

Teste de desvio padrão para 1 amostra

Visão geral

O teste de desvio padrão para 1 amostra é utilizada para estimar a variabilidade do processo e para comparar a variabilidade com um valor alvo. Em geral, a variabilidade é medida utilizando-se a variância ou, de forma equivalente, o desvio padrão.

Muitos métodos estatísticos foram desenvolvidos para avaliar a variância de uma população, cada um com suas próprias pontos fortes e limitações. O método do qui-quadrado clássico usado para testar a variância é provavelmente o mais comumente utilizado, mas é extremamente sensível à suposição de normalidade e pode produzir resultados extremamente imprecisos quando os dados são assimétricos ou com caudas pesadas. Outros métodos foram desenvolvidos, mas eles também apresentam inconvenientes. Por exemplo, alguns métodos são válidos apenas para amostras grandes ou para os dados de distribuição simétrica (consulte Anexo A).

No Minitab 15, utilizamos um método de amostra grande alternativa que extraímos a partir de uma aproximação do qui-quadrado escalonado para a distribuição da variância da amostra por Box (1953). Este método, conhecido como os graus de liberdade ajustados método (AdjDF), é menos sensível à suposição normal para amostras suficientemente grandes e demonstrou capacidade de produzir intervalos de confiança mais precisos do que outros métodos (Hummel, Banga, & Hettmansperger, 2005) . Mais recentemente, no entanto, foi desenvolvido um método estatístico revisado por Bonett (2006) que parece proporcionar melhores intervalos de confiança aproximados.

Neste artigo, nós avaliamos o desempenho do método de Bonett. Além disso, para o planejamento do tamanho da amostra, investigamos a função de poder para o procedimento de teste equivalente associado aos intervalos de confiança de Bonett. Com base em nossos resultados, usamos o método de Bonett para o teste de desvio padrão para 1 amostra no Assistente. Também examinamos as seguintes verificações de dados que são executadas automaticamente e exibidas no Cartão de Relatório do Assistente e explicam como elas afetam os resultados:

- Dados incomuns
- Validade do teste
- Tamanho amostral

Métodos de desvio padrão para 1 amostra

Método de Bonett comparado com o método AdjDF

Antes da publicação do método de Bonett (2006), o processo mais robusto para realizar inferências sobre a variância de uma população era, muito provavelmente, o método AdjDF. Os resultados publicados por Bonett, no entanto, mostram que o método de Bonett fornece os níveis de confiança estáveis que estão perto do nível alvo quando amostras de tamanho moderado são extraídas de populações não normais. Portanto, o método de Bonett pode ser preferível para realizar inferências sobre o desvio padrão ou a variância de uma população.

Objetivo

Desejávamos comparar o desempenho do método de Bonett com o método AdjDF ao fazer inferências sobre a variância de uma única população. Especificamente, queríamos determinar qual método produz intervalos de confiança mais precisos para a variância (ou o desvio padrão) quando as amostras de vários tamanhos são geradas a partir de populações não normais.

Comparamos os intervalos de confiança porque o método de Bonett se aplica diretamente a intervalos de confiança. É possível obter o procedimento de teste de hipóteses equivalentes associado com intervalos de confiança de Bonett. No entanto, para comparar diretamente os nossos resultados com os resultados publicados em Bonett (2006), examinamos os intervalos de confiança em vez de os testes de hipóteses.

Método

O método AdjDF e método de Bonett estão ambos formalmente definidos no Anexo B. Para comparar a precisão dos intervalos de confiança para cada método, foram realizadas as simulações a seguir. Em primeiro lugar, nós geramos amostras aleatórias de vários tamanhos a partir de distribuições com propriedades diferentes, como distribuições assimétricas e com caudas pesadas, simétricas e com caudas pesadas, e simétricas e com cauda leve. Para cada tamanho de amostra, 10.000 replicações de amostras foram extraídas de cada distribuição, e foram calculados de intervalos de confiança de 95% bilateral para a verdadeira variância de distribuição usando cada um dos métodos. Em seguida, calculamos a proporção dos 10.000 intervalos que continham a verdadeira variância, conhecida como a probabilidade de cobertura simulada. Se os intervalos de confiança forem precisos, a probabilidade de cobertura simulada deve estar próxima sa probabilidade de cobertura alvo de 0,95. Além disso, calculamos as larguras médias associadas aos intervalos de confiança para cada um dos métodos. Se os intervalos de confiança dos dois métodos têm aproximadamente as mesmas probabilidades de cobertura simuladas, o método que produz intervalos mais curtos (em média) é mais preciso. Para obter mais detalhes, consulte o Anexo C.

Resultados

O método de Bonett geralmente produz melhores probabilidades de cobertura e intervalos de confiança mais precisos do que o método AdjDF. Como resultado, os testes estatísticos para a variância com base no método de Bonett geram as taxas de erro Tipo I e tipo II menores. Por esta razão, o teste de desvio padrão para 1 amostra no Assistente é baseado no método de Bonett.

Além disso, nossos resultados mostram que, se a distribuição tem caudas de moderadas a pesadas, o método de Bonett exige amostras maiores para atingir o nível alvo de precisão:

- Para distribuições com caudas normais ou leves, um tamanho de amostra de 20 é suficiente.
- Para distribuições com caudas moderadamente pesadas, o tamanho da amostra deve ser de pelo menos 80.
- Para distribuições com caudas pesadas, o tamanho da amostra deve ser de pelo menos 200.

Portanto, para garantir que os resultados do teste de desvio padrão para 1 amostra ou dos intervalos de confiança sejam válidos para os seus dados, o Assistente inclui uma verificação de dados para avaliar simultaneamente o tamanho da amostra e as caudas da distribuição de dados (consulte a Validade da verificação dos dados de teste abaixo).

Desempenho do poder teórico

O método de Bonett se aplica diretamente a intervalos de confiança para a variância (ou desvio padrão). No entanto, com a relação estatística entre testes de hipóteses e intervalos de confiança, podemos extrair o teste equivalente que está associado aos intervalos de confiança aproximados de Bonett. Como não existe uma função de potência exata disponível para este teste, era necessário que nós a extraíssemos. Além disso, queríamos avaliar a sensibilidade da função de poder teórica para a suposição de normalidade.

Objetivo

Desejávamos determinar se é possível usar a função de poder teórica do teste associado aos intervalos de confiança de Bonett para avaliar os requisitos de poder e de tamanho da amostra para o teste de desvio padrão para 1 amostra no Assistente. Para fazer isso, foi necessário avaliar se esta função de poder teórico reflete com precisão o poder real do teste quando são analisados dados normais e não normais.

Método

A função de poder teórico do teste utilizando o método de Bonett é extraída no Anexo C. Realizamos simulações para estimar os níveis reais de poder (aos quais nos referimos como níveis de poder simulados) usando o método de Bonett. Em primeiro lugar, nós geramos amostras aleatórias de vários tamanhos a partir das distribuições descritas em estudos anteriores: distribuições assimétricas e com caudas pesadas, simétricas e com caudas pesadas, e simétrica e com cauda leve. Para cada distribuição, foi realizado o teste em cada uma das 10.000 réplicas de amostra. Para cada tamanho de amostra, foi calculado o poder

simulado do teste para detectar uma determinada diferença, como por exemplo, a fração das 10.000 amostras para as quais o teste é significativo. Para efeito de comparação, também calculamos o nível de poder correspondente quando usada a função de poder teórico do teste. Se a função de poder teórico não for muito sensível à normalidade, os níveis de poder teóricos e simulados devem estar próximos aos dados normais e não normais. Para obter mais detalhes, consulte o Anexo D.

Resultados

Nossas simulações mostraram que, quando a amostra for proveniente de uma distribuição normal ou com caudas leves, o poder teórico e simulado do teste utilizando o método de Bonett são quase iguais. Entretanto, quando a amostra for proveniente de uma distribuição com caudas pesadas, a função de poder teórico pode ser conservadora e superestimar o tamanho da amostra necessário para atingir um determinado poder. Portanto, a função de poder teórica do teste garante que o tamanho da amostra seja amplo o suficiente para detectar uma diferença importante em termos práticos para o desvio padrão, independentemente da distribuição. No entanto, se os dados forem provenientes de distribuições com caudas pesadas, o tamanho estimado da amostra pode ser maior do que o tamanho que é realmente necessário, o que pode levar a custos superiores aos necessários quando forem extraídos os itens de amostragem.

Verificações dos dados

Dados incomuns

Dados incomuns são valores de dados extremamente grandes ou pequenos, também conhecidos como outliers. Dados incomuns podem exercer forte influência sobre os resultados da análise e podem afetar as chances de encontrar resultados estatisticamente significativos, especialmente quando a amostra for pequena. Os dados incomuns podem indicar problemas com a coleta de dados, ou podem ser devidos a um comportamento incomum do processo que está sendo estudado. Portanto, esses pontos de dados são, muitas vezes, dignos de investigação e devem ser corrigidos quando possível.

Objetivo

Desejávamos desenvolver um método para procurar por valores de dados muito grandes ou muito pequenos em relação ao total da amostra e que possam afetar os resultados da análise.



Método

Desenvolvemos um método para procurar por dados não usuais com base no método descrito por Hoaglin, Iglewicz e Tukey (1986), que é usado para identificar outliers em boxplots.

