entonces el análisis de potencia y tamaño de la muestra basado en la función de potencia teórica debería producir resultados exactos.

## Fórmula D1: Función de potencia teórica para el método de Bonett

Como se describió anteriormente, el método de Bonett se basa en el enfoque clásico ampliamente conocido, en el cual se utilizan el teorema del límite central y el método  $\delta$  de Cramer para hallar una distribución asintótica de la transformación logarítmica de la varianza de la muestra. Más específicamente, se establece que en muestras grandes,  $\frac{\ln S^2 - \ln \sigma^2}{se}$  está distribuido aproximadamente como la distribución normal estándar. El denominador,  $se$ , es el error estándar de muestras grandes de la varianza de la muestra con transformación logarítmica y se calcula como

$$se = \sqrt{\frac{\gamma - (n-3)/n}{n-1}}$$

donde  $\gamma$  es la curtosis de la población original desconocida.

Por consiguiente, una función de potencia aproximada con un nivel de significancia aproximado para la prueba bilateral usando el método de Bonett puede calcularse en función del tamaño de la muestra, la relación  $\rho = \sigma/\sigma_0$  y la curtosis de la población original  $\gamma$  como

$$\pi(n, \rho, \gamma) = 1 - \Phi\left(z_{\alpha/2} - \frac{\ln \rho^2}{\sqrt{\frac{\gamma - 1 + 3/n}{n-1}}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\ln \rho^2}{\sqrt{\frac{\gamma - 1 + 3/n}{n-1}}}\right)$$

donde  $\sigma_0$  es el valor hipotético de la desviación estándar desconocida,  $\Phi$  es la CDF de la distribución normal estándar y  $z_\alpha$  es el punto percentil  $\alpha$  superior de la distribución normal estándar. Las funciones unilaterales de potencia también pueden obtenerse con estos cálculos.

Tenga en cuenta que al planificar el tamaño de la muestra para un estudio, puede utilizarse una estimación de la curtosis en lugar de la curtosis real. Esta estimación por lo general se basa en las opiniones de expertos o en los resultados de experimentos anteriores. Si esa información no está disponible, llevar a cabo un estudio piloto pequeño para planificar el estudio principal suele ser una práctica recomendada. Usando una muestra del estudio piloto, la curtosis puede estimarse como



$$\hat{\gamma} = \frac{n}{(n-1)^2} \sum_{i=1}^n \left( \frac{x_i - m}{s} \right)^4$$

donde  $m$  es una media recortada con la proporción del recorte igual a  $1/2\sqrt{n-4}$ .

## Simulación D1: Comparación de la potencia real con la potencia teórica

Diseñamos una simulación para comparar los niveles estimados de potencia real (también conocidos como niveles de potencia simulada) con los niveles de potencia teórica (también denominados niveles de potencia aproximada) cuando se usa el método de Bonett para probar la varianza.

En cada experimento, generamos 10,000 réplicas de muestra, cada una con un tamaño de  $n$ , donde  $n = 20, 30, 40, 50, \dots, 120$ , de cada una de las distribuciones descritas en la simulación C1 (consulte el Apéndice C). Para cada distribución y tamaño de muestra  $n$ , calculamos el nivel de potencia simulada como la fracción de las 10,000 réplicas de muestra aleatorias para la cual la prueba bilateral con nivel de significancia  $\alpha = 0.05$  era significativa. Durante el cálculo de la potencia simulada, usamos  $\rho = \sigma/\sigma_0 = 1.25$  para obtener niveles de potencia relativamente pequeños. Posteriormente, para hacer una comparación, calculamos los niveles de potencia correspondientes utilizando la función de potencia teórica.

Los resultados se muestran en las tablas 3 y 4 y se representan gráficamente en la figura 2.

**Tabla 3** Niveles de potencia simulada (evaluados en  $\rho = \sigma/\sigma_0 = 1.25$ ) de una prueba bilateral para la varianza basada en el método de Bonett en comparación con los niveles de potencia teórica (aproximación a la normal). Las muestras se generaron a partir de distribuciones simétricas con colas livianas, normales, casi normales o pesadas.

$n$	Potencia	Distribuciones simétricas con colas livianas, normales o casi normales				Distribuciones simétricas con colas pesadas			
		U(0,1)	B(3,3)	N(0,1)	t(10)	Lpl	CN (.8,3)	CN (.9,3)	t(5)
20	Simul.	0.521	0.390	0.310	0.237	0.178	0.152	0.139	0.172
	Aprox.	0.514	0.359	0.264	0.195	0.137	0.117	0.109	0.104
30	Simul.	0.707	0.551	0.441	0.337	0.225	0.186	0.169	0.228
	Aprox.	0.717	0.519	0.382	0.276	0.186	0.154	0.143	0.135
40	Simul.	0.831	0.679	0.526	0.427	0.285	0.266	0.203	0.285
	Aprox.	0.846	0.651	0.490	0.356	0.236	0.192	0.176	0.165
50	Simul.	0.899	0.753	0.621	0.505	0.332	0.255	0.238	0.340
	Aprox.	0.921	0.754	0.586	0.431	0.284	0.229	0.210	0.196

