

1-Sample Standard Deviation Test

Overview

The 1-Sample Standard Deviation test is used to estimate the variability of your process and to compare the variability with a target value. Typically, variability is measured using the variance or, equivalently, the standard deviation.

Many statistical methods have been developed to evaluate the variance of a population, each with its own strengths and limitations. The classical chi-square method used to test the variance is likely the most commonly used, but it is extremely sensitive to the assumption of normality and can produce extremely inaccurate results when the data are skewed or heavy-tailed. Other methods have also been developed, but these too have drawbacks. For example, some methods are valid only for large samples or for data from a symmetric distribution (see Appendix A).

In Minitab 15, we use an alternative large-sample method that we derived from a scaled chi-squared approximation to the distribution of the sample variance by Box (1953). This method, referred to as the adjusted degrees of freedom (AdjDF) method, is less sensitive to the normal assumption for sufficiently large samples and has been shown to produce more accurate confidence intervals than other methods (Hummel, Banga, & Hettmansperger, 2005). More recently, however, a revised statistical method by Bonett (2006) has been developed that appears to provide better approximate confidence intervals.

In this paper, we evaluate the performance of Bonett's method. In addition, for sample size planning, we investigate the power function for the equivalent testing procedure associated with Bonett's confidence intervals. Based on our results, we use the Bonett method for the 1-Sample Standard Deviation test in the Assistant. We also examine the following data checks that are automatically performed and displayed in the Assistant Report Card and explain how they affect the results:

- Unusual data

- Validity of test
- Sample size

1-sample standard deviation methods

Bonett's method versus AdjDF method

Before the publication of Bonett's method (2006), the most robust procedure for making inferences on the variance of a population was most likely the AdjDF method. The published results by Bonett, however, show that Bonett's method provides stable confidence levels that are near the target level when samples of moderate size are drawn from nonnormal populations. Therefore, Bonett's method may be preferable for making inferences on the standard deviation or variance of a population.

Objective

We wanted to compare the performance of Bonett's method with the AdjDF method when making inferences on the variance of a single population. Specifically, we wanted to determine which method produces more accurate confidence intervals for the variance (or the standard deviation) when samples of various sizes are generated from nonnormal populations.

We compare confidence intervals because Bonett's method directly applies to confidence intervals. The equivalent hypothesis testing procedure associated with Bonett's confidence intervals can be derived. However, to directly compare our results to those published in Bonett (2006), we examined the confidence intervals rather than the hypothesis tests.

Method

The AdjDF method and Bonett's method are both formally defined in Appendix B. To compare the accuracy of the confidence intervals for each method, we performed the following simulations. First, we generated random samples of various sizes from distributions with different properties, such as skewed and heavy-tailed, symmetric and heavy-tailed, and symmetric and light-tailed distributions. For each sample size, 10,000 sample replicates were drawn from each distribution, and two-sided 95% confidence intervals for the true variance of the distribution were calculated using each method. Then we calculated the proportion of the 10,000 intervals that contained the true variance, referred to as the simulated coverage probability. If the confidence intervals are accurate, the simulated coverage probability should be close to the target coverage probability of 0.95. In addition, we calculated the average widths associated with the confidence intervals for each method. If the confidence intervals of the two methods have about the same simulated coverage probabilities, then the method that produces shorter intervals (on average) is more precise. For more details, see Appendix C.

Results

Bonett's method generally yields better coverage probabilities and more precise confidence intervals than the AdjDF method. As a result, the statistical tests for the variance based on

Bonett's method generate lower Type I and Type II error rates. For this reason, the 1-Sample Standard Deviation test in the Assistant is based on Bonett's method.

In addition, our results show that if the distribution has moderate to heavy tails, Bonett's method requires larger sample sizes to achieve the target level of accuracy:

- For distributions with normal or light tails, a sample size of 20 is sufficient.
- For distributions with moderately heavy tails, the sample size should be at least 80.
- For distributions with heavy tails, the sample size should be at least 200.

Therefore, to ensure that the 1-Sample Standard Deviation test or confidence intervals results are valid for your data, the Assistant includes a data check to simultaneously evaluate both the sample size and the tails of the data distribution (see the Validity of test data check below).

Performance of theoretical power

Bonett's method directly applies to confidence intervals for the variance (or standard deviation). However, using the statistical relationship between hypothesis tests and confidence intervals, we can derive the equivalent test that is associated with Bonett's approximate confidence intervals. Because an exact power function for this test is unavailable, we needed to derive it. In addition, we wanted to evaluate the sensitivity of the theoretical power function to the assumption of normality.

Objective

We wanted to determine whether we could use the theoretical power function of the test associated with Bonett's confidence intervals to evaluate the power and sample size requirements for the 1-Sample Standard Deviation test in the Assistant. To do this, we needed to evaluate whether this theoretical power function accurately reflects the actual power of the test when normal and nonnormal data are analyzed.

Method

The theoretical power function of the test using Bonett's method is derived in Appendix C. We performed simulations to estimate the actual power levels (which we refer to as simulated power levels) using Bonett's method. First, we generated random samples of various sizes from the distributions described in the previous study: skewed and heavy-tailed, symmetric and heavy-tailed, and symmetric and light-tailed distributions. For each distribution, we performed the test on each of 10,000 sample replicates. For each sample size, we calculated the simulated power of the test to detect a given difference as the fraction of the 10,000 samples for which the test is significant. For comparison, we also calculated the corresponding power level using the theoretical power function of the test. If the theoretical power function is not too sensitive to normality, the theoretical and simulated power levels should be close for normal and nonnormal data. For more details, see Appendix D.

Results

Our simulations showed that, when the sample comes from a distribution with normal or light tails, the theoretical and simulated power of the test using Bonett's method are nearly equal. When the sample comes from a distribution with heavy tails, however, the theoretical power function may be conservative and overestimate the sample size required to achieve a given power. Therefore, the theoretical power function of the test ensures that the samples size is large enough to detect a practically important difference in the standard deviation regardless of the distribution. However, if the data come from heavy-tailed distributions, the estimated sample size may be larger than the size that is actually required, which may lead to higher than necessary costs when sampling items.

Data checks

Unusual data

Unusual data are extremely large or small data values, also known as outliers. Unusual data can have a strong influence on the results of the analysis and can affect the chances of finding statistically significant results, especially when the sample is small. Unusual data can indicate problems with data collection, or may be due to unusual behavior of the process you are studying. Therefore, these data points are often worth investigating and should be corrected when possible.