Resultados

O assistente identifica um ponto de dados como incomum se ele for mais do que 1,5 vezes o intervalo interquartil além do quartil inferior ou superior da distribuição. Os quartis superiores e inferiores são o 25º e 75º percentis dos dados. O intervalo interquartil é a diferença entre os dois quartis. Este método funciona bem, mesmo quando há vários outliers porque torna possível a detecção de cada outlier específico.

Quando procura por dados incomuns, o Cartão de Relatório do Assistente exibe os indicadores de status a seguir no Cartão de Relatório:

Status	Condição
	Não há pontos de dados incomuns.
	Pelo menos um ponto de dados é incomum e pode exercer forte influência sobre os resultados.

Validade do teste

Na seção Métodos de desvio padrão para 1 amostra acima, mostramos que o método de Bonett geralmente proporciona melhores resultados do que o método AdjDF. No entanto, quando as caudas de uma distribuição são mais pesadas, o método de Bonett requer amostras maiores para obter resultados precisos. Assim, um método para avaliar a validade do teste deve se basear não só o tamanho da amostra, mas também no peso das caudas da distribuição hierárquicamente superior. Gel et al. (2007) desenvolveram um teste para determinar se uma amostra é proveniente de uma distribuição com caudas pesadas. Este teste, chamado o teste de SJ, é baseado na relação entre o desvio padrão da amostra (s) e o J estimador da cauda (para obter mais detalhes, consulte o Anexo E).

Objetivo

Para uma determinada amostra de dados, foi necessário desenvolver uma regra para avaliar a validade do método de Bonett avaliando o peso das caudas nos dados.


Método


Foram realizadas simulações para investigar o poder do teste de SJ para identificar as distribuições de caudas pesadas. Se o teste de SJ for potente para amostras moderadamente grandes, então ele poderá ser usado para discriminar entre as distribuições com caudas pesadas e caudas leves para os nossos propósitos. Para obter mais detalhes, consulte o Anexo F.

Resultados

Nossas simulações mostraram que, quando as amostras são grandes o suficiente, o teste de SJ pode ser usado para discriminar entre as distribuições de cauda pesada e cauda leve. Para amostras de tamanho moderado ou grande, os valores de p menores indicam caudas mais pesadas e valores de p maiores indicam caudas mais leves. No entanto, como as amostras maiores tendem a ter valores de p menores do que amostras menores, também consideramos o tamanho da amostra para determinar o peso das caudas. Por isso, desenvolvemos o nosso conjunto de regras para o Assistente classificar as caudas da distribuição para cada amostra com base no tamanho da amostra e no valor de p do teste de SJ. Para conhecer os intervalos específicos de valores de p e tamanhos de amostra associados a distribuições com caudas leves, moderadas e pesadas, consulte o Anexo F.

Com base nesses resultados, o Assistente do Cartão de Relatório exibe os seguintes indicadores de status para avaliar a validade do teste de 1 desvio padrão (método de Bonett) para seus dados da amostra:

Status	Condição
	Não há evidências de que sua amostra tenha caudas pesadas. O tamanho de sua amostra é grande o suficiente para procurar confiavelmente por essa condição. OU Sua amostra tem caudas moderadamente pesadas ou pesadas. No entanto, o tamanho da amostra é grande o suficiente para compensar, de modo que o valor de p deve ser preciso.

Status	Condição
	<p>Sua amostra tem caudas moderadamente pesadas ou pesadas. Sua amostra não é grande o suficiente para compensar. Use de cautela quando interpretar os resultados</p> <p>OU</p> <p>O tamanho de sua amostra não é grande o suficiente para procurar confiavelmente por caudas pesadas. Use de cautela quando interpretar os resultados</p>

Tamanho amostral

Normalmente, um teste de hipótese estatística é realizado para reunir evidências para rejeitar a hipótese nula de "nenhuma diferença". Se a amostra for pequena demais, o poder do teste pode não ser suficiente para detectar uma diferença que realmente existe, o que resulta num erro de Tipo II. Assim, é fundamental garantir que os tamanhos de amostra sejam suficientemente grandes para detectar diferenças importantes em termos práticos com alta probabilidade.

Objetivo

Se os dados não fornecem evidências suficientes contra a hipótese nula, queremos determinar se os tamanhos de amostra são grandes o suficiente para o teste detectar diferenças práticas de interesse com alta probabilidade. Embora o objetivo do planejamento do tamanho da amostra seja garantir que os tamanhos de amostra são grandes o suficiente para detectar diferenças importantes com grande probabilidade, as amostras não devem ser tão grandes de forma que diferenças insignificantes se tornem estatisticamente significativas com alta probabilidade.






Método

A análise do poder e do tamanho da amostra para o teste de desvio padrão para 1 amostra é baseado na função de poder teórico do teste. Esta função de poder fornece boas estimativas quando os dados têm caudas quase normais ou caudas leves, mas podem produzir estimativas conservadoras quando os dados têm caudas pesadas (consulte os resultados da simulação resumidos em Desempenho da Função Poder Teórico na seção Métodos de desvio padrão para 1 amostra acima).

Resultados

Quando os dados não proporcionam evidência suficiente contra a hipótese nula, o Assistente utiliza a função de poder do teste de aproximação normal para calcular as diferenças práticas que podem ser detectadas com probabilidade de 80% e de 90% para determinado tamanho de amostra. Além disso, se o usuário fornecer uma diferença prática de interesse específico, o assistente usa a função de poder do teste de aproximação normal para calcular o tamanho das amostras que produzem 80% e 90% de chance de detecção da diferença.

Para ajudar a interpretar os resultados, o Cartão de Relatório do Assistente para o teste de desvio padrão para 1 amostra exibe os indicadores de status a seguir durante a verificação de poder e tamanho da amostra:

Status	Condição
	<p>O teste encontra uma diferença entre o desvio padrão e o valor alvo, de modo que o poder não é um problema.</p> <p>OU</p> <p>O poder é suficiente. O teste não encontrou nenhuma diferença entre o desvio padrão e o valor alvo, mas a amostra é grande o suficiente para fornecer pelo menos 90% de chance de detectar a determinada diferença.</p>
	<p>O poder pode ser suficiente. O teste não encontrou nenhuma diferença entre o desvio padrão e o valor alvo, mas a amostra é grande o suficiente para fornecer de 80% a 90% de chance de detectar a determinada diferença. O tamanho da amostra necessária para atingir 90% de poder é informado.</p>
	<p>Talvez o poder não seja suficiente. O teste não encontrou nenhuma diferença entre o desvio padrão e o valor alvo, sendo que a amostra é grande o suficiente para fornecer de 60% a 80% de chance de detectar a determinada diferença. Os tamanhos das amostras necessárias para atingir 80% e 90% de poder são relatados.</p>
	<p>O poder não é suficiente (< 60%). O teste não encontrou nenhuma diferença entre o desvio padrão e o valor alvo, sendo que a amostra é grande o suficiente para fornecer pelo menos 60% de chance de detectar a determinada diferença. Os tamanhos das amostras necessárias para atingir 80% e 90% de poder são relatados.</p>
	<p>O teste não encontrou uma diferença entre o desvio padrão e o valor alvo. Você não especificou uma diferença prática a ser detectada. Dependendo dos seus dados, o relatório pode indicar as diferenças que você poderia detectar com 80% e 90% de chance, com base em seu tamanho da amostra e alfa.</p>

Referências

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., & Tibshirani, R. J. (1993). An introduction to the bootstrap. Boca Raton, FL: Chapman and Hall/CRC.

Gel, Y. R., Miao, W., & Gastwirth, J. L. (2007). Robust directed tests of normality against heavy-tailed alternatives. *Computational Statistics & Data Analysis*, 51, 2734-2746.

Hummel, R., Banga, S., & Hettmansperger, T.P. (2005). Better confidence intervals for the variance in a random sample. *Minitab Technical Report*.

Lee, S.J., & Ping, S. (1996). Testing the variance of symmetric heavy-tailed distributions. *Journal of Statistical Computation and Simulation*, 56, 39-52.

Anexo A: Métodos para testar variância (ou desvio padrão)

A tabela abaixo resume os pontos fortes e fracos associados a vários métodos para testar a variância.

Método	Comentário
Procedimento qui-quadrado clássico	Suposição extremamente sensível à normalidade. Mesmo pequenos desvios da normalidade podem produzir resultados imprecisos, independentemente de o quão grande é a amostra. Na verdade, quando os dados se desviam da normalidade, aumentar a dimensão da amostra diminui a precisão do processo.
O método de amostra grande com base na distribuição normal assintótica da transformação logarítmica da variância da amostra	Geralmente melhor do que o método do qui-quadrado clássico, mas requer amostras maiores para ser insensível à suposição de normalidade.
Método de amostras grandes com base na expansão de Edgeworth para testes unilaterais (cauda superior) Consulte Lee e Ping (1996).	Produz as taxas de erro tipo I aceitáveis, mas exige que os dados sejam provenientes de uma distribuição simétrica.
Método de amostras grandes com base numa aproximação da distribuição da variância da amostra por uma distribuição do Qui-quadrado escalonado. O método é citado como o método de graus de liberdade ajustadas (AdjDF). Consulte Hummel, Banga e Hettmansperger (2005).	Oferece melhor probabilidade de cobertura do que o método baseado na distribuição normal assintótica da transformação logarítmica da variância da amostra e o método bootstrap de aproximação não paramétrica ABC para os intervalos de confiança (Efron e Tibshirani, 1993). Usado para o teste de variância 1 no Minitab 15.
A distribuição assintótica ajustada de Bonett da transformação logarítmica da variância da amostra Consulte Bonett (2006).	Oferece boa probabilidade de cobertura para intervalos de confiança, até mesmo para amostras moderadamente grandes. No entanto, requer amostras muito maiores quando os dados são provenientes de distribuições com caudas pesadas. Usado para o teste de 1 variância e o teste de desvio padrão para 1 amostra do Assistente no Minitab 16.