n	Potencia	Distribuciones simétricas con colas livianas, normales o casi normales				Distribuciones simétricas con colas pesadas			
		U(0,1)	B(3,3)	N(0,1)	t(10)	Lpl	CN (.8,3)	CN (.9,3)	t(5)
60	Simul.	0.942	0.822	0.701	0.570	0.380	0.285	0.274	0.384
	Aprox.	0.961	0.830	0.668	0.501	0.332	0.266	0.243	0.227
70	Simul.	0.964	0.866	0.757	0.632	0.424	0.327	0.314	0.439
	Aprox.	0.981	0.885	0.737	0.566	0.379	0.303	0.276	0.257
80	Simul.	0.981	0.909	0.815	0.689	0.481	0.372	0.347	0.483
	Aprox.	0.991	0.923	0.794	0.624	0.423	0.340	0.309	0.288
90	Simul.	0.988	0.937	0.851	0.724	0.514	0.400	0.377	0.523
	Aprox.	0.996	0.950	0.840	0.676	0.467	0.375	0.342	0.318
100	Simul.	0.994	0.961	0.880	0.779	0.558	0.430	0.411	0.566
	Aprox.	0.998	0.967	0.876	0.722	0.508	0.410	0.373	0.347
110	Simul.	0.997	0.967	0.909	0.803	0.591	0.471	0.449	0.592
	Aprox.	0.999	0.979	0.905	0.763	0.547	0.443	0.404	0.376
120	Simul.	0.999	0.982	0.929	0.844	0.629	0.502	0.476	0.630
	Aprox.	1.000	0.987	0.928	0.799	0.584	0.476	0.434	0.405

Tabla 4 Niveles de potencia simulada (evaluados en  $\rho = \sigma/\sigma_0 = 1.25$ ) de una prueba bilateral para la varianza basada en el método de Bonett en comparación con los niveles de potencia teórica (aproximación a la normal). Las muestras se generaron a partir de distribuciones asimétricas con colas casi normales, moderadamente pesadas o pesadas.

n	Potencia	Distribuciones asimétricas con colas casi normales o moderadamente pesadas			Distribuciones asimétricas con colas pesadas	
		Chi(10)	B(8,1)	Chi(5)	Chi(3)	Exp
20	Simul.	0.222	0.166	0.172	0.139	0.128
	Aprox.	0.186	0.152	0.149	0.123	0.104
30	Simul.	0.314	0.216	0.234	0.190	0.151
	Aprox.	0.263	0.263	0.205	0.164	0.135
40	Simul.	0.387	0.266	0.292	0.223	0.186
	Aprox.	0.338	0.266	0.261	0.204	0.165

<i>n</i>	Potencia	Distribuciones asimétricas con colas casi normales o moderadamente pesadas			Distribuciones asimétricas con colas pesadas	
		Chi(10)	B(8,1)	Chi(5)	Chi(3)	Exp
<b>50</b>	<b>Simul.</b>	0.455	0.324	0.349	0.263	0.208
	<b>Aprox.</b>	0.409	0.323	0.316	0.245	0.196
<b>60</b>	<b>Simul.</b>	0.521	0.376	0.399	0.302	0.239
	<b>Aprox.</b>	0.477	0.377	0.369	0.286	0.227
<b>70</b>	<b>Simul.</b>	0.583	0.419	0.463	0.361	0.269
	<b>Aprox.</b>	0.539	0.430	0.420	0.325	0.257
<b>80</b>	<b>Simul.</b>	0.646	0.473	0.499	0.394	0.299
	<b>Aprox.</b>	0.597	0.479	0.469	0.365	0.288
<b>90</b>	<b>Simul.</b>	0.688	0.517	0.561	0.428	0.327
	<b>Aprox.</b>	0.649	0.526	0.516	0.403	0.318
<b>100</b>	<b>Simul.</b>	0.738	0.561	0.591	0.469	0.368
	<b>Aprox.</b>	0.695	0.571	0.560	0.440	0.347
<b>110</b>	<b>Simul.</b>	0.779	0.608	0.637	0.495	0.394
	<b>Aprox.</b>	0.737	0.611	0.600	0.475	0.376
<b>120</b>	<b>Simul.</b>	0.810	0.635	0.679	0.538	0.416
	<b>Aprox.</b>	0.774	0.650	0.638	0.509	0.405