Objective

We wanted to develop a method to check for data values that are very large or very small relative to the overall sample and that may affect the results of the analysis.

Method

We developed a method to check for unusual data based on the method described by Hoaglin, Iglewicz, and Tukey (1986) that is used to identify outliers in boxplots.

Results

The Assistant identifies a data point as unusual if it is more than 1.5 times the interquartile range beyond the lower or upper quartile of the distribution. The lower and upper quartiles are the 25th and 75th percentiles of the data. The interquartile range is the difference between the two quartiles. This method works well even when there are multiple outliers because it makes it possible to detect each specific outlier.

When checking for unusual data, the Assistant displays the following status indicators in the Report Card:

| Status | Condition |
|---|--|
|  | There are no unusual data points. |
|  | At least one data point is unusual and may have a strong influence on the results. |

Validity of test

In the 1-Sample Standard Deviation Methods section above, we showed that Bonett’s method generally provides better results than the AdjDF method. However, when the tails of a distribution are heavier, Bonett’s method requires larger sample sizes to achieve accurate results. Thus, a method for assessing the validity of the test must be based on not only the sample size but also on the heaviness of the tails of the parent distribution. Gel et al. (2007) developed a test to determine whether a sample comes from a distribution with heavy tails. This test, called the SJ test, is based on the ratio of the sample standard deviation (s) and the tail estimator J (for details, see Appendix E).

Objective

For a given sample of data, we needed to develop a rule to assess the validity of Bonett’s method by evaluating the heaviness of the tails in the data.

Method

We performed simulations to investigate the power of the SJ test to identify heavy-tailed distributions. If the SJ test is powerful for moderately large samples, then it can be used to discriminate between heavy-tailed and light-tailed distributions for our purposes. For more details, see Appendix F.

Results

Our simulations showed that when samples are large enough, the SJ test can be used to discriminate between heavy-tailed and light-tailed distributions. For samples of moderate or large size, smaller p-values indicate heavier tails and larger p-values indicate lighter tails. However, because larger samples tend to have smaller p-values than smaller samples, we also consider the sample size when determining the heaviness of the tails. Therefore, we developed our set of rules for the Assistant to classify the tails of the distribution for each sample based on the both sample size and the p-value of the SJ test. To see the specific ranges of p-values and sample sizes associated with light, moderate, and heavy-tailed distributions, see Appendix F.

Based on these results, the Assistant Report Card displays the following status indicators to evaluate the validity of the 1-Standard Deviation test (Bonett’s method) for your sample data:

| Status | Condition |
|---|---|
|  | There is no evidence that your sample has heavy tails. Your sample size is large enough to reliably check for this condition. OR Your sample has moderately heavy or heavy tails. However, your sample size is large enough to compensate, so the p-value should be accurate. |

| Status | Condition |
|---|---|
|  | <p>Your sample has moderately heavy or heavy tails. Your sample size is not large enough to compensate. Use caution when interpreting the results.</p> <p>OR</p> <p>Your sample is not large enough to reliably check for heavy tails. Use caution when interpreting the results.</p> |

Sample size

Typically, a statistical hypothesis test is performed to gather evidence to reject the null hypothesis of “no difference”. If the sample is too small, the power of the test may not be adequate to detect a difference that actually exists, which results in a Type II error. It is therefore crucial to ensure that the sample sizes are sufficiently large to detect practically important differences with high probability.

Objective

If the data does not provide sufficient evidence to reject the null hypothesis, we want to determine whether the sample sizes are large enough for the test to detect practical differences of interest with high probability. Although the objective of sample size planning is to ensure that sample sizes are large enough to detect important differences with high probability, they should not be so large that meaningless differences become statistically significant with high probability.

Method

The power and sample size analysis for the 1-Sample Standard Deviation test is based on the theoretical power function of the test. This power function provides good estimates when data have nearly normal tails or light tails, but may produce conservative estimates when the data have heavy tails (see the simulation results summarized in Performance of Theoretical Power Function in the 1-Sample Standard Deviation Methods section above).

Results

When the data does not provide enough evidence against the null hypothesis, the Assistant uses the power function of the normal approximation test to calculate the practical differences that can be detected with an 80% and a 90% probability for the given sample size. In addition, if the user provides a particular practical difference of interest, the Assistant uses the power function of the normal approximation test to calculate sample sizes that yield an 80% and a 90% chance of detection of the difference.

To help interpret the results, the Assistant Report Card for the 1-Sample Standard Deviation test displays the following status indicators when checking for power and sample size:

| Status | Condition |
|---|--|
|  | <p>The test finds a difference between the standard deviation and the target value, so power is not an issue.</p> <p>OR</p> <p>Power is sufficient. The test did not find a difference between the standard deviation and the target value, but the sample is large enough to provide at least a 90% chance of detecting the given difference.</p> |
|  | <p>Power may be sufficient. The test did not find a difference between the standard deviation and the target value, but the sample is large enough to provide an 80% to 90% chance of detecting the given difference. The sample size required to achieve 90% power is reported.</p> |
|  | <p>Power might not be sufficient. The test did not find a difference between the standard deviation and the target value, and the sample is large enough to provide a 60% to 80% chance of detecting the given difference. The sample sizes required to achieve 80% power and 90% power are reported.</p> |
|  | <p>Power is not sufficient. The test did not find a difference between the standard deviation and the target value, and the sample is not large enough to provide at least a 60% chance of detecting the given difference. The sample sizes required to achieve 80% power and 90% power are reported.</p> |
|  | <p>The test did not find a difference between the standard deviation and the target value. You did not specify a practical difference to detect. Depending on your data, the report may indicate the differences that you could detect with 80% and 90% chance, based on your sample size and alpha.</p> |

References

- 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.