Anexo B: Definição do método de Bonette do método AdjDF

Permita que x_1, \dots, x_n seja uma amostra de tamanho aleatória observada n de uma população com um quarto momento finito. Permita que \bar{x} e s sejam a média de amostra e o desvio padrão observados, respectivamente. Além disso, permita que γ e γ_e sejam a curtose da população e o excesso de curtose, respectivamente, de modo que $\gamma_e = \gamma - 3$. Portanto, para uma população normal, $\gamma = 3$ e $\gamma_e = 0$. Da mesma forma, permita que σ^2 seja a variância da população desconhecida. Nas seções a seguir, apresentamos dois métodos para fazer uma inferência sobre o método de graus de liberdade ajustados (AdjDF) e método de Bonett.

Fórmula B1: Método AdjDF

O método AdjDF é baseado em uma aproximação da distribuição da variância da amostra por uma distribuição qui-quadrado escalonada (consulte Box, 1953). Mais especificamente, os dois primeiros momentos da variância da amostra são combinados com os momentos de uma distribuição qui-quadrado escalonada para determinar a escala desconhecida e os graus de liberdade. Esta abordagem oferece intervalo de confiança percentual bilateral $(1 - \alpha)100$ aproximado para a variância:

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

em que

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

Esta estimativa do excesso de curtose é idêntica à usada para os comandos de Estatística Básica do Minitab.

Fórmula B2: Método de Bonett

O método de Bonett depende da abordagem clássica bem conhecida, que utiliza o teorema do limite central e do método de Cramer δ para obter uma distribuição assintótica da transformação por log da variância da amostra. A transformação por log é usada para acelerar a convergência para a normalidade. Com essa abordagem, o intervalo de confiança percentual bilateral $(1 - \alpha)100$ aproximado para a variância é definido como:

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

onde z_α é o percentil superior da distribuição normal padrão, e se é uma estimativa assintótica do erro padrão da variação da amostra transformada por log, dada como:

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

Em trabalhos anteriores, Hummel et al. (2005) realizaram estudos de simulação que demonstraram que o método AdjDF é superior a esta abordagem clássica. No entanto, Bonett prevê duas adaptações para a abordagem clássica a fim de superar suas limitações. A primeira adaptação envia a estimativa da curtose. Para estimar a curtose, Bonett usa a seguinte fórmula:

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

em que m é uma média aparada com proporção de ajuste igual a $1/2\sqrt{n-4}$. Esta estimativa de curtose tende a melhorar a precisão dos níveis de confiança para distribuições com caudas pesadas (simétricas ou assimétricas).

Para o segundo ajuste, Bonett determina empiricamente um multiplicador constante para a variância da amostra e o erro padrão. Este multiplicador constante equaliza aproximadamente as probabilidades da cauda quando a amostra é pequena, e é dada como:

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

Esses ajustes produzem o intervalo de confiança percentual bilateral $(1 - \alpha)100$ aproximado de Bonett para a variância:

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

Anexo C: Precisão do método de Bonett contra o método AdjDF

Simulação C1: Comparação dos intervalos de confiança

Queríamos comparar a precisão dos intervalos de confiança para variância que são calculados usando o método AdjDF e método de Bonett. Geramos amostras aleatórias de diferentes tamanhos ($n = 20, 30, 40, 50, 60, 80, 100, 150, 200, 250, 300$) a partir de várias distribuições e calculamos os intervalos de confiança com cada método. A distribuição incluiu:

- Distribuição normal padrão ($N(0, 1)$)
- Distribuições simétricas e com cauda leve, incluindo a distribuição uniforme ($U(0, 1)$) e a distribuição beta, com ambos os parâmetros definidos para 3 ($B(3, 3)$)
- Distribuições simétrica e com caudas pesadas, incluindo distribuições t com 5 e 10 graus de liberdade ($t(5)$, $t(10)$), e a distribuição de Laplace com localização 0 e escala de 1 (LPL)
- Distribuições assimétricas e com caudas pesadas, incluindo a distribuição exponencial com escala de 1 (Exp) e distribuições qui-quadrado com 3, 5 e 10 graus de liberdade (Chi (3), Chi (5), Chi (10))
- Distribuição assimétrica à esquerda e com cauda pesada; especificamente, a distribuição beta, com os parâmetros definidos para 8 e 1, respectivamente ($B(8, 1)$)

Além disso, para avaliar o efeito direto de outliers, geramos amostras de distribuições normais contaminadas definidos como

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

em que p é o parâmetro de mistura e $1 - p$ é a proporção de contaminação (que é igual à proporção de outliers). Selecionamos duas populações contaminadas normais para o estudo: $CN(0,9, 3)$, em que 10% da população são outliers; e $CN(0,8, 3)$, em que 20% da população são outliers. Estas duas distribuições são simétricas e têm caudas longas, devido aos outliers.

Para cada tamanho de amostra, 10.000 replicações de amostra foram extraídas de cada distribuição e os intervalos de confiança bilaterais de 95% foram calculados utilizando-se cada um dos métodos. O gerador de amostras aleatórias recebeu uma semente de modo que ambos os métodos foram aplicados para as mesmas amostras. Com base nesses intervalos de confiança, em seguida, calculamos as probabilidades de cobertura simulada (CovP) e larguras médias de intervalo (AveW) para cada método. Se os intervalos de confiança dos dois métodos têm aproximadamente as mesmas probabilidades de cobertura simuladas, o método que produz intervalos mais curtos (em média) é mais preciso. Como usamos um nível de confiança alvo de 95%, o erro da simulação foi $\sqrt{0,95(0,05)/10.000} = 0,2\%$.

Os resultados da simulação foram registrados nas Tabelas 1 e 2 abaixo.

Tabela 1 Probabilidades de cobertura simulada do intervalo de confiança bilateral de 95% para a variância calculada usando-se os métodos AdjDF e de Bonett. Estas amostras foram gerados a partir de distribuições simétricas com caudas leves, normais, quase normais ou pesadas.

Distribuição		Distribuições simétricas com caudas leves, normais, quase normais				Distribuições simétricas com caudas pesadas			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lpl	CN(0,8, 3)	CN (0,9, 3)	T(5)
Assimetria		0	0	0	0	0	0	0	0
Curtose (γ_e)		-1,20	-0,667	0	1,00	3,00	4,544	5,333	6,00
n = 10									
AdjDF	CovP	0,91	0,909	0,903	0,883	0,853	0,793	0,815	0,858
	AveW	0,154	0,087	3,276	5,16	13,924	21,658	14,913	11,742
Bonett	CovP	0,972	0,967	0,962	0,952	0,919	0,891	0,92	0,935
	AveW	0,242	0,115	3,71	5,134	10,566	15,335	10,367	8,578
n = 20									
AdjDF	CovP	0,937	0,937	0,923	0,909	0,881	0,819	0,817	0,868
	AveW	0,080	0,045	1,572	2,463	5,781	9,265	6,539	5,151
Bonett	CovP	0,953	0,954	0,946	0,934	0,909	0,856	0,864	0,904
	AveW	0,10	0,051	1,683	2,422	4,932	7,282	4,945	4,026
n = 30									
AdjDF	CovP	0,946	0,942	0,933	0,917	0,894	0,851	0,823	0,882
	AveW	0,061	0,034	1,17	1,764	4,117	6,33	4,557	3,667
Bonett	CovP	0,951	0,95	0,947	0,933	0,909	0,869	0,852	0,907
	AveW	0,070	0,037	1,221	1,75	3,654	5,383	3,736	2,997
n = 40									
AdjDF	CovP	0,953	0,947	0,932	0,922	0,904	0,867	0,833	0,89
	AveW	0,051	0,028	0,971	1,489	3,246	5,131	3,654	3,024
Bonett	CovP	0,954	0,951	0,941	0,936	0,914	0,879	0,856	0,907
	AveW	0,057	0,030	1,002	1,469	2,994	4,519	3,128	2,542
n = 50									
AdjDF	CovP	0,951	0,945	0,937	0,925	0,911	0,878	0,838	0,893
	AveW	0,045	0,025	0,849	1,291	2,789	4,357	3,091	2,603