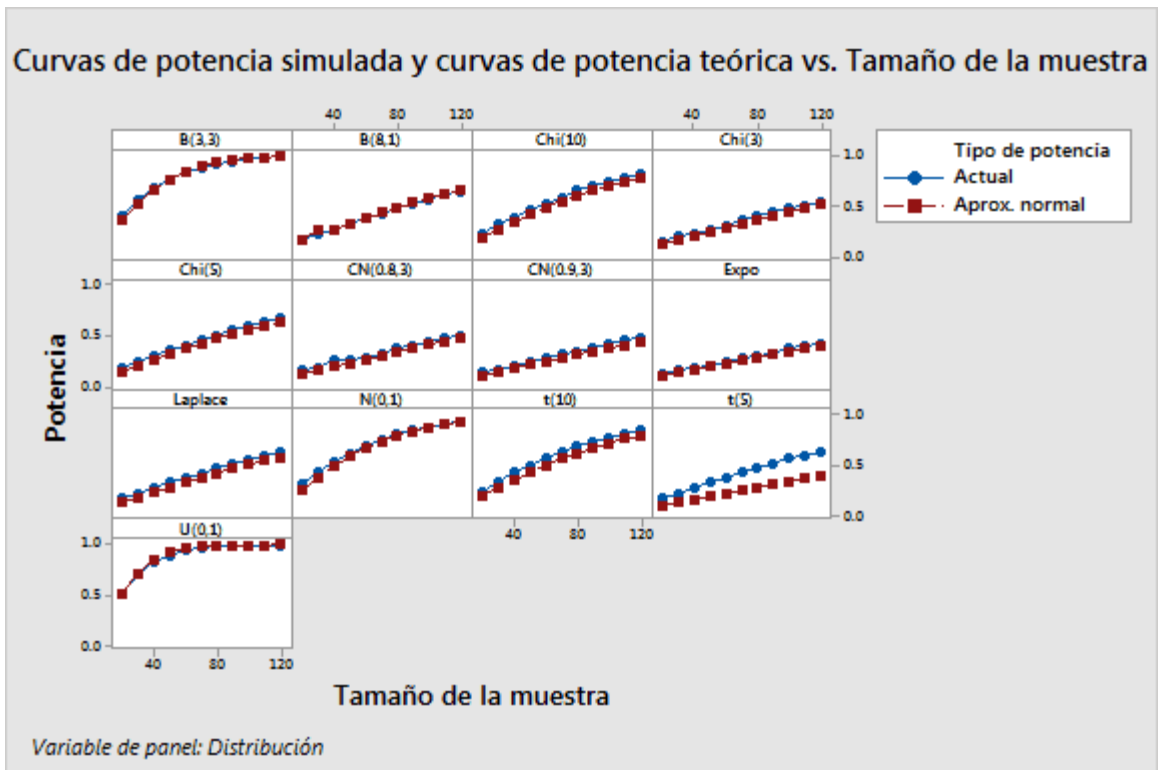


Figura 2 Curvas de potencia simulada en comparación con las curvas de potencia teórica para diferentes distribuciones

Los resultados que se muestran en las tablas 3 y 4 y en la figura 2 indican que cuando las muestras se obtienen de distribuciones con colas más livianas (distribuciones tipo L, tal como se define en el Apéndice C), como la distribución uniforme, la distribución Beta (3,3), la distribución normal, la distribución t con 10 grados de libertad y la distribución de chi-cuadrado con 10 grados de libertad, los valores de potencia teórica y los niveles de potencia simulada son prácticamente indistinguibles.

Sin embargo, para distribuciones con colas pesadas (distribuciones tipo H), las curvas de potencia simulada están considerablemente por encima de las curvas de potencia teórica cuando las muestras son pequeñas. Estas distribuciones de colas pesadas incluyen la distribución t con 5 grados de libertad, la distribución de Laplace, la distribución de chi-cuadrado con 3 grados de libertad, la distribución exponencial y las dos distribuciones normales contaminadas, CN(0.9,3) y CN(0.8,3). Por lo tanto, al planificar el tamaño de la muestra para un estudio y la muestra proviene de una distribución con colas pesadas, el tamaño de la muestra estimado por la función de potencia teórica podría ser más grande que el tamaño de la muestra real necesario para lograr una determinada potencia objetivo.

# Apéndice E: La prueba SJ para colas normales en comparación con colas pesadas

Los resultados del estudio de simulación descrito en el Apéndice C revelaron que cuando las colas de la distribución son más pesadas, se necesitan tamaños de muestra más grandes para que la probabilidad de cobertura simulada de los intervalos de confianza de Bonett se acerquen al nivel objetivo. Sin embargo, la asimetría no parecía tener un efecto significativo sobre las probabilidades de cobertura simulada.