Appendix A: Methods for testing variance (or standard deviation)

The table below summarizes strengths and weaknesses associated with various methods for testing the variance.

| Method | Comment |
|---|---|
| Classical chi-square procedure | Extremely sensitive to normality assumption. Even small departures from normality can produce inaccurate results regardless of how large the sample is. In fact, when the data deviate from normality, increasing the sample size decreases the accuracy of the procedure. |
| Large-sample method based on the asymptotic normal distribution of the logarithmic-transform of the sample variance | Generally better than the classical chi-square method but requires larger sample sizes to be insensitive to the normality assumption. |
| Large-sample method based on Edgeworth expansion for one-sided (upper tail) tests See Lee and Ping (1996). | Produces acceptable Type I error rates, but requires that data come from a symmetric distribution. |
| Large-sample method based on an approximation of the distribution of the sample variance by a scaled chi-square distribution. The method is referred to as the Adjusted degrees of freedom (AdjDF) method. See Hummel, Banga, and Hettmansperger (2005). | Provides better coverage probability than the method based on the asymptotic normal distribution of the logarithmic-transform of the sample variance and the nonparametric ABC bootstrap approximation method for confidence intervals (Efron and Tibshirani, 1993). Used for the 1 Variance test in Minitab 15. |
| Bonett's Adjusted asymptotic distribution of the logarithmic-transform of the sample variance See Bonett (2006). | Provides good coverage probability for confidence intervals even for moderately large samples. However, requires much larger samples when data come from heavy-tailed distributions. Used for the 1 Variance test and the Assistant 1-Sample Standard Deviation test in Minitab 16. |

Appendix B: Definition of Bonett's method and the AdjDF method

Let x_1, \dots, x_n be an observed random sample of size n from a population with a finite fourth moment. Let \bar{x} and s be the observed sample mean and standard deviation, respectively. Also, let γ and γ_e be the population kurtosis and kurtosis excess, respectively, so that $\gamma_e = \gamma - 3$. Thus, for a normal population, $\gamma = 3$ and $\gamma_e = 0$. Also, let σ^2 be the unknown population variance. In the sections that follow, we present two methods for making an inference about σ^2 , the adjusted degrees of freedom (AdjDF) method and Bonett's method.

Formula B1: AdjDF method

The AdjDF method is based upon an approximation of the distribution of the sample variance by a scaled chi-square distribution (see Box, 1953). More specifically, the first two moments of the sample variance are matched with the moments of a scaled chi-square distribution to determine the unknown scale and degrees of freedom. This approach yields the following approximate two-sided $(1 - \alpha)100$ percent confidence interval for the variance:

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

where

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

This estimate of the kurtosis excess is identical to the one used for the Basic Statistics commands in Minitab.

Formula B2: Bonett's method

Bonett's method relies on the well-known classical approach, which uses the central limit theorem and the Cramer δ method to obtain an asymptotic distribution of the log-transform of the sample variance. The log-transformation is used to accelerate convergence to normality. Using this approach, the approximate two-sided $(1 - \alpha)100$ percent confidence interval for the variance is defined as:

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

where z_α is the upper percentile of the standard normal distribution, and se is an asymptotic estimate of the standard error of the log-transformed sample variance, given as:

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

Previously, Hummel et al. (2005) performed simulation studies that demonstrated that the AdjDF method is superior to this classical approach. However, Bonett makes two adjustments to the classical approach to overcome its limitations.

The first adjustment involves the estimate of the kurtosis. To estimate kurtosis, Bonett uses the following formula:

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

where m is a trimmed mean with trim-proportion equal to $1/2\sqrt{n-4}$. This estimate of kurtosis tends to improve the accuracy of the confidence levels for heavy-tailed (symmetric or skewed) distributions.

For the second adjustment, Bonett empirically determines a constant multiplier for the sample variance and the standard error. This constant multiplier approximately equalizes the tail probabilities when the sample is small, and is given as:

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

These adjustments yield Bonett's approximate two-sided $(1 - \alpha)100$ percent confidence interval for the variance:

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

Appendix C: Accuracy of Bonett's method versus AdjDF method

Simulation C1: Comparison of confidence intervals

We wanted to compare the accuracy of the confidence intervals for the variance that are calculated using AdjDF method and Bonett's method. We generated random samples of different sizes ($n = 20, 30, 40, 50, 60, 80, 100, 150, 200, 250, 300$) from several distributions and calculated the confidence intervals using each method. The distributions included:

- Standard normal distribution ($N(0,1)$)
- Symmetric and light-tailed distributions, including the uniform distribution ($U(0,1)$) and the Beta distribution with both parameters set to 3 ($B(3,3)$)
- Symmetric and heavy-tailed distributions, including t-distributions with 5 and 10 degrees of freedom ($t(5), t(10)$), and the Laplace distribution with location 0 and scale 1 (Lpl)
- Skewed and heavy-tailed distributions, including the exponential distribution with scale 1 (Exp) and chi-square distributions with 3, 5, and 10 degrees of freedom ($Chi(3), Chi(5), Chi(10)$)
- Left-skewed and heavy-tailed distribution; specifically, the Beta distribution with the parameters set to 8 and 1, respectively ($B(8,1)$)

In addition, to assess the direct effect of outliers, we generated samples from contaminated normal distributions defined as

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

where p is the mixing parameter and $1 - p$ is the proportion of contamination (which equals the proportion of outliers). We selected two contaminated normal populations for the study: $CN(0.9,3)$, where 10% of the population are outliers; and $CN(0.8,3)$, where 20% of the population are outliers. These two distributions are symmetric and have long tails due to the outliers.

For each sample size, 10,000 sample replicates were drawn from each distribution and the two-sided 95% confidence intervals were calculated using each method. The random sample generator was seeded so that both methods were applied to the same samples. Based on these confidence intervals, we then calculated the simulated coverage probabilities (CovP) and average interval widths (AveW) for each method. If the confidence intervals of the two methods have about the same simulated coverage probabilities, then the method that produces shorter intervals (on average) is more precise. Because we used a target confidence level of 95%, the simulation error was $\sqrt{0.95(0.05)/10,000} = 0.2\%$.

The simulation results are recorded in Tables 1 and 2 below.

Table 1 Simulated coverage probabilities of 95% two-sided confidence intervals for the variance calculated using the AdjDF and Bonett’s methods. These samples were generated from symmetric distributions with light, normal, nearly normal, or heavy tails.