Distribuição		Distribuições simétricas com caudas leves, normais, quase normais				Distribuições simétricas com caudas pesadas			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lpl	CN(0,8, 3)	CN (0,9, 3)	T(5)
Assimetria		0	0	0	0	0	0	0	0
Curtose (γ_e)		-1,20	-0,667	0	1,00	3,00	4,544	5,333	6,00
Bonett	CovP	0,951	0,947	0,944	0,938	0,918	0,888	0,855	0,908
	AveW	0,049	0,026	0,87	1,28	2,613	3,939	2,729	2,24

n = 60

AdjDF	CovP	0,949	0,943	0,938	0,926	0,913	0,89	0,853	0,899
	AveW	0,040	0,022	0,766	1,155	2,49	3,857	2,768	2,283
Bonett	CovP	0,949	0,947	0,943	0,935	0,918	0,896	0,868	0,91
	AveW	0,043	0,023	0,781	1,147	2,354	3,552	2,498	2,023

n = 70

AdjDF	CovP	0,948	0,945	0,94	0,93	0,913	0,89	0,858	0,896
	AveW	0,037	0,020	0,701	1,056	2,283	3,458	2,475	2,049
Bonett	CovP	0,947	0,946	0,944	0,938	0,918	0,894	0,868	0,905
	AveW	0,039	0,021	0,713	1,049	2,174	3,227	2,272	1,828

n = 80

AdjDF	CovP	0,947	0,949	0,938	0,929	0,918	0,905	0,869	0,902
	AveW	0,034	0,019	0,652	0,988	2,089	3,205	2,30	1,906
Bonett	CovP	0,946	0,95	0,942	0,935	0,923	0,907	0,877	0,911
	AveW	0,036	0,019	0,662	0,982	2,005	3,014	2,133	1,716

n = 90

AdjDF	CovP	0,946	0,947	0,948	0,929	0,918	0,908	0,869	0,901
	AveW	0,032	0,018	0,611	0,921	1,951	2,982	2,124	1,874
Bonett	CovP	0,945	0,948	0,952	0,936	0,92	0,91	0,874	0,909
	AveW	0,034	0,018	0,618	0,916	1,882	2,822	1,984	1,646

n = 100

AdjDF	CovP	0,947	0,951	0,945	0,933	0,92	0,91	0,885	0,912
	AveW	0,030	0,017	0,576	0,873	1,83	2,801	2,017	1,658
Bonett	CovP	0,946	0,953	0,948	0,937	0,923	0,912	0,891	0,916
	AveW	0,032	0,017	0,583	0,869	1,772	2,666	1,899	1,522

Distribuição	Distribuições simétricas com caudas leves, normais, quase normais				Distribuições simétricas com caudas pesadas			
	U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lpl	CN(0,8, 3)	CN (0,9, 3)	T(5)
Assimetria	0	0	0	0	0	0	0	0
Curtose (γ_e)	-1,20	-0,667	0	1,00	3,00	4,544	5,333	6,00

n = 150

AdjDF	CovP	0,949	0,951	0,947	0,936	0,932	0,925	0,896	0,912
	AveW	0,024	0,014	0,464	0,70	1,47	2,228	1,602	1,325
Bonett	CovP	0,948	0,952	0,949	0,939	0,933	0,924	0,898	0,915
	AveW	0,025	0,014	0,467	0,698	1,438	2,156	1,539	1,251

n = 200

AdjDF	CovP	0,943	0,949	0,948	0,938	0,927	0,93	0,914	0,918
	AveW	0,021	0,012	0,40	0,605	1,265	1,906	1,373	1,178
Bonett	CovP	0,942	0,951	0,949	0,94	0,928	0,93	0,915	0,92
	AveW	0,021	0,012	0,402	0,603	1,245	1,86	1,333	1,106

n = 250

AdjDF	CovP	0,952	0,952	0,949	0,942	0,938	0,929	0,909	0,915
	AveW	0,019	0,010	0,355	0,538	1,12	1,69	1,219	1,037
Bonett	CovP	0,951	0,952	0,949	0,944	0,941	0,929	0,909	0,916
	AveW	0,019	0,010	0,357	0,537	1,106	1,657	1,19	0,986

n = 300

AdjDF	CovP	0,95	0,948	0,951	0,94	0,938	0,936	0,92	0,914
	AveW	0,017	0,009	0,324	0,49	1,019	1,544	1,115	0,933
Bonett	CovP	0,95	0,947	0,951	0,942	0,937	0,929	0,92	0,916
	AveW	0,017	0,010	0,325	0,489	1,009	1,657	1,093	0,897

Tabela 2 Probabilidades de cobertura simulada do intervalo de confiança bilateral de 95% para a variância calculada usando-se os métodos AdjDF e de Bonett. Estas amostras foram gerados a partir de distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas.

Distribuição		Distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas			Distribuições assimétricas com caudas pesadas	
		Chi(10)	B(8, 1)	Chi(5)	Chi(3)	Exp
Assimetria		0,894	-1,423	1,265	1,633	2
Curtose (γ_e)		1,20	2,284	2,40	4,00	6
n = 10						
AdjDF	CovP	0,869	0,815	0,836	0,797	0,758
	AveW	93,383	0,065	61,994	47,821	10,711
Bonett	CovP	0,95	0,917	0,938	0,911	0,882
	AveW	91,006	0,058	53,83	38,137	7,498
n = 20						
AdjDF	CovP	0,889	0,862	0,862	0,833	0,811
	AveW	41,497	0,026	25,479	20,099	4,293
Bonett	CovP	0,932	0,912	0,913	0,893	0,877
	AveW	41,60	0,026	24,094	17,232	3,37
n = 30						
AdjDF	CovP	0,901	0,881	0,88	0,864	0,838
	AveW	30,021	0,018	18,182	13,63	2,844
Bonett	CovP	0,931	0,92	0,914	0,906	0,885
	AveW	30,462	0,019	17,858	12,634	2,441
n = 40						
AdjDF	CovP	0,909	0,882	0,885	0,867	0,862
	AveW	24,459	0,015	14,577	10,649	2,193
Bonett	CovP	0,93	0,915	0,913	0,904	0,898
	AveW	24,952	0,015	14,504	1,991	1,991
n = 50						
AdjDF	CovP	0,912	0,90	0,892	0,871	0,868
	AveW	21,373	0,013	12,694	9,115	1,861

Distribuição		Distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas			Distribuições assimétricas com caudas pesadas	
		Chi(10)	B(8, 1)	Chi(5)	Chi(3)	Exp
Assimetria		0,894	-1,423	1,265	1,633	2
Curtose (γ_e)		1,20	2,284	2,40	4,00	6
Bonett	CovP	0,93	0,927	0,916	0,903	0,901
	AveW	21,814	0,013	12,741	8,897	1,735

n = 60

AdjDF	CovP	0,915	0,908	0,901	0,89	0,875
	AveW	18,928	0,011	11,338	8,211	1,645
Bonett	CovP	0,93	0,933	0,923	0,917	0,90
	AveW	19,369	0,012	11,456	8,093	1,554

n = 70

AdjDF	CovP	0,915	0,91	0,904	0,898	0,881
	AveW	17,513	0,010	10,307	7,461	1,488
Bonett	CovP	0,932	0,932	0,922	0,919	0,906
	AveW	17,906	0,011	10,464	7,408	1,429

n = 80

AdjDF	CovP	0,92	0,916	0,911	0,904	0,89
	AveW	16,157	0,009	9,604	6,892	1,349
Bonett	CovP	0,935	0,936	0,929	0,924	0,915
	AveW	16,537	0,010	9,765	6,882	1,314

n = 90

AdjDF	CovP	0,924	0,918	0,911	0,897	0,894
	AveW	15,25	0,009	9,007	6,323	1,255
Bonett	CovP	0,938	0,936	0,929	0,918	0,913
	AveW	15,609	0,009	9,175	6,366	1,23

n = 100

AdjDF	CovP	0,926	0,919	0,915	0,908	0,895
	AveW	14,332	0,008	8,451	6,016	1,171
Bonett	CovP	0,935	0,936	0,931	0,924	0,916
	AveW	14,664	0,009	8,625	6,063	1,158

Distribuição	Distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas			Distribuições assimétricas com caudas pesadas	
	Chi(10)	B(8, 1)	Chi(5)	Chi(3)	Exp
Assimetria	0,894	-1,423	1,265	1,633	2
Curtose (γ_e)	1,20	2,284	2,40	4,00	6

n = 150

AdjDF	CovP	0,933	0,925	0,923	0,913	0,911
	AveW	11,606	0,007	6,781	4,792	0,933
Bonett	CovP	0,943	0,941	0,936	0,929	0,928
	AveW	11,846	0,007	6,942	4,875	0,937

n = 200

AdjDF	CovP	0,935	0,934	0,926	0,916	0,915
	AveW	9,973	0,006	5,849	4,127	0,799
Bonett	CovP	0,942	0,948	0,936	0,93	0,931
	AveW	10,185	0,006	5,991	4,212	0,808

n = 250

AdjDF	CovP	0,938	0,939	0,934	0,926	0,922
	AveW	8,899	0,005	5,231	3,652	0,705
Bonett	CovP	0,946	0,951	0,944	0,936	0,931
	AveW	9,078	0,005	5,355	3,735	0,716

n = 300

AdjDF	CovP	0,942	0,938	0,934	0,931	0,922
	AveW	8,156	0,005	4,749	3,344	0,64
Bonett	CovP	0,947	0,948	0,943	0,941	0,933
	AveW	8,314	0,005	4,862	3,419	0,651

Nossos resultados são muito consistentes com os publicados por Bonett (2006). Como mostrado nas Tabelas 1 e 2, os intervalos de confiança calculados utilizando-se o método de Bonett são superiores aos intervalos de confiança calculados usando-se o método AdjDF porque eles produzem probabilidades de cobertura mais próximas do nível pretendido de 0,95 e intervalos de confiança mais estreitos, em média. Se os intervalos de confiança dos dois métodos têm aproximadamente as mesmas probabilidades de cobertura simuladas, o método que produz intervalos mais curtos (em média) é mais preciso. Isto significa que o teste estatístico de variância com base no método de Bonett apresenta melhor desempenho e os resultados nas taxas de erro Tipo I inferior e Tipo II. Quando os tamanhos das amostras

são grandes, os dois métodos de produzem resultados quase idênticos, mas tamanhos de amostra de pequena a moderada, o método de Bonett é superior.