Por lo tanto, teníamos que crear un criterio para evaluar la validez del método de Bonett basado en el tamaño de la muestra y en el peso de las colas de la distribución de la cual se extrajo la muestra. Afortunadamente, Gel et al. (2007) proporcionan una prueba razonablemente potente para probar directamente la hipótesis nula de que la distribución tiene colas normales con respecto a la hipótesis alternativa de que la distribución tiene colas pesadas. La prueba, a la cual nos referimos como la prueba SJ, se basa en el siguiente estadístico:

$$\hat{R} = \frac{S}{\hat{j}}$$

donde  $S$  es la desviación estándar de la muestra,  $\hat{j}$  es la estimación de la desviación absoluta de la media de muestra de la mediana,  $m$ , y se calcula como

$$\hat{j} = \frac{\sqrt{\pi/2}}{n} \sum_{i=1}^n |X_i - m|$$

Una prueba con un tamaño aproximado de  $\alpha$  con respecto a la hipótesis alternativa de colas pesadas rechaza la hipótesis nula de colas normales si

$$\frac{\sqrt{n}(\hat{R} - 1)}{\sigma_R} \geq z_\alpha$$

donde  $z_\alpha$  es el percentil  $\alpha$  superior de una distribución normal estándar y  $\sigma_R = (\pi - 3)/2$ .

Gel et al. (2007) han demostrado que, al reemplazar el percentil  $\alpha$  superior de la distribución normal estándar por el de la distribución t con  $(\sqrt{n} + 3)/2$  grados de libertad, se obtienen mejores aproximaciones para los tamaños de muestra moderados. Por lo tanto, al aplicar la prueba SJ para la validez de la prueba en las verificaciones de los datos, reemplazamos  $z_\alpha$  por  $t_{d,\alpha}$ , el percentil  $\alpha$  superior de la distribución t con  $d = (\sqrt{n} + 3)/2$  grados de libertad.

# Apéndice F: Validez de la prueba

## Simulación F1: Uso de la potencia simulada de la prueba SJ para determinar clasificaciones de las distribuciones

Realizamos simulaciones para investigar la potencia de la prueba SJ. Generamos muestras de diversos tamaños ( $n = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200$ ) de diferentes distribuciones. Las distribuciones tenían colas normales, livianas, moderadas o pesadas, y son las mismas que se describen en la simulación C1 (consulte el Apéndice C). Para cada tamaño de muestra dado, se extrajeron 10,000 réplicas de muestra de cada distribución. Calculamos la potencia simulada de la prueba SJ como la proporción de casos para los cuales se rechazó la hipótesis nula (de que la distribución original tiene colas normales). Además, calculamos los valores  $R$  promedio (AveR) y los valores  $p$  promedio (AvePV).

Los resultados de la simulación se muestran a continuación, en las tablas 5 y 6.

**Tabla 5** Niveles de potencia simulada de la prueba SJ. Las muestras se generaron a partir de distribuciones simétricas con colas livianas, normales, casi normales o pesadas.

Distribución		Distribuciones simétricas con colas livianas, normales o casi normales				Distribuciones simétricas con colas pesadas			
		U(0,1)	B(3,3)	N(0,1)	t(10)	Lpl	CN (.8,3)	CN (.9,3)	t(5)
$n$	Rreal	0.921	0.965	1.0	1.032	1.128	1.152	1.118	1.085
10	Potencia	0.021	0.041	0.075	0.103	0.249	0.264	0.198	0.161
	AveR	1.010	1.036	1.060	1.073	1.129	1.131	1.106	1.096
	AvePV	0.482	0.401	0.341	0.314	0.219	0.228	0.272	0.278
15	Potencia	0.009	0.027	0.071	0.121	0.350	0.389	0.283	0.215
	AveR	0.986	1.018	1.043	1.063	1.130	1.140	1.110	1.093
	AvePV	0.572	0.440	0.357	0.302	0.171	0.181	0.240	0.247
20	Potencia	0.002	0.016	0.066	0.144	0.428	0.465	0.331	0.253
	AveR	0.966	1.001	1.030	1.054	1.127	1.137	1.104	1.086
	AvePV	0.669	0.503	0.382	0.311	0.147	0.161	0.236	0.244