| Distribution | | Symmetric Distributions with Light, Normal, or Nearly Normal Tails | | | | Symmetric Distributions with Heavy Tails | | | |
|-------------------------|------|--|--------|--------|-------|--|------------|-------------|--------|
| | | U(0,1) | B(3,3) | N(0,1) | t(10) | Lpl | CN(0.8, 3) | CN (0.9, 3) | T(5) |
| Skewness | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Kurtosis (γ_e) | | -1.200 | -0.667 | 0 | 1.000 | 3.000 | 4.544 | 5.333 | 6.000 |
| n = 10 | | | | | | | | | |
| AdjDF | CovP | 0.910 | 0.909 | 0.903 | 0.883 | 0.853 | 0.793 | 0.815 | 0.858 |
| | AveW | 0.154 | 0.087 | 3.276 | 5.160 | 13.924 | 21.658 | 14.913 | 11.742 |
| Bonett | CovP | 0.972 | 0.967 | 0.962 | 0.952 | 0.919 | 0.891 | 0.920 | 0.935 |
| | AveW | 0.242 | 0.115 | 3.710 | 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 | 0.100 | 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 | 1.170 | 1.764 | 4.117 | 6.330 | 4.557 | 3.667 |
| Bonett | CovP | 0.951 | 0.950 | 0.947 | 0.933 | 0.909 | 0.869 | 0.852 | 0.907 |
| | AveW | 0.070 | 0.037 | 1.221 | 1.750 | 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 | 0.890 |
| | 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 |

| Distribution | Symmetric Distributions with Light, Normal, or Nearly Normal Tails | | | | Symmetric Distributions with Heavy Tails | | | |
|-------------------------|--|--------|--------|-------|--|------------|-------------|-------|
| | U(0,1) | B(3,3) | N(0,1) | t(10) | Lpl | CN(0.8, 3) | CN (0.9, 3) | T(5) |
| Skewness | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Kurtosis (γ_e) | -1.200 | -0.667 | 0 | 1.000 | 3.000 | 4.544 | 5.333 | 6.000 |

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 |
| Bonett | CovP | 0.951 | 0.947 | 0.944 | 0.938 | 0.918 | 0.888 | 0.855 | 0.908 |
| | AveW | 0.049 | 0.026 | 0.870 | 1.280 | 2.613 | 3.939 | 2.729 | 2.240 |

n = 60

| | | | | | | | | | |
|--------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| AdjDF | CovP | 0.949 | 0.943 | 0.938 | 0.926 | 0.913 | 0.890 | 0.853 | 0.899 |
| | AveW | 0.040 | 0.022 | 0.766 | 1.155 | 2.490 | 3.857 | 2.768 | 2.283 |
| Bonett | CovP | 0.949 | 0.947 | 0.943 | 0.935 | 0.918 | 0.896 | 0.868 | 0.910 |
| | 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 | 0.940 | 0.930 | 0.913 | 0.890 | 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 | 2.300 | 1.906 |
| Bonett | CovP | 0.946 | 0.950 | 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 | 0.920 | 0.910 | 0.874 | 0.909 |
| | AveW | 0.034 | 0.018 | 0.618 | 0.916 | 1.882 | 2.822 | 1.984 | 1.646 |

| Distribution | Symmetric Distributions with Light, Normal, or Nearly Normal Tails | | | | Symmetric Distributions with Heavy Tails | | | |
|-------------------------|--|--------|--------|-------|--|------------|-------------|-------|
| | U(0,1) | B(3,3) | N(0,1) | t(10) | Lpl | CN(0.8, 3) | CN (0.9, 3) | T(5) |
| Skewness | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Kurtosis (γ_e) | -1.200 | -0.667 | 0 | 1.000 | 3.000 | 4.544 | 5.333 | 6.000 |

n = 100

| | | | | | | | | | |
|--------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| AdjDF | CovP | 0.947 | 0.951 | 0.945 | 0.933 | 0.920 | 0.910 | 0.885 | 0.912 |
| | AveW | 0.030 | 0.017 | 0.576 | 0.873 | 1.830 | 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 |

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 | 0.700 | 1.470 | 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 | 0.930 | 0.914 | 0.918 |
| | AveW | 0.021 | 0.012 | 0.400 | 0.605 | 1.265 | 1.906 | 1.373 | 1.178 |
| Bonett | CovP | 0.942 | 0.951 | 0.949 | 0.940 | 0.928 | 0.930 | 0.915 | 0.920 |
| | AveW | 0.021 | 0.012 | 0.402 | 0.603 | 1.245 | 1.860 | 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 | 1.120 | 1.690 | 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 | 1.190 | 0.986 |

n = 300

| | | | | | | | | | |
|--------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| AdjDF | CovP | 0.950 | 0.948 | 0.951 | 0.940 | 0.938 | 0.936 | 0.920 | 0.914 |
| | AveW | 0.017 | 0.009 | 0.324 | 0.490 | 1.019 | 1.544 | 1.115 | 0.933 |
| Bonett | CovP | 0.950 | 0.947 | 0.951 | 0.942 | 0.937 | 0.929 | 0.920 | 0.916 |
| | AveW | 0.017 | 0.010 | 0.325 | 0.489 | 1.009 | 1.657 | 1.093 | 0.897 |

Table 2 Simulated coverage probabilities of 95% two-sided confidence intervals for the variance calculated using the AdjDF and Bonett's methods. These samples were generated from skew distributions with nearly normal, moderately heavy, or heavy tails.

| Distribution | | Skew Distributions with Nearly Normal or Moderately Heavy Tails | | | Skew Distributions with Heavy Tails | |
|-------------------------|------|---|--------|--------|-------------------------------------|--------|
| | | Chi(10) | B(8,1) | Chi(5) | Chi(3) | Exp |
| Skewness | | 0.894 | -1.423 | 1.265 | 1.633 | 2 |
| Kurtosis (γ_e) | | 1.200 | 2.284 | 2.400 | 4.000 | 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 | 0.950 | 0.917 | 0.938 | 0.911 | 0.882 |
| | AveW | 91.006 | 0.058 | 53.830 | 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 | 41.600 | 0.026 | 24.094 | 17.232 | 3.370 |
| n = 30 | | | | | | |
| AdjDF | CovP | 0.901 | 0.881 | 0.880 | 0.864 | 0.838 |
| | AveW | 30.021 | 0.018 | 18.182 | 13.630 | 2.844 |
| Bonett | CovP | 0.931 | 0.920 | 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 | 0.930 | 0.915 | 0.913 | 0.904 | 0.898 |
| | AveW | 24.952 | 0.015 | 14.504 | 1.991 | 1.991 |

| Distribution | Skew Distributions with Nearly Normal or Moderately Heavy Tails | | | Skew Distributions with Heavy Tails | |
|-------------------------|---|--------|--------|-------------------------------------|-----|
| | Chi(10) | B(8,1) | Chi(5) | Chi(3) | Exp |
| Skewness | 0.894 | -1.423 | 1.265 | 1.633 | 2 |
| Kurtosis (γ_e) | 1.200 | 2.284 | 2.400 | 4.000 | 6 |