Embora o método de Bonett geralmente apresente um desempenho melhor do que o método AdjDF, ele produz consistentemente probabilidades de cobertura abaixo do objetivo de cobertura de 0,95 para distribuições com caudas pesadas (simétricas ou assimétricas) mesmo para amostras extremamente grandes ($n > 100$). Isto é ilustrado na Figura 1 abaixo, que representa graficamente as probabilidades de cobertura simulados para o método de Bonett contra o excesso de curtose verdadeiro da população para os tamanhos de amostras pequenos, moderados e grandes.

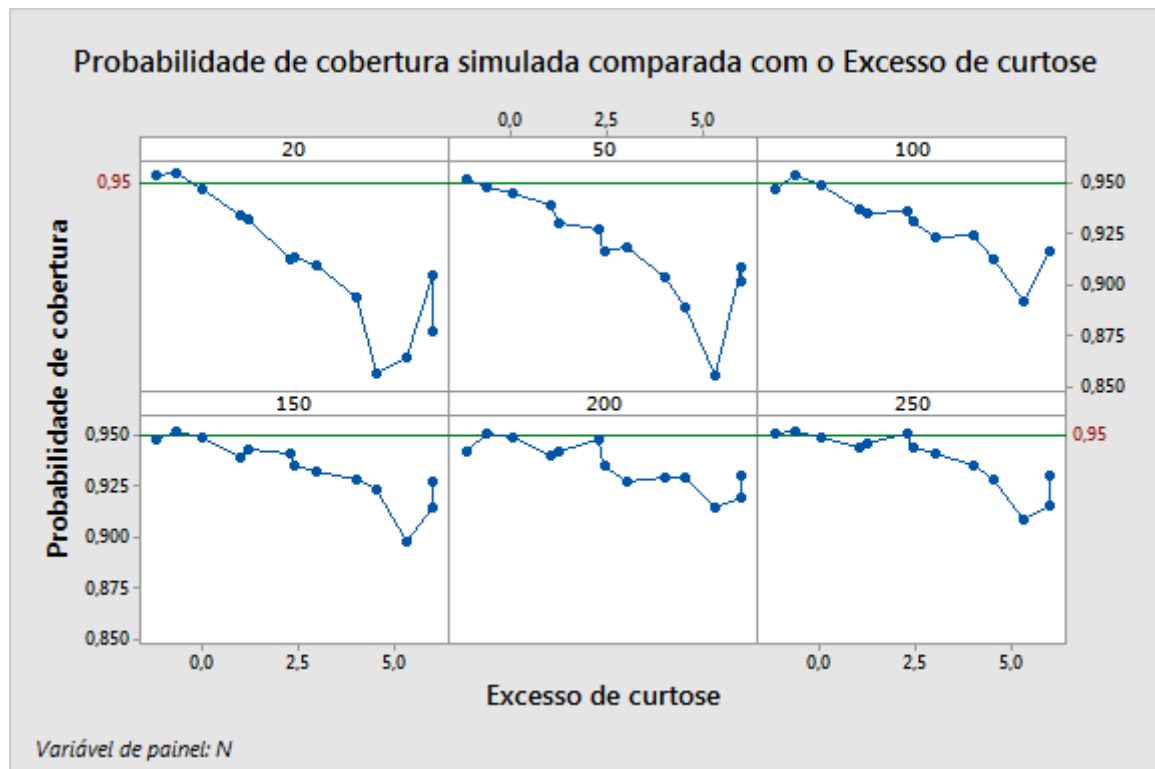


Figura 1 Probabilidades de cobertura simulada para intervalos de confiança de 95% Bonett representado graficamente em função do excesso de curtose de cada distribuição em vários tamanhos de amostra.

Como mostrado na Figura 1, quanto maior a curtose, maior é o tamanho da amostra necessário para fazer com que as probabilidades de cobertura simuladas se aproximem do nível alvo. Como observado anteriormente, as probabilidades de cobertura simuladas para o método de Bonett são baixas para distribuições com caudas pesadas. No entanto, para distribuições de cauda leves, como as distribuições Beta(3,3), as probabilidades de cobertura simuladas de distribuição são estáveis e no alvo para tamanhos de amostra pequenos como 20. Portanto, fundamentamos nossos critérios para determinar a validade do método de Bonett no tamanho da amostra e no peso das caudas da distribuição a partir das quais a amostra é retirada.

Como um primeiro passo para o desenvolvimento destes critérios, classificamos as distribuições em três categorias de acordo com o peso de suas caudas:

- **Distribuições de cauda leve e normal (tipo L):** estas são as distribuições para as quais os intervalos de confiança de Bonett produzem probabilidades de cobertura estáveis próximas do nível de cobertura alvo. Para essas distribuições, tamanhos de amostra pequenos, na faixa de 20, produzem resultados precisos. Os exemplos incluem a distribuição uniforme, a distribuição Beta(3, 3), a distribuição normal, a distribuição t com 10 graus de liberdade e a distribuição do qui-quadrado com 10 graus de liberdade.
- **Distribuições com cauda moderadamente pesadas (tipo M):** para essas distribuições, o método de Bonett requer uma amostra mínima de 80 para que as probabilidades de cobertura simuladas fiquem próximas à cobertura alvo. Os exemplos incluem a distribuição de qui-quadrado com 5 graus de liberdade, e a distribuição Beta(8, 1).
- **Distribuições com caudas pesadas (Tipo H):** estas são as distribuições para as quais os intervalos de confiança de Bonett produzem probabilidades de cobertura que estão muito abaixo da cobertura alvo, a menos que os tamanhos de amostra sejam extremamente grandes ($n \geq 200$). Os exemplos incluem a distribuição t com 5 graus de liberdade, a distribuição de Laplace, a distribuição do qui-quadrado com 3 graus de liberdade, a distribuição exponencial e as duas distribuições normais contaminadas, CN(0,9, 3) e CN(0,8, 3) .

Assim, uma regra geral para avaliar a validade do método de Bonett exige que nós desenvolvamos um procedimento para identificar de qual dos 3 tipos de distribuição dos dados da amostra são provenientes. Desenvolvemos este procedimento como parte da verificação da Validade dos dados de teste. Para obter mais detalhes, consulte o Anexo E.

Anexo D: Poder teórico

Extraímos a função de poder teórico do teste associado ao método de Bonett e simulações realizadas para comparar o poder teórico e simulado do teste. Se as curvas de poder teórico e simulado estiverem próximas umas das outras, a análise do poder e do tamanho de amostra baseada na função de poder teórico deve produzir resultados exatos.

Fórmula D1: Função de poder teórico para o método de Bonett

Como descrito anteriormente, o método de Bonett se baseia na abordagem clássica bem conhecido, na qual o teorema do limite central e o método Cramer δ são usados para encontrar uma distribuição assintótica da variância da amostra transformada por log. Mais especificamente, estabeleceu que, em amostras grandes, $\frac{\ln S^2 - \ln \sigma^2}{se}$ é aproximadamente distribuída como a distribuição normal padrão. O denominador, se , é o erro padrão da amostra grande da variância da amostra transformada por log e é dado como

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

em que γ é a curtose da população hierarquicamente superior desconhecida.

Ele preconiza que uma função de poder aproximado com um nível alfa aproximado para o teste bilateral usando-se o método de Bonett pode ser dado como uma função do tamanho da amostra, o $\rho = \sigma/\sigma_0$ da proporção e a curtose da população hierarquicamente superior γ como

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

em σ_0 é o valor hipotético do desvio padrão desconhecido, Φ é a CDF de uma distribuição normal padrão, e z_{α} é o ponto superior α do percentil da distribuição normal padrão. As funções de poder unilaterais também podem ser obtidas a partir destes cálculos.

Observe que, ao planejar o tamanho da amostra para um estudo, uma estimativa da curtose pode ser usada no lugar da verdadeiro curtose. Esta estimativa baseia-se normalmente nas opiniões dos especialistas ou nos resultados das experiências anteriores. Se essa informação não estiver disponível, muitas vezes é uma boa prática realizar um pequeno estudo piloto para desenvolver os planos para o estudo principal. Usando-se uma amostra a partir do estudo piloto, a curtose pode ser estimada como

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

em que m é uma média aparada com proporção de ajuste igual a $1/2\sqrt{n-4}$.