Distribución		Distribuciones simétricas con colas livianas, normales o casi normales				Distribuciones simétricas con colas pesadas			
		U(0,1)	B(3,3)	N(0,1)	t(10)	Lpl	CN (.8,3)	CN (.9,3)	t(5)
<i>n</i>	Rreal	0.921	0.965	1.0	1.032	1.128	1.152	1.118	1.085
25	Potencia	0.002	0.011	0.065	0.153	0.500	0.550	0.397	0.293
	AveR	0.959	0.995	1.025	1.050	1.128	1.141	1.107	1.086
	AvePV	0.721	0.535	0.391	0.305	0.120	0.128	0.208	0.223
30	Potencia	0.001	0.010	0.060	0.170	0.561	0.603	0.431	0.334
	AveR	0.951	0.989	1.019	1.046	1.127	1.141	1.106	1.084
	AvePV	0.773	0.570	0.409	0.304	0.103	0.112	0.197	0.209
40	Potencia	0.000	0.006	0.058	0.190	0.665	0.709	0.513	0.401
	AveR	0.944	0.984	1.015	1.043	1.126	1.145	1.109	1.084
	AvePV	0.840	0.616	0.420	0.287	0.073	0.076	0.162	0.179
50	Potencia	0.000	0.004	0.058	0.208	0.746	0.785	0.590	0.462
	AveR	0.939	0.980	1.012	1.040	1.126	1.146	1.111	1.084
	AvePV	0.886	0.654	0.427	0.279	0.053	0.055	0.131	0.156
60	Potencia	0.000	0.002	0.060	0.231	0.813	0.836	0.647	0.518
	AveR	0.936	0.978	1.010	1.039	1.127	1.146	1.112	1.084
	AvePV	0.913	0.686	0.430	0.267	0.039	0.039	0.109	0.134
70	Potencia	0.000	0.002	0.054	0.247	0.863	0.879	0.702	0.554
	AveR	0.934	0.975	1.009	1.037	1.127	1.147	1.112	1.083
	AvePV	0.935	0.716	0.437	0.259	0.028	0.029	0.091	0.123
80	Potencia	0.000	0.001	0.054	0.265	0.896	0.912	0.729	0.591
	AveR	0.933	0.974	1.007	1.037	1.128	1.147	1.111	1.083
	AvePV	0.950	0.740	0.440	0.241	0.021	0.021	0.079	0.105
90	Potencia	0.000	0.001	0.054	0.281	0.933	0.934	0.771	0.633
	AveR	0.932	0.973	1.007	1.036	1.128	1.148	1.113	1.083
	AvePV	0.962	0.759	0.445	0.237	0.014	0.016	0.067	0.093

Distribución		Distribuciones simétricas con colas livianas, normales o casi normales				Distribuciones simétricas con colas pesadas			
		U(0,1)	B(3,3)	N(0,1)	t(10)	Lpl	CN (.8,3)	CN (.9,3)	t(5)
<i>n</i>	Rreal	0.921	0.965	1.0	1.032	1.128	1.152	1.118	1.085
100	Potencia	0.000	0.001	0.057	0.301	0.947	0.954	0.805	0.661
	AveR	0.930	0.972	1.006	1.036	1.127	1.148	1.113	1.083
	AvePV	0.971	0.779	0.446	0.224	0.012	0.011	0.055	0.083
120	Potencia	0.000	0.000	0.052	0.334	0.974	0.974	0.852	0.732
	AveR	0.929	0.971	1.005	1.035	1.128	1.149	1.114	1.083
	AvePV	0.982	0.809	0.452	0.206	0.006	0.007	0.041	0.064
140	Potencia	0.000	0.000	0.052	0.336	0.986	0.988	0.894	0.785
	AveR	0.928	0.971	1.004	1.034	1.127	1.150	1.116	1.084
	AvePV	0.989	0.834	0.454	0.192	0.004	0.003	0.027	0.048
160	Potencia	0.000	0.000	0.054	0.402	0.993	0.992	0.916	0.819
	AveR	0.927	0.970	1.004	1.034	1.128	1.150	1.114	1.084
	AvePV	0.993	0.858	0.457	0.177	0.002	0.002	0.021	0.040
180	Potencia	0.000	0.000	0.052	0.416	0.998	0.996	0.934	0.853
	AveR	0.926	0.969	1.003	1.034	1.128	1.149	1.115	1.084
	AvePV	0.995	0.874	0.461	0.167	0.001	0.001	0.016	0.033
200	Potencia	0.000	0.000	0.053	0.448	0.998	0.998	0.954	0.884
	AveR	0.926	0.969	1.003	1.034	1.127	1.150	1.116	1.083
	AvePV	0.997	0.890	0.461	0.153	0.001	0.001	0.011	0.025



Tabla 6 Niveles de potencia simulada de la prueba SJ. Las muestras se generaron a partir de distribuciones asimétricas con colas casi normales, moderadamente pesadas o pesadas.