n = 50

| | | | | | | |
|--------|------|--------|-------|--------|-------|-------|
| AdjDF | CovP | 0.912 | 0.900 | 0.892 | 0.871 | 0.868 |
| | AveW | 21.373 | 0.013 | 12.694 | 9.115 | 1.861 |
| Bonett | CovP | 0.930 | 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 | 0.890 | 0.875 |
| | AveW | 18.928 | 0.011 | 11.338 | 8.211 | 1.645 |
| Bonett | CovP | 0.930 | 0.933 | 0.923 | 0.917 | 0.900 |
| | AveW | 19.369 | 0.012 | 11.456 | 8.093 | 1.554 |

n = 70

| | | | | | | |
|--------|------|--------|-------|--------|-------|-------|
| AdjDF | CovP | 0.915 | 0.910 | 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 | 0.920 | 0.916 | 0.911 | 0.904 | 0.890 |
| | 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 | 15.250 | 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 | 1.230 |

| Distribution | Skew Distributions with Nearly Normal or Moderately Heavy Tails | | | Skew Distributions with Heavy Tails | |
|-------------------------|---|--------|--------|-------------------------------------|-----|
| | Chi(10) | B(8,1) | Chi(5) | Chi(3) | Exp |
| Skewness | 0.894 | -1.423 | 1.265 | 1.633 | 2 |
| Kurtosis (γ_e) | 1.200 | 2.284 | 2.400 | 4.000 | 6 |

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 |

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 | 0.930 | 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 | 0.640 |
| Bonett | CovP | 0.947 | 0.948 | 0.943 | 0.941 | 0.933 |
| | AveW | 8.314 | 0.005 | 4.862 | 3.419 | 0.651 |

Our results are very consistent with those published by Bonett (2006). As shown in Tables 1 and 2, the confidence intervals calculated using Bonett’s method are superior to the confidence intervals calculated using the AdjDF method because they yield coverage probabilities closer to the target level of 0.95 and narrower confidence intervals, on average. If the confidence intervals of the two methods have about the same simulated coverage probabilities, then the method that produces shorter intervals (on average) is more precise. This means that the statistical test for the variance based on Bonett’s method performs better and results in lower Type I and Type II error rates. When sample sizes are large, the two methods yield almost identical results, but for small to moderate sample sizes, Bonett’s method is superior.

Although Bonett’s method generally performs better than the AdjDF method, it consistently yields coverage probabilities below the target coverage of 0.95 for heavy-tailed distributions (symmetric or skewed) even for extremely large samples ($n > 100$). This is illustrated in Figure 1 below, which plots the simulated coverage probabilities for Bonett’s method against the true kurtosis excess of the population for small, moderate, and large sample sizes.

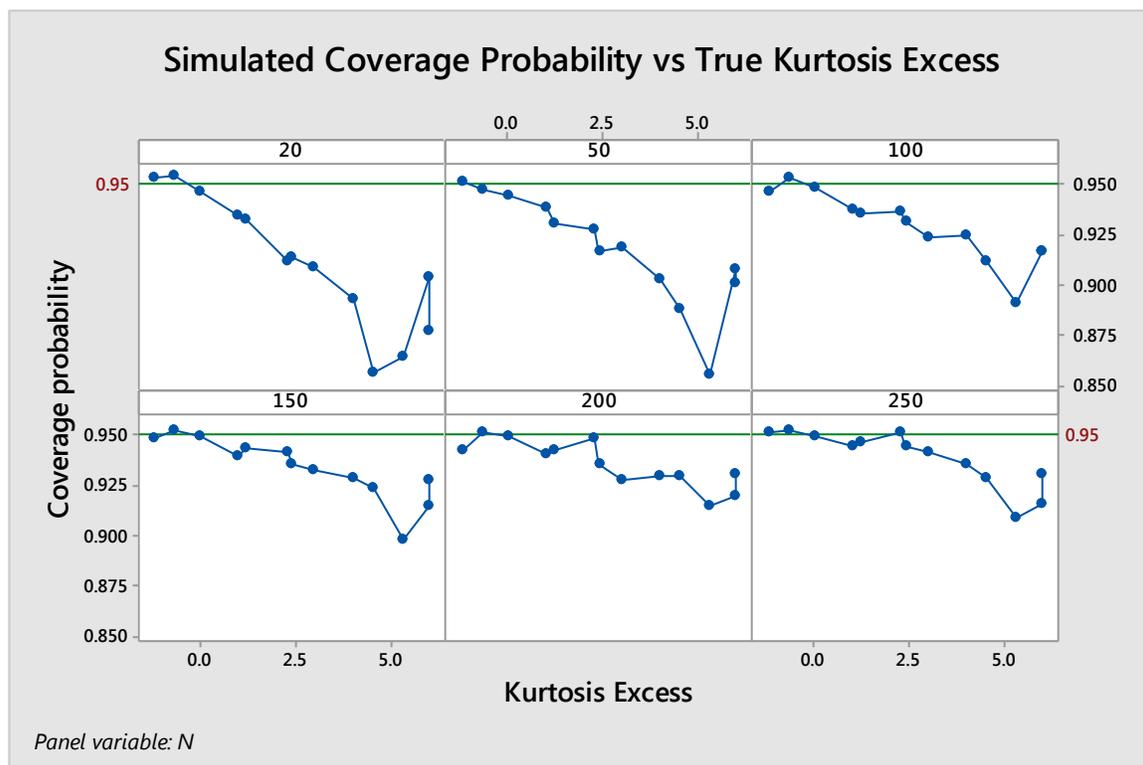


Figure 1 Simulated coverage probabilities for Bonett’s 95% confidence intervals plotted against the kurtosis excess of each distribution at various sample sizes.

As shown in Figure 1, the greater the kurtosis, the larger the sample size that is needed to make the simulated coverage probabilities approach the target level. As noted previously, the simulated coverage probabilities for Bonett’s method are low for heavy-tailed distributions. However, for lighter tailed distributions, such as the uniform and the Beta(3,3) distributions, the

simulated coverage probabilities are stable and on target for sample sizes as small as 20. Therefore, we base our criterion to determine the validity of Bonett's method upon both the sample size and the heaviness of the tails of the distribution from which the sample is drawn.