Simulação D1: Comparação do poder real com o poder teórico

Projetamos uma simulação para comparar níveis de poder reais estimados (citados como níveis de poder simulados) para os níveis de poder teórico (citados como níveis de poder aproximados) quando se utiliza o método de Bonett para testar a variância.

Em cada experimento, foram geradas 10.000 réplicas de amostra, cada um do tamanho n , em que $n = 20, 30, 40, 50, \dots, 120$, a partir de cada uma das distribuições descritas em Simulação C1 (consulte o Anexo C). Para cada distribuição e tamanho de amostra, foi calculado o nível de poder simulado como a fração de 10.000 réplicas de amostra aleatória para as quais o teste bilateral com nível alfa $\alpha = 0,05$ foi significativo. Ao calcular o poder simulado, foi utilizado $\rho = \sigma/\sigma_0 = 1,25$ para obter os níveis de poder relativamente pequenos. Para efeito de comparação, calculamos os níveis de poder correspondente quando usada a função de poder teórico.

Os resultados são mostrados nas Tabelas 3 e 4 e representados graficamente na Figura 2, abaixo.

Tabela 3 Níveis de poder simulado (avaliados em $\rho = \sigma/\sigma_0 = 1,25$) de um teste bilateral para a variância com base no método de Bonett em comparação com níveis de poder teórico (aproximação normal). As amostras foram gerados a partir de distribuições simétricas com caudas leves, normais, quase normais ou pesadas.

n	Poder	Distribuições simétricas com caudas leves, normais, quase normais				Distribuições simétricas com caudas pesadas			
		U(0,1)	B(3, 3)	N(0, 1)	t(10)	Lpl	CN(0,8, 3)	CN(0,9, 3)	t(5)
20	Simul.	0,521	0,39	0,31	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	0,49	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	0,34
	Aprox.	0,921	0,754	0,586	0,431	0,284	0,229	0,21	0,196
60	Simul.	0,942	0,822	0,701	0,57	0,38	0,285	0,274	0,384
	Aprox.	0,961	0,83	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	0,34	0,309	0,288

n	Poder	Distribuições simétricas com caudas leves, normais, quase normais				Distribuições simétricas com caudas pesadas			
		U(0,1)	B(3, 3)	N(0, 1)	t(10)	Lpl	CN(0,8, 3)	CN(0,9, 3)	t(5)
90	Simul.	0,988	0,937	0,851	0,724	0,514	0,40	0,377	0,523
	Aprox.	0,996	0,95	0,84	0,676	0,467	0,375	0,342	0,318
100	Simul.	0,994	0,961	0,88	0,779	0,558	0,43	0,411	0,566
	Aprox.	0,998	0,967	0,876	0,722	0,508	0,41	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	0,63
	Aprox.	1,00	0,987	0,928	0,799	0,584	0,476	0,434	0,405

Tabela 4 Níveis de poder simulado (avaliados em $\rho = \sigma/\sigma_0 = 1,25$) de um teste bilateral para a variância com base no método de Bonett em comparação com níveis de poder teórico (aproximação normal). As amostras foram gerados a partir de distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas.

n	Poder	Distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas			Distribuições assimétricas com caudas 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	0,19	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
50	Simul.	0,455	0,324	0,349	0,263	0,208
	Aprox.	0,409	0,323	0,316	0,245	0,196
60	Simul.	0,521	0,376	0,399	0,302	0,239
	Aprox.	0,477	0,377	0,369	0,286	0,227
70	Simul.	0,583	0,419	0,463	0,361	0,269
	Aprox.	0,539	0,43	0,42	0,325	0,257
80	Simul.	0,646	0,473	0,499	0,394	0,299
	Aprox.	0,597	0,479	0,469	0,365	0,288

n	Poder	Distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas			Distribuições assimétricas com caudas pesadas	
		Chi(10)	B(8, 1)	Chi(5)	Chi(3)	Exp
90	Simul.	0,688	0,517	0,561	0,428	0,327
	Aprox.	0,649	0,526	0,516	0,403	0,318
100	Simul.	0,738	0,561	0,591	0,469	0,368
	Aprox.	0,695	0,571	0,56	0,44	0,347
110	Simul.	0,779	0,608	0,637	0,495	0,394
	Aprox.	0,737	0,611	0,60	0,475	0,376
120	Simul.	0,81	0,635	0,679	0,538	0,416
	Aprox.	0,774	0,65	0,638	0,509	0,405

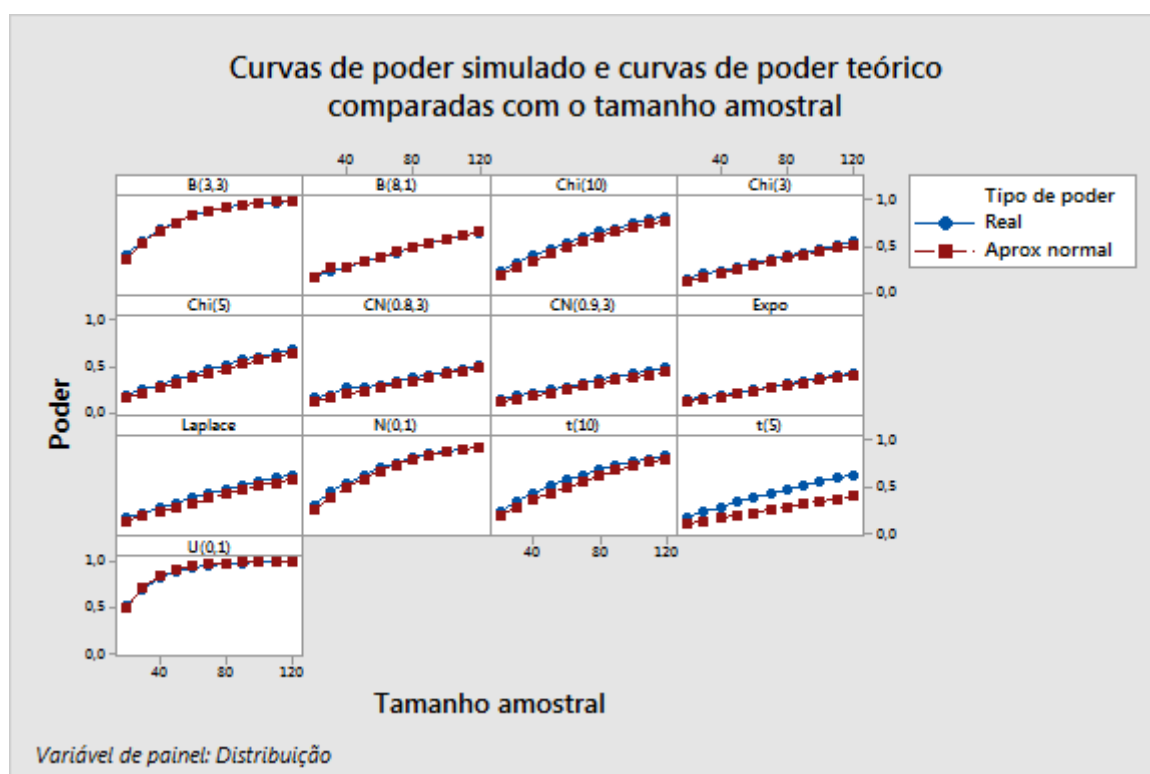


Figura 2 As curvas de poder simulados comparado com as curvas de poder teórico para várias distribuições

Os resultados nas Tabelas 3 e 4 e na Figura 2 mostram que, quando as amostras são geradas a partir de distribuições com caudas leves (distribuições do tipo L, como definido no Anexo C), como a distribuição uniforme, a distribuição Beta (3, 3), a distribuição normal, a distribuição t com 10 graus de liberdade e da distribuição qui-quadrado com 10 graus de liberdade, os valores de poder teórico e os níveis de poder simulado são praticamente indistinguíveis.

No entanto, para distribuições com caudas pesadas (distribuições do tipo H), as curvas de poder simulado são marcadamente acima das curvas de poder teórico quando as amostras são pequenas. As distribuições com cauda pesada incluem a distribuição t com 5 graus de liberdade, a distribuição de Laplace, a distribuição do qui-quadrado com 3 graus de liberdade, a distribuição exponencial e as duas distribuições normais contaminadas, $CN(0,9, 3)$ e $CN(0,8, 3)$. Portanto, quando planejar o tamanho da amostra para um estudo e a amostra for proveniente de uma distribuição com caudas pesadas, o tamanho da amostra estimada pela função de poder teórico pode ser maior do que o tamanho da amostra real necessário para alcançar um determinado poder alvo.

Anexo E: O teste SJ para caudas normais contra caudas pesadas

Os resultados do estudo de simulação no Anexo C mostraram que, quando as caudas da distribuição são mais pesadas, são necessárias amostras maiores para que a probabilidade de cobertura simulada dos intervalos de confiança de Bonett possa se aproximar do alvo. A assimetria, no entanto, não parece ter um efeito significativo sobre as probabilidades de cobertura simulada.