Distribuciones		Distribuciones asimétricas con colas casi normales o moderadamente pesadas			Distribuciones asimétricas con colas pesadas	
		Chi(10)	B(8,1)	Chi(5)	Chi(3)	Exp
<i>n</i>	Rreal	1.028	1.075	1.059	1.098	1.151
10	Potencia	0.120	0.213	0.161	0.218	0.283
	AveR	1.072	1.105	1.088	1.108	1.136
	AvePV	0.326	0.284	0.304	0.279	0.251
15	Potencia	0.139	0.270	0.205	0.292	0.377
	AveR	1.062	1.105	1.082	1.110	1.141
	AvePV	0.320	0.261	0.286	0.245	0.209
20	Potencia	0.152	0.295	0.223	0.328	0.449
	AveR	1.051	1.089	1.070	1.101	1.142
	AvePV	0.335	0.260	0.296	0.242	0.186
25	Potencia	0.160	0.336	0.255	0.374	0.515
	AveR	1.043	1.084	1.068	1.101	1.144
	AvePV	0.337	0.236	0.281	0.219	0.156
30	Potencia	0.171	0.370	0.285	0.414	0.564
	AveR	1.043	1.084	1.065	1.097	1.142
	AvePV	0.329	0.228	0.274	0.206	0.139
40	Potencia	0.193	0.440	0.331	0.490	0.651
	AveR	1.039	1.085	1.064	1.098	1.143
	AvePV	0.321	0.188	0.246	0.171	0.106
50	Potencia	0.215	0.484	0.370	0.556	0.720
	AveR	1.037	1.081	1.064	1.100	1.143
	AvePV	0.314	0.173	0.220	0.140	0.080

Distribuciones		Distribuciones asimétricas con colas casi normales o moderadamente pesadas			Distribuciones asimétricas con colas pesadas	
		Chi(10)	B(8,1)	Chi(5)	Chi(3)	Exp
<i>n</i>	Rreal	1.028	1.075	1.059	1.098	1.151
60	Potencia	0.224	0.527	0.395	0.607	0.778
	AveR	1.035	1.079	1.062	1.099	1.146
	AvePV	0.303	0.152	0.208	0.119	0.062
70	Potencia	0.241	0.568	0.438	0.648	0.822
	AveR	1.034	1.079	1.061	1.098	1.146
	AvePV	0.292	0.134	0.191	0.104	0.048
80	Potencia	0.259	0.612	0.474	0.689	0.855
	AveR	1.034	1.079	1.062	1.098	1.148
	AvePV	0.280	0.115	0.170	0.089	0.036
90	Potencia	0.284	0.643	0.501	0.733	0.890
	AveR	1.034	1.079	1.060	1.099	1.148
	AvePV	0.270	0.104	0.163	0.075	0.028
100	Potencia	0.285	0.675	0.527	0.757	0.912
	AveR	1.032	1.078	1.060	1.098	1.147
	AvePV	0.267	0.094	0.151	0.067	0.022
120	Potencia	0.323	0.728	0.572	0.816	0.942
	AveR	1.032	1.077	1.060	1.098	1.149
	AvePV	0.246	0.074	0.129	0.050	0.014
140	Potencia	0.344	0.769	0.621	0.852	0.963
	AveR	1.031	1.077	1.060	1.099	1.148
	AvePV	0.232	0.060	0.112	0.036	0.009
160	Potencia	0.363	0.815	0.666	0.887	0.978
	AveR	1.031	1.077	1.060	1.098	1.150
	AvePV	0.217	0.047	0.093	0.027	0.005

Distribuciones		Distribuciones asimétricas con colas casi normales o moderadamente pesadas			Distribuciones asimétricas con colas pesadas	
		Chi(10)	B(8,1)	Chi(5)	Chi(3)	Exp
<i>n</i>	Rreal	1.028	1.075	1.059	1.098	1.151
180	Potencia	0.385	0.843	0.692	0.910	0.986
	AveR	1.031	1.077	1.059	1.099	1.148
	AvePV	0.209	0.039	0.083	0.021	0.004
200	Potencia	0.410	0.877	0.727	0.931	0.989
	AveR	1.030	1.077	1.059	1.098	1.149
	AvePV	0.196	0.030	0.071	0.016	0.003

Los resultados de nuestra simulación presentados en las tablas 5 y 6 son consistentes con los publicados en Gel et al. (2007). Cuando las muestras provienen de poblaciones normales, los niveles de potencia simulada (que en este caso representan el nivel de significancia real de la prueba) no están lejos el nivel objetivo, incluso para tamaños de muestra tan bajos como 25. Cuando las muestras provienen de distribuciones de colas pesadas, la potencia de la prueba es baja para los tamaños de muestra pequeños, pero aumenta a por lo menos 40% cuando el tamaño de la muestra llega a 40. Específicamente, con un tamaño de muestra de 40, la potencia es de 40.1% para la distribución t con 5 grados de libertad, 66.5% para la distribución de Laplace y 65.1% de la distribución exponencial.

Para las distribuciones con colas livianas (las distribuciones Beta(3,3) y uniforme), la potencia de la prueba es cerca de 0 para muestras pequeñas y disminuye aún más a medida que aumenta el tamaño de la muestra. Esto no causa sorpresa, porque la evidencia con respecto a estas distribuciones en realidad apoya la hipótesis alternativa de una distribución con colas más livianas, en lugar de la hipótesis alternativa de una distribución con colas más pesadas.