As a first step for developing this criterion, we classify the distributions into three categories according to the heaviness of their tails:

- **Light-tailed or normal-tailed distributions (L-type):** These are distributions for which Bonett's confidence intervals yield stable coverage probabilities near the target coverage level. For these distributions, sample sizes as low as 20 produce accurate results. Examples include the uniform distribution, the Beta(3,3) distribution, the normal distribution, the t distribution with 10 degrees of freedom, and the chi-square distribution with 10 degrees of freedom.
- **Moderately heavy-tailed distributions (M-type):** For these distributions, Bonett's method requires a minimum sample size of 80 for the simulated coverage probabilities to be close to the target coverage. Examples include the chi-square distribution with 5 degrees of freedom, and the Beta(8,1) distribution.
- **Heavy-tailed distributions (H-type):** These are distributions for which Bonett's confidence intervals yield coverage probabilities that are far below the targeted coverage, unless the sample sizes are extremely large ($n \geq 200$). Examples include the t distribution with 5 degrees of freedom, the Laplace distribution, the chi-square distribution with 3 degrees of freedom, the exponential distribution, and the two contaminated normal distributions, CN(0.9,3) and CN(0.8,3).

Thus, a general rule for evaluating the validity of Bonett's method requires that we develop a procedure to identify which of the 3 distribution types the sample data comes from. We developed this procedure as part of the Validity of test data check. For more details, see Appendix E.

Appendix D: Theoretical power

We derived the theoretical power function of the test associated with Bonett's method and performed simulations to compare the theoretical and simulated power of the test. If the theoretical and simulated power curves are close to each other, then the power and sample size analysis based upon on the theoretical power function should yield accurate results.

Formula D1: Theoretical power function for Bonett's method

As described earlier, Bonett's method is based upon the well-known classical approach, in which the central limit theorem and the Cramer δ method are used to find an asymptotic distribution of the log-transformed sample variance. More specifically, it is established that in large samples, $\frac{\ln S^2 - \ln \sigma^2}{se}$ is approximately distributed as the standard normal distribution. The denominator, se , is the large sample standard error of the log-transformed sample variance and is given as

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

where γ is the kurtosis of the unknown parent population.

It follows that an approximate power function with an approximate alpha level for the two-sided test using Bonett's method may be given as a function of the sample size, the ratio $\rho = \sigma/\sigma_0$ and the parent population kurtosis γ as

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

where σ_0 is the hypothesized value of the unknown standard deviation, Φ is the CDF of the standard normal distribution, and z_{α} is the upper α percentile point of the standard normal distribution. The one-sided power functions can also be obtained from these calculations.

Note that when planning the sample size for a study, an estimate of the kurtosis may be used in place of the true kurtosis. This estimate is usually based on the opinions of experts or the results of previous experiments. If that information is not available, it is often a good practice to perform a small pilot study to develop the plans for the major study. Using a sample from the pilot study, the kurtosis may be estimated as

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

where m is a trimmed mean with trim-proportion equal to $1/2\sqrt{n-4}$.

Simulation D1: Comparison of actual power versus theoretical power

We designed a simulation to compare estimated actual power levels (referred to as simulated power levels) to the theoretical power levels (referred to as approximate power levels) when using Bonett's method to test the variance.

In each experiment, we generated 10,000 sample replicates, each of size n , where $n = 20, 30, 40, 50, \dots, 120$, from each of the distributions described in Simulation C1 (see Appendix C). For each distribution and sample size n , we calculated the simulated power level as the fraction of 10,000 random sample replicates for which the two-sided test with alpha level $\alpha = 0.05$ was significant. When calculating the simulated power, we used $\rho = \sigma/\sigma_0 = 1.25$ to obtain relatively small power levels. We then calculated the corresponding power levels using the theoretical power function for comparison.

The results are shown in Tables 3 and 4 and graphically represented in Figure 2 below.

Table 3 Simulated power levels (evaluated at $\rho = \sigma/\sigma_0 = 1.25$) of a two-sided test for the variance based on Bonett's method compared with theoretical (normal approximation) power levels. The samples were generated from symmetric distributions with light, normal, nearly normal, or heavy tails.

| n | Power | Symmetric Distributions with Light, Normal, or Nearly Normal Tails | | | | Symmetric Distributions with Heavy Tails | | | |
|-----|---------|--|--------|--------|-------|--|-----------|-----------|-------|
| | | 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 |
| | Approx. | 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 |
| | Approx. | 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 |
| | Approx. | 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 |
| | Approx. | 0.921 | 0.754 | 0.586 | 0.431 | 0.284 | 0.229 | 0.210 | 0.196 |
| 60 | Simul. | 0.942 | 0.822 | 0.701 | 0.570 | 0.380 | 0.285 | 0.274 | 0.384 |
| | Approx. | 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 |
| | Approx. | 0.981 | 0.885 | 0.737 | 0.566 | 0.379 | 0.303 | 0.276 | 0.257 |

| n | Power | Symmetric Distributions with Light, Normal, or Nearly Normal Tails | | | | Symmetric Distributions with Heavy Tails | | | |
|-----|---------|--|--------|--------|-------|--|-----------|-----------|-------|
| | | U(0,1) | B(3,3) | N(0,1) | t(10) | Lpl | CN (.8,3) | CN (.9,3) | t(5) |
| 80 | Simul. | 0.981 | 0.909 | 0.815 | 0.689 | 0.481 | 0.372 | 0.347 | 0.483 |
| | Approx. | 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 |
| | Approx. | 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 |
| | Approx. | 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 |
| | Approx. | 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 |
| | Approx. | 1.000 | 0.987 | 0.928 | 0.799 | 0.584 | 0.476 | 0.434 | 0.405 |

Table 4 Simulated power levels (evaluated at $\rho = \sigma/\sigma_0 = 1.25$) of a two-sided test for the variance based on Bonett's method compared with theoretical (normal approximation) power levels. The samples were generated from skew distributions with nearly normal, moderately heavy, or heavy tails.

| n | Power | Skewed Distributions with Nearly Normal or Moderately Heavy Tails | | | Skewed Distributions with Heavy Tails | |
|----|---------|---|--------|--------|---------------------------------------|-------|
| | | Chi(10) | B(8,1) | Chi(5) | Chi(3) | Exp |
| 20 | Simul. | 0.222 | 0.166 | 0.172 | 0.139 | 0.128 |
| | Approx. | 0.186 | 0.152 | 0.149 | 0.123 | 0.104 |
| 30 | Simul. | 0.314 | 0.216 | 0.234 | 0.190 | 0.151 |
| | Approx. | 0.263 | 0.263 | 0.205 | 0.164 | 0.135 |
| 40 | Simul. | 0.387 | 0.266 | 0.292 | 0.223 | 0.186 |
| | Approx. | 0.338 | 0.266 | 0.261 | 0.204 | 0.165 |
| 50 | Simul. | 0.455 | 0.324 | 0.349 | 0.263 | 0.208 |
| | Approx. | 0.409 | 0.323 | 0.316 | 0.245 | 0.196 |
| 60 | Simul. | 0.521 | 0.376 | 0.399 | 0.302 | 0.239 |
| | Approx. | 0.477 | 0.377 | 0.369 | 0.286 | 0.227 |