Portanto, foi necessário desenvolver um critério para avaliar a validade do método de Bonett com base no tamanho da amostra e no peso das caudas da distribuição a partir da qual a amostra é retirada. Felizmente, Gel et al. (2007) fornecem um teste razoavelmente potente para testar diretamente a hipótese nula de que a distribuição tem caudas normais contra a hipótese alternativa de que a distribuição tem caudas pesadas. O teste, ao que chamamos de teste de SJ, baseia-se na seguinte estatística:

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

em que S é o desvio padrão da amostra, \hat{j} é a estimativa do desvio absoluto da média da amostra da mediana, m , e é dada como

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

Um teste de α de tamanho aproximado contra a hipótese alternativa de caudas pesadas rejeita a hipótese nula de caudas normais se

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

em que z_α é o percentil de α superior de uma distribuição normal padrão e $\sigma_R = (\pi - 3)/2$.

Gel et al. (2007) mostrou que a substituição da parte superior do percentil de α da distribuição normal padrão com o da distribuição t com $(\sqrt{n} + 3)/2$ graus de liberdade fornece melhores aproximações para tamanhos de amostra moderados. Portanto, ao aplicar o teste de SJ para a validade da verificação de dados de teste, substituímos z_α por $t_{d,\alpha}$ o percentil de α superior da distribuição t com $d = (\sqrt{n} + 3)/2$ graus de liberdade.

Anexo F: Validade do teste

Simulação F1: Usando o poder simulado o teste de SJ para determinar as classificações de distribuição

Foram realizadas simulações para investigar o poder do teste de SJ. Geramos amostras de vários tamanhos ($n = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200$) a partir de várias distribuições. As distribuições tinham caudas normais, leves, moderadas ou pesadas, e são iguais às descritas na simulação C1 (consulte Anexo C). Para cada tamanho de determinada amostra, 10.000 réplicas de amostra foram retiradas de cada distribuição. Foi calculado o poder simulado do teste de SJ como a proporção de casos em que a hipótese nula (de que a distribuição dos níveis hierarquicamente superiores têm caudas normais) tenha sido rejeitada. Além disso, foram calculados os valores de R médios (AveR) e os valores de p médios (AvePV).

Os resultados da simulação são mostrados nas Tabelas 5 e 6 abaixo.

Tabela 5 Níveis de poder simulado do teste de SJ. As amostras foram gerados a partir de distribuições simétricas com caudas leves, normais, quase normais ou pesadas.

Distribuição		Distribuições simétricas com caudas leves, normais, quase normais				Distribuições simétricas com caudas pesadas			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lpl	CN(0,8, 3)	CN(0,9, 3)	t(5)
n	TrueR	0,921	0,965	1,0	1,032	1,128	1,152	1,118	1,085
10	Poder	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	Poder	0,009	0,027	0,071	0,121	0,35	0,389	0,283	0,215
	AveR	0,986	1,018	1,043	1,063	1,13	1,14	1,11	1,093
	AvePV	0,572	0,44	0,357	0,302	0,171	0,181	0,24	0,247
20	Poder	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
25	Poder	0,002	0,011	0,065	0,153	0,50	0,55	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	0,12	0,128	0,208	0,223
30	Poder	0,001	0,010	0,060	0,17	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	0,57	0,409	0,304	0,103	0,112	0,197	0,209

Distribuição		Distribuições simétricas com caudas leves, normais, quase normais				Distribuições simétricas com caudas pesadas			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lpl	CN(0,8, 3)	CN(0,9, 3)	t(5)
n	TrueR	0,921	0,965	1,0	1,032	1,128	1,152	1,118	1,085
40	Poder	0,00	0,006	0,058	0,19	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	0,84	0,616	0,42	0,287	0,073	0,076	0,162	0,179
50	Poder	0,00	0,004	0,058	0,208	0,746	0,785	0,59	0,462
	AveR	0,939	0,98	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	Poder	0,00	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	0,43	0,267	0,039	0,039	0,109	0,134
70	Poder	0,00	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	Poder	0,00	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	0,95	0,74	0,44	0,241	0,021	0,021	0,079	0,105
90	Poder	0,00	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
100	Poder	0,00	0,001	0,057	0,301	0,947	0,954	0,805	0,661
	AveR	0,93	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	Poder	0,00	0,00	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	Poder	0,00	0,00	0,052	0,336	0,986	0,988	0,894	0,785
	AveR	0,928	0,971	1,004	1,034	1,127	1,15	1,116	1,084
	AvePV	0,989	0,834	0,454	0,192	0,004	0,003	0,027	0,048
160	Poder	0,00	0,00	0,054	0,402	0,993	0,992	0,916	0,819
	AveR	0,927	0,97	1,004	1,034	1,128	1,15	1,114	1,084
	AvePV	0,993	0,858	0,457	0,177	0,002	0,002	0,021	0,040

Distribuição		Distribuições simétricas com caudas leves, normais, quase normais				Distribuições simétricas com caudas pesadas			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lpl	CN(0,8, 3)	CN(0,9, 3)	t(5)
n	TrueR	0,921	0,965	1,0	1,032	1,128	1,152	1,118	1,085
180	Poder	0,00	0,00	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	Poder	0,00	0,00	0,053	0,448	0,998	0,998	0,954	0,884
	AveR	0,926	0,969	1,003	1,034	1,127	1,15	1,116	1,083
	AvePV	0,997	0,89	0,461	0,153	0,001	0,001	0,011	0,025

Tabela 6 Níveis de poder simulado do teste de SJ. As amostras foram gerados a partir de distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas.

Distribuições		Distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas			Distribuições assimétricas com caudas pesadas	
		Chi(10)	B(8, 1)	Chi(5)	Chi(3)	Exp
n	TrueR	1,028	1,075	1,059	1,098	1,151
10	Poder	0,12	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	Poder	0,139	0,27	0,205	0,292	0,377
	AveR	1,062	1,105	1,082	1,11	1,141
	AvePV	0,32	0,261	0,286	0,245	0,209
20	Poder	0,152	0,295	0,223	0,328	0,449
	AveR	1,051	1,089	1,070	1,101	1,142
	AvePV	0,335	0,26	0,296	0,242	0,186
25	Poder	0,16	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	Poder	0,171	0,37	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	Poder	0,193	0,44	0,331	0,49	0,651
	AveR	1,039	1,085	1,064	1,098	1,143
	AvePV	0,321	0,188	0,246	0,171	0,106

Distribuições		Distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas			Distribuições assimétricas com caudas pesadas	
		Chi(10)	B(8, 1)	Chi(5)	Chi(3)	Exp
<i>n</i>	TrueR	1,028	1,075	1,059	1,098	1,151
50	Poder	0,215	0,484	0,37	0,556	0,72
	AveR	1,037	1,081	1,064	1,10	1,143
	AvePV	0,314	0,173	0,22	0,14	0,080
60	Poder	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	Poder	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	Poder	0,259	0,612	0,474	0,689	0,855
	AveR	1,034	1,079	1,062	1,098	1,148
	AvePV	0,28	0,115	0,17	0,089	0,036
90	Poder	0,284	0,643	0,501	0,733	0,89
	AveR	1,034	1,079	1,060	1,099	1,148
	AvePV	0,27	0,104	0,163	0,075	0,028
100	Poder	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	Poder	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	Poder	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	Poder	0,363	0,815	0,666	0,887	0,978
	AveR	1,031	1,077	1,060	1,098	1,15
	AvePV	0,217	0,047	0,093	0,027	0,005
180	Poder	0,385	0,843	0,692	0,91	0,986
	AveR	1,031	1,077	1,059	1,099	1,148
	AvePV	0,209	0,039	0,083	0,021	0,004

Distribuições		Distribuições assimétricas com caudas quase normais, moderadamente pesadas ou pesadas			Distribuições assimétricas com caudas pesadas	
		Chi(10)	B(8, 1)	Chi(5)	Chi(3)	Exp
<i>n</i>	TrueR	1,028	1,075	1,059	1,098	1,151
200	Poder	0,41	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

Os resultados da simulação na Tabelas 5 e 6 são consistentes com os publicados em Gel et al. (2007). Quando as amostras são das populações normais, os níveis de poder simulado (que, neste caso, representa o nível de significância real do teste), não estão longe do nível alvo, mesmo para tamanhos de amostra baixos como 25. Quando as amostras são de distribuições com caudas pesadas, o poder do teste é baixo para amostras de tamanhos pequenos, mas aumenta para pelo menos 40% quando o tamanho da amostra chega a 40. Especificamente, o poder no tamanho da amostra de 40 é de aproximadamente 40,1% para a distribuição t com 5 graus de liberdade, de 66,5% para a distribuição de Laplace, e de 65,1% para a distribuição exponencial.

Para distribuições de cauda leve (a versão Beta (3,3) e as distribuições uniformes), o poder do teste está próximo de 0 para amostras pequenas e diminui ainda mais quando o tamanho da amostra aumenta. Isto não é surpreendente, porque a evidência para estas distribuições efetivamente sustenta a hipótese alternativa de uma distribuição com cauda mais leve, e não a hipótese alternativa de uma distribuição de cauda mais pesada.

Quando as amostras são de distribuições com caudas um pouco mais pesadas, como a distribuição t com 10 graus de liberdade ou a distribuição qui-quadrado com 10 graus de liberdade, os níveis de poder são baixos para tamanhos de amostra de moderados a grandes. Para os nossos propósitos, este é realmente um bom resultado, pois o teste para uma variância (desvio padrão) funciona bem para essas distribuições e não queremos essas distribuições sejam marcadas como cauda pesada. No entanto, conforme o tamanho da amostra aumenta, o poder do teste aumenta, de modo que essas distribuições com caudas um pouco pesadas são detectadas como distribuições com caudas pesadas.