Cuando las muestras provienen de distribuciones con colas ligeramente más pesadas, como la distribución t con 10 grados de libertad o la distribución de chi-cuadrado con 10 grados de libertad, los niveles de potencia son bajos para los tamaños de muestra de moderados a grandes. Para nuestros propósitos, en realidad se trata de un buen resultado porque la prueba para una varianza (desviación estándar) funciona adecuadamente para estas distribuciones y no queremos que estas distribuciones se marquen como distribuciones de colas pesadas. Sin embargo, a medida que aumenta el tamaño de la muestra, la potencia de la prueba aumenta, así que estas distribuciones de colas ligeramente pesadas se detectan como distribuciones de colas pesadas.

Por lo tanto, las reglas para evaluar el peso de la cola de la distribución para esta prueba también deben considerar el tamaño de la muestra. Un enfoque para hacer esto es calcular un

intervalo de confianza para la medición del peso de la cola; sin embargo, la distribución del estadístico SJ es extremadamente sensible a la distribución original de la muestra. Un enfoque alternativo consiste en evaluar el peso de las colas de la distribución con base en la fuerza del rechazo de la hipótesis nula de la prueba SJ y en el tamaño de la muestra. Más específicamente, los valores p más pequeños indican colas más pesadas y los valores p más grandes indican colas más livianas. Sin embargo, las muestras más grandes tienden a tener valores p más pequeños que las muestras más pequeñas. Por lo tanto, con base en los niveles de potencia simulada, los tamaños de las muestras y los valores p promedio de la tabla 3, ideamos un conjunto general de reglas para evaluar las colas de una distribución para cada muestra usando la prueba SJ.

Para tamaños de muestra de moderados a grandes ( $40 \leq n \leq 100$ ), si el valor p está entre 0.01 y 0.05, consideramos que existe evidencia moderada en contra de la hipótesis nula. Es decir, la distribución de la muestra se clasifica como una distribución (tipo M) con colas moderadamente pesadas. Por otro lado, si el valor p está por debajo de 0.01, entonces existe fuerte evidencia en contra de la hipótesis nula y la distribución original de la muestra se clasifica como una distribución con colas pesadas (tipo H).

Para muestras grandes ( $n > 100$ ), podemos clasificar la distribución original como una distribución tipo M si el valor p está entre 0.005 y 0.01 y como una distribución tipo H si el valor p es extremadamente pequeño (menos de 0.005). Tenga en cuenta que cuando el tamaño de la muestra está por debajo de 40, por lo general la potencia de la prueba SJ es demasiado baja para determinar eficazmente la distribución de la muestra.

Las reglas generales de clasificación para la validez de la prueba de 1 varianza usando el método de Bonett se resumen a continuación, en la tabla 7.

**Tabla 7** Reglas de clasificación para identificar la distribución original de cada muestra ( $p$  es el valor p de la prueba SJ)

Condición	Tipo de distribución
$n < 40$	No se determinó ninguno
$100 \geq n \geq 40$ y $p > 0.05$	Distribución tipo L
$n > 100$ y $p > 0.01$	Distribución tipo L
$40 \leq n \leq 100$ y $0.01 < p \leq 0.05$	Distribución tipo M
$n > 100$ y $0.005 < p \leq 0.01$	Distribución tipo M
$40 \leq n \leq 100$ y $p \leq 0.01$	Distribución tipo H
$n > 100$ y $p \leq 0.005$	Distribución tipo H

Como se indicó anteriormente, con base en los resultados de las tablas 1 y 2 en la simulación C1, el tamaño mínimo aproximado de la muestra que se necesita para lograr un mínimo de probabilidad de cobertura de 0.93 cuando las muestras provienen de una distribución tipo L,

tipo M y tipo H es de 20, 80 y 200, respectivamente. Sin embargo, debido a que la potencia de la prueba SJ es baja para muestras pequeñas, el requisito de tamaño mínimo de la muestra para las distribuciones tipo L se establece en 40.

## Simulación F2: Verificación de las reglas para clasificar las distribuciones

Generado muestras de algunas de las distribuciones descritas en la simulación C1 y utilizamos la prueba SJ para determinar las proporciones de las muestras que se clasificaron en uno de los tres grupos de distribuciones: tipo L, tipo M y tipo H. Los resultados de la simulación se muestran en la tabla 8.