| <i>n</i> | Power | Skewed Distributions with Nearly Normal or Moderately Heavy Tails | | | Skewed Distributions with Heavy Tails | |
|-----------------|---------|---|--------|--------|---------------------------------------|-------|
| | | Chi(10) | B(8,1) | Chi(5) | Chi(3) | Exp |
| 70 | Simul. | 0.583 | 0.419 | 0.463 | 0.361 | 0.269 |
| | Approx. | 0.539 | 0.430 | 0.420 | 0.325 | 0.257 |
| 80 | Simul. | 0.646 | 0.473 | 0.499 | 0.394 | 0.299 |
| | Approx. | 0.597 | 0.479 | 0.469 | 0.365 | 0.288 |
| 90 | Simul. | 0.688 | 0.517 | 0.561 | 0.428 | 0.327 |
| | Approx. | 0.649 | 0.526 | 0.516 | 0.403 | 0.318 |
| 100 | Simul. | 0.738 | 0.561 | 0.591 | 0.469 | 0.368 |
| | Approx. | 0.695 | 0.571 | 0.560 | 0.440 | 0.347 |
| 110 | Simul. | 0.779 | 0.608 | 0.637 | 0.495 | 0.394 |
| | Approx. | 0.737 | 0.611 | 0.600 | 0.475 | 0.376 |
| 120 | Simul. | 0.810 | 0.635 | 0.679 | 0.538 | 0.416 |
| | Approx. | 0.774 | 0.650 | 0.638 | 0.509 | 0.405 |

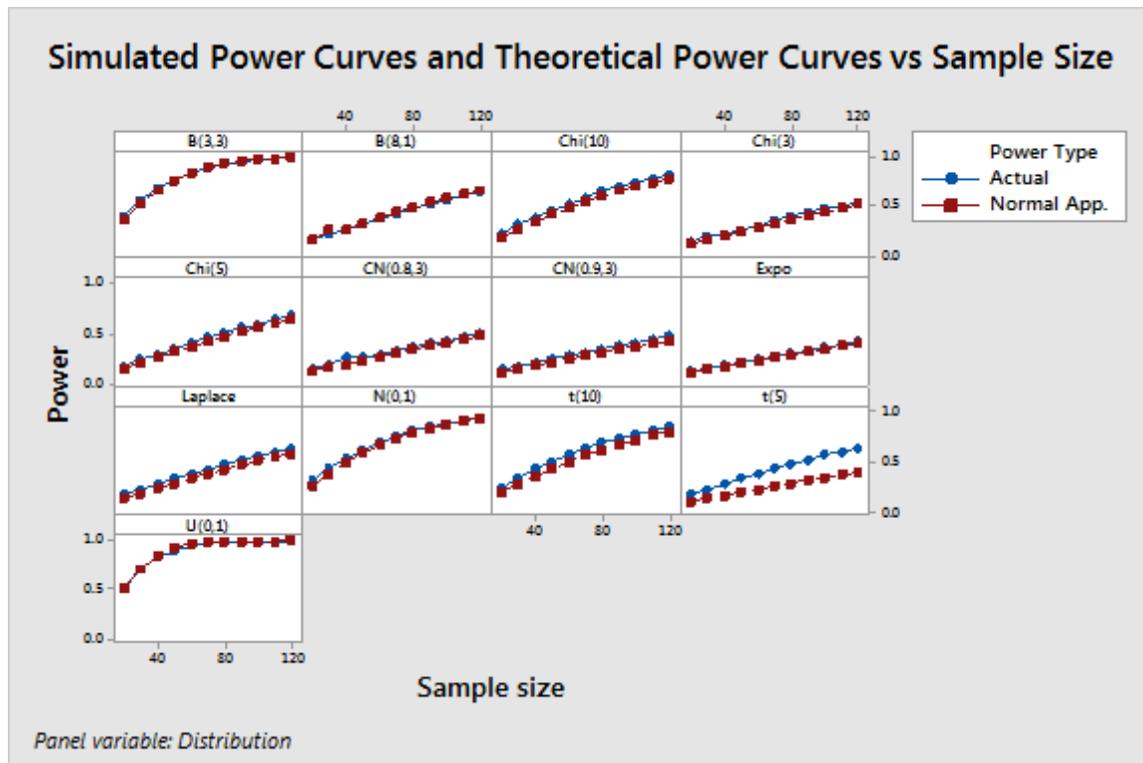


Figure 2 Simulated power curves compared with theoretical power curves for various distributions

The results in Tables 3 and 4 and Figure 2 show that when samples are generated from distributions with lighter tails (L-type distributions, as defined in Appendix C), such as the uniform distribution, the Beta (3,3) distribution, the normal distribution, the t distribution with 10 degrees of freedom, and the chi-square distribution with 10 degrees of freedom, the theoretical power values and the simulated power levels are practically undistinguishable.

However, for distributions with heavy tails (H-type distributions), the simulated power curves are markedly above the theoretical power curves when the samples are small. These heavy-tailed distributions include the t distribution with 5 degrees of freedom, the Laplace distribution, the chi-square distribution with 3 degrees of freedom, the exponential distribution, and the two contaminated normal distributions, CN(0.9,3) and CN(0.8,3). Therefore, when planning the sample size for a study and the sample comes from a distribution with heavy tails, the sample size estimated by the theoretical power function may be larger than the actual sample size required to achieve a given target power.

Appendix E: The SJ test for normal versus heavy tails

The results of the simulation study in Appendix C showed that when the tails of the distribution are heavier, larger sample sizes are required for the simulated coverage probability of Bonett's confidence intervals to approach the target level. Skewness, however, did not appear to have a significant effect on the simulated coverage probabilities.

Therefore, we needed to develop a criterion to assess the validity of Bonett's method based both on the size of the sample and the heaviness of the tails of the distribution from which the sample is drawn. Fortunately, Gel et al. (2007) provide a reasonably powerful test for directly testing the null hypothesis that the distribution has normal tails against the alternative hypothesis that the distribution has heavy tails. The test, which we refer to as the SJ test, is based upon the following statistic:

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

where S is the sample standard deviation, \hat{j} is the estimate of the sample mean absolute deviation from the median, m , and is given as

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

An approximate size- α test against the alternative hypothesis of heavy tails rejects the null hypothesis of normal tails if

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

where z_α is the upper α -percentile of a standard normal distribution and $\sigma_R = (\pi - 3)/2$.