Portanto, as regras para a avaliação do peso da cauda da distribuição para esse teste também deve levar em consideração o tamanho da amostra. Uma abordagem para fazer isso é calcular um intervalo de confiança para a medida do peso de cauda; entretanto, a distribuição da estatística de SJ é extremamente sensível à distribuição hierarquicamente superior da amostra. Uma abordagem alternativa é a de avaliar o peso das caudas de distribuição com base na força da rejeição da hipótese nula do teste de SJ e no tamanho da amostra. Mais especificamente, os valores de p menores indicam caudas mais pesadas e valores de p maiores indicam caudas mais leves. No entanto, amostras maiores tendem a ter valores de p menores do que amostras menores. Portanto, com base nos níveis de poder simulado, tamanhos de amostra e valores de p médios na Tabela 3, elaboramos um conjunto geral de normas para avaliar as caudas de uma distribuição para cada amostra utilizando o teste de SJ.

Para tamanhos de amostra moderados a grandes ($40 \leq n \leq 100$), se o valor de p estiver entre 0,01 e 0,05, consideramos que não há evidências brandas contra a hipótese nula. Isto é, a distribuição da amostra é classificada como uma distribuição de cauda moderadamente pesada (tipo M). Por outro lado, se o valor de p for inferior a 0,01, então existe uma forte evidência contra a hipótese nula, e a distribuição do nível hierarquicamente superior da amostra é classificada como uma distribuição com caudas pesadas (H-tipo).

Para amostras grandes ($n > 100$), categorizamos a distribuição do nível hierarquicamente superior como uma distribuição do tipo M, se o valor de p ficar entre 0,005 e 0,01, e como uma distribuição do tipo H se o valor de p for extremamente pequeno (abaixo de 0,005). Observe que, quando o tamanho da amostra está abaixo de 40, o poder do teste de SJ é geralmente baixo demais para que a distribuição da amostra possa ser determinada de forma eficaz.

As regras gerais de classificação para a validade do teste de 1-variância pelo método de Bonett estão resumidas na Tabela 7 abaixo.

Tabela 7 Regras de classificação para identificar a distribuição hierarquicamente superior de cada amostra (é o valor de p do teste SJ)

Condição	Tipo de distribuição
$n < 40$	Nenhum é determinado
$100 \geq n \geq 40$ e $p > 0,05$	Distribuição tipo L
$n > 100$ e $p > 0,01$	Distribuição tipo L
$40 \leq n \leq 100$ e $0,01 < p \leq 0,05$	Distribuição tipo M
$n > 100$ e $0,005 < p \leq 0,01$	Distribuição tipo M
$40 \leq n \leq 100$ e $p \leq 0,01$	Distribuição tipo H
$n > 100$ e $p \leq 0,005$	Distribuição tipo H

Como indicado anteriormente, com base nos resultados das Tabelas 1 e 2 na simulação C1, os tamanhos de amostra mínimos aproximados necessários para atingir no mínimo 0,93 de probabilidade de cobertura quando as amostras são geradas a partir de uma distribuição tipo L, tipo M e tipo H são de 20, 80 e 200, respectivamente. No entanto, como o poder do teste de SJ é acessível para pequenas amostras, a exigência de tamanho mínimo de amostra para distribuição do tipo L foi fixada em 40.

Simulação F2: Verificação das regras para distribuições de classificação

Geramos amostras de algumas das distribuições descritas na simulação C1 e utilizamos o teste de SJ para determinar as proporções das amostras que foram classificadas em um dos três grupos de distribuição: tipo L, tipo M e tipo H. Os resultados de simulação são mostrados na Tabela 8.

Tabela 8 Fração de 10.000 amostras de tamanhos diferentes das várias distribuições que são identificadas como tipo L, tipo M e tipo H

n	Distribuições	Tipo L				Tipo M	Tipo H	
		B(3, 3)	N(0, 1)	t(10)	Chi(10)	Chi(5)	Lpl	Exp
40	% do tipo L	99,6	94,0	81,5	80,3	66,6	33,0	34,4
	% do tipo M	0,4	5,5	14,0	14,0	20,0	31,9	22,9
	% do tipo H	0,0	0,5	4,5	5,7	13,4	35,1	42,8
50	% do tipo L	99,7	94,4	78,7	79,1	64,0	25,1	28,0
	% do tipo M	0,3	5,1	15,6	14,2	20,0	29,9	20,7
	% do tipo H	0,0	0,5	5,7	6,7	16,0	45,0	51,3
60	% do tipo L	99,7	94,5	77,3	77,3	59,1	18,5	22,6
	% do tipo M	0,3	5,1	16,4	15,0	22,0	27,4	19,2
	% do tipo H	0,0	0,5	6,3	7,7	18,9	54,1	58,2
70	% do tipo L	99,8	94,4	74,5	75,2	55,9	14,0	18,1
	% do tipo M	0,2	5,0	18,1	16,0	22,2	24,0	17,5
	% do tipo H	0,0	0,6	7,4	8,8	21,9	62,0	64,4
80	% do tipo L	99,9	94,3	74,1	74,4	53,0	10,0	13,9
	% do tipo M	0,1	5,1	17,8	16,7	22,8	21,0	15,5
	% do tipo H	0,0	0,6	8,2	8,9	24,2	69,0	70,6
90	% do tipo L	99,9	94,4	71,2	72,1	49,5	7,5	11,1
	% do tipo M	0,1	5,0	19,1	17,2	22,6	16,5	13,7
	% do tipo H	0,0	0,6	9,7	10,7	27,9	76,0	75,3
100	% do tipo L	99,9	94,5	70,8	70,3	47,3	4,8	8,9
	% do tipo M	0,1	4,9	19,5	17,9	22,7	14,3	11,8
	% do tipo H	0,0	0,6	9,7	11,8	30,0	80,9	79,4
120	% do tipo L	100,0	99,4	87,4	87,2	64,8	12,0	14,4
	% do tipo M	0,0	0,4	5,0	4,5	7,9	7,8	5,6
	% do tipo H	0,0	0,2	7,6	8,4	27,4	80,4	80,0
140	% do tipo L	100,0	99,3	86,0	85,1	60,5	7,0	9,9
	% do tipo M	0,0	0,5	5,2	5,0	8,6	5,6	4,1
	% do tipo H	0,0	0,2	8,8	9,9	30,9	87,4	86,0
160	% do tipo L	100,0	99,4	83,4	83,0	55,6	4,0	6,9
	% do tipo M	0,0	0,5	6,3	5,8	9,5	3,5	3,0
	% do tipo H	0,0	0,1	10,4	11,2	34,9	92,5	90,1

n	Distribuições	Tipo L				Tipo M	Tipo H	
		B(3, 3)	N(0, 1)	t(10)	Chi(10)	Chi(5)	Lpl	Exp
180	% do tipo L	100,0	99,3	81,1	81,7	51,0	2,5	4,6
	% do tipo M	0,0	0,5	6,8	5,9	9,4	1,9	2,2
	% do tipo H	0,0	0,2	12,1	12,4	39,6	95,6	93,2
200	% do tipo L	100,0	99,5	79,0	80,5	47,2	1,3	3,0
	% do tipo M	0,0	0,4	7,6	6,1	9,4	1,6	1,7
	% do tipo H	0,0	0,1	13,4	13,4	43,4	97,1	95,3

Os resultados na Tabela 8 mostram que, quando as amostras são de distribuições de cauda leve (tipo L) e de cauda pesada (tipo H), uma proporção maior das amostras é classificada corretamente. Por exemplo, quando as amostras de tamanho 40 foram geradas a partir de distribuição Beta(3, 3), 99,6% das amostras foram classificadas corretamente como tendo caudas leves; quando amostras de tamanho 90 foram geradas a partir da distribuição de Laplace, 76,0% foram classificadas corretamente como tendo caudas pesadas. Como resultado, as mensagens de aviso no Cartão de relatório sobre a validade do teste não são erradamente emitidas quando as amostras são realmente de distribuições de caudas leves, e são corretamente emitidas quando a amostra vem de uma distribuição com caudas pesadas e a exigência de tamanho mínimo da amostra não é atendida. Além disso, para as amostras a partir de distribuições com caudas moderadamente pesadas (tipo M), como a distribuição do Qui-quadrado (5), uma proporção mais elevada de amostras é classificada erroneamente como sendo de cauda leve (do tipo L) quando as amostras são pequenas (para um tamanho de amostra de 40, 66% das amostras são classificadas erroneamente como distribuição de cauda leve). Consequentemente, para esses casos, as mensagens de advertência no Cartão de Relatório pode não ser emitida mesmo que as distribuições hierarquicamente superiores tenham caudas moderadamente pesadas. No entanto, quando o tamanho da amostra é superior a 80, a classificação errônea como uma distribuição do tipo L não tem efeito porque o requisito de tamanho mínimo da amostra já foi atendido.

© 2015, 2017 Minitab Inc. All rights reserved.

Minitab®, Quality. Analysis. Results.® and the Minitab® logo are all registered trademarks of Minitab, Inc., in the United States and other countries. See minitab.com/legal/trademarks for more information.