**Tabla 8** Fracción de las 10,000 muestras de diferentes tamaños provenientes de diversas distribuciones que se identifican como tipo L, tipo M y tipo H

<i>n</i>	Distribuciones	Tipo L				Tipo M	Tipo H	
		B(3,3)	N(0,1)	t(10)	Chi(10)	Chi(5)	Lpl	ExpExpP
40	%Tipo L	99.6	94.0	81.5	80.3	66.6	33.0	34.4
	%Tipo M	0.4	5.5	14.0	14.0	20.0	31.9	22.9
	%Tipo H	0.0	0.5	4.5	5.7	13.4	35.1	42.8
50	%Tipo L	99.7	94.4	78.7	79.1	64.0	25.1	28.0
	%Tipo M	0.3	5.1	15.6	14.2	20.0	29.9	20.7
	%Tipo H	0.0	0.5	5.7	6.7	16.0	45.0	51.3
60	%Tipo L	99.7	94.5	77.3	77.3	59.1	18.5	22.6
	%Tipo M	0.3	5.1	16.4	15.0	22.0	27.4	19.2
	%Tipo H	0.0	0.5	6.3	7.7	18.9	54.1	58.2
70	%Tipo L	99.8	94.4	74.5	75.2	55.9	14.0	18.1
	%Tipo M	0.2	5.0	18.1	16.0	22.2	24.0	17.5
	%Tipo H	0.0	0.6	7.4	8.8	21.9	62.0	64.4
80	%Tipo L	99.9	94.3	74.1	74.4	53.0	10.0	13.9
	%Tipo M	0.1	5.1	17.8	16.7	22.8	21.0	15.5
	%Tipo H	0.0	0.6	8.2	8.9	24.2	69.0	70.6
90	%Tipo L	99.9	94.4	71.2	72.1	49.5	7.5	11.1
	%Tipo M	0.1	5.0	19.1	17.2	22.6	16.5	13.7
	%Tipo H	0.0	0.6	9.7	10.7	27.9	76.0	75.3

n	Distribuciones	Tipo L				Tipo M	Tipo H	
		B(3,3)	N(0,1)	t(10)	Chi(10)	Chi(5)	Lpl	ExpExp P
100	%Tipo L	99.9	94.5	70.8	70.3	47.3	4.8	8.9
	%Tipo M	0.1	4.9	19.5	17.9	22.7	14.3	11.8
	%Tipo H	0.0	0.6	9.7	11.8	30.0	80.9	79.4
120	%Tipo L	100.0	99.4	87.4	87.2	64.8	12.0	14.4
	%Tipo M	0.0	0.4	5.0	4.5	7.9	7.8	5.6
	%Tipo H	0.0	0.2	7.6	8.4	27.4	80.4	80.0
140	%Tipo L	100.0	99.3	86.0	85.1	60.5	7.0	9.9
	%Tipo M	0.0	0.5	5.2	5.0	8.6	5.6	4.1
	%Tipo H	0.0	0.2	8.8	9.9	30.9	87.4	86.0
160	%Tipo L	100.0	99.4	83.4	83.0	55.6	4.0	6.9
	%Tipo M	0.0	0.5	6.3	5.8	9.5	3.5	3.0
	%Tipo H	0.0	0.1	10.4	11.2	34.9	92.5	90.1
180	%Tipo L	100.0	99.3	81.1	81.7	51.0	2.5	4.6
	%Tipo M	0.0	0.5	6.8	5.9	9.4	1.9	2.2
	%Tipo H	0.0	0.2	12.1	12.4	39.6	95.6	93.2
200	%Tipo L	100.0	99.5	79.0	80.5	47.2	1.3	3.0
	%Tipo M	0.0	0.4	7.6	6.1	9.4	1.6	1.7
	%Tipo H	0.0	0.1	13.4	13.4	43.4	97.1	95.3

Los resultados de la tabla 8 muestran que cuando las muestras provienen de distribuciones de colas livianas (tipo L) y de colas pesadas (tipo H), una mayor proporción de las muestras se clasifica de manera correcta. Por ejemplo, cuando se generaron muestras con un tamaño de 40 a partir de la distribución Beta(3,3), el 99.6% de las muestras se clasificó correctamente como una distribución con colas más livianas; cuando se generaron muestras con un tamaño de 90 a partir de la distribución de Laplace, el 76.0% se clasificó correctamente como una distribución con colas pesadas. Como resultado, los mensajes de advertencia en la Tarjeta de informe acerca de la validez de la prueba no se emiten de manera errónea cuando las muestras provienen realmente de distribuciones con colas más livianas y se emiten correctamente cuando la muestra proviene de una distribución con colas pesadas y no se cumple el requisito del tamaño mínimo de la muestra. Además, para las muestras provenientes de distribuciones con colas moderadamente pesadas (tipo M), como la distribución de chi-cuadrado (5), una mayor proporción de las muestras se clasifican erróneamente como de colas livianas (tipo L) cuando las muestras son pequeñas (para un tamaño de muestra de 40, el 66% de las muestras se clasifican erróneamente como una distribución de colas livianas). En consecuencia, para estos casos, los

mensajes de advertencia en la Tarjeta de informe podrían no emitirse aun cuando las distribuciones originales tengan colas moderadamente pesadas. Sin embargo, cuando el tamaño de la muestra es mayor que 80, la clasificación errónea como una distribución tipo L no tiene ningún efecto, porque ya se cumplió el requisito de tamaño mínimo de la muestra.

© 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](http://www.minitab.com). All other marks referenced remain the property of their respective owners.