Gel et al. (2007) have shown that replacing the upper α -percentile of the standard normal distribution with that of the t distribution with $(\sqrt{n} + 3)/2$ degrees of freedom provides better approximations for moderate sample sizes. Therefore, when applying the SJ test for the Validity of test data check, we replace z_α with $t_{d,\alpha}$, the upper α -percentile of the t-distribution with $d = (\sqrt{n} + 3)/2$ degrees of freedom.

Appendix F: Validity of test

Simulation F1: Using simulated power of SJ test to determine distribution classifications

We performed simulations to investigate the power of the SJ test. We generated samples of various sizes ($n = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200$) from various distributions. The distributions had normal, light, moderate, or heavy tails, and are the same as those described in simulation C1 (see Appendix C). For each given sample size, 10,000 sample replicates were drawn from each distribution. We calculated the simulated power of the SJ test as the proportion of cases for which the null hypothesis (that the parent distribution has normal tails) was rejected. In addition, we calculated the average R values (AveR) and the average p -values (AvePV).

The simulation results are shown in Tables 5 and 6 below.

Table 5 Simulated power levels of the SJ test. The samples were generated from symmetric distributions with light, normal, nearly normal, or heavy tails.

| Distribution | | Symmetric Distributions with Light, Normal, or Nearly Normal Tails | | | | Symmetric Distributions with Heavy Tails | | | |
|-----------------------|-------|--|--------|--------|-------|--|-----------|-----------|-------|
| | | U(0,1) | B(3,3) | N(0,1) | t(10) | Lpl | CN (.8,3) | CN (.9,3) | t(5) |
| n | TrueR | 0.921 | 0.965 | 1.0 | 1.032 | 1.128 | 1.152 | 1.118 | 1.085 |
| 10 | Power | 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 | Power | 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 | Power | 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 | Power | 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 |

| Distribution | | Symmetric Distributions with Light, Normal, or Nearly Normal Tails | | | | Symmetric Distributions with Heavy Tails | | | |
|--------------|-------|--|--------|--------|-------|--|-----------|-----------|-------|
| | | U(0,1) | B(3,3) | N(0,1) | t(10) | Lpl | CN (.8,3) | CN (.9,3) | t(5) |
| n | TrueR | 0.921 | 0.965 | 1.0 | 1.032 | 1.128 | 1.152 | 1.118 | 1.085 |
| 30 | Power | 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 | Power | 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 | Power | 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 | Power | 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 | Power | 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 | Power | 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 | Power | 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 |
| 100 | Power | 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 | Power | 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 |

| Distribution | | Symmetric Distributions with Light, Normal, or Nearly Normal Tails | | | | Symmetric Distributions with Heavy Tails | | | |
|--------------|-------|--|--------|--------|-------|--|-----------|-----------|-------|
| | | U(0,1) | B(3,3) | N(0,1) | t(10) | Lpl | CN (.8,3) | CN (.9,3) | t(5) |
| n | TrueR | 0.921 | 0.965 | 1.0 | 1.032 | 1.128 | 1.152 | 1.118 | 1.085 |
| 140 | Power | 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 | Power | 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 | Power | 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 | Power | 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 |

Table 6 Simulated power levels of the SJ test. The samples were generated from skew distributions with nearly normal, moderately heavy, or heavy tails.

| Distributions | | Skew Distributions with Nearly Normal or Moderately Heavy Tails | | | Skew Distributions with Heavy Tails | |
|---------------|-------|---|--------|--------|-------------------------------------|-------|
| | | Chi(10) | B(8,1) | Chi(5) | Chi(3) | Exp |
| n | TrueR | 1.028 | 1.075 | 1.059 | 1.098 | 1.151 |
| 10 | Power | 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 | Power | 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 | Power | 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 |

| Distributions | | Skew Distributions with Nearly Normal or Moderately Heavy Tails | | | Skew Distributions with Heavy Tails | |
|---------------|-------|---|--------|--------|-------------------------------------|-------|
| | | Chi(10) | B(8,1) | Chi(5) | Chi(3) | Exp |
| n | TrueR | 1.028 | 1.075 | 1.059 | 1.098 | 1.151 |
| 25 | Power | 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 | Power | 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 | Power | 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 | Power | 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 |
| 60 | Power | 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 | Power | 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 | Power | 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 | Power | 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 | Power | 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 |

| Distributions | | Skew Distributions with Nearly Normal or Moderately Heavy Tails | | | Skew Distributions with Heavy Tails | |
|---------------|-------|---|--------|--------|-------------------------------------|-------|
| | | Chi(10) | B(8,1) | Chi(5) | Chi(3) | Exp |
| n | TrueR | 1.028 | 1.075 | 1.059 | 1.098 | 1.151 |
| 120 | Power | 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 | Power | 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 | Power | 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 |
| 180 | Power | 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 | Power | 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 |

Our simulation results in Tables 5 and 6 are consistent with those published in Gel et al. (2007). When the samples are from the normal populations, the simulated power levels (which in this case represent the actual significance level of the test) are not far from the target level, even for sample sizes as low as 25. When the samples are from heavy-tailed distributions, the power of the test is low for small sample sizes but increases to at least 40% when the sample size reaches 40. Specifically, the power at sample size 40 is about 40.1% for the t-distribution with 5 degrees of freedom, 66.5% for the Laplace distribution, and 65.1% for the exponential distribution.

For light-tailed distributions (the Beta(3,3) and the uniform distributions), the power of the test is near 0 for small samples and decreases even further as the sample size increases. This is not surprising because the evidence for these distributions actually supports the alternative hypothesis of a lighter tailed distribution, rather than the alternative hypothesis of a heavier-tailed distribution.

When the samples are from distributions with slightly heavier tails, such as the t-distribution with 10 degrees of freedom or the chi-square distribution with 10 degrees of freedom, the power levels are low for moderate to large sample sizes. For our purposes, this is actually a good result because the test for one variance (standard deviation) performs well for these

distributions and we do not want these distributions to be flagged as heavy-tailed. However, as the sample size increases, the power of the test increases, so these slightly heavy-tailed distributions are detected as heavy-tailed distributions.

Therefore, the rules for evaluating the tail weight of the distribution for this test must also take into consideration the size of the sample. One approach for doing this is to calculate a confidence interval for the measure of the tail weight; however, the distribution of the SJ statistic is extremely sensitive to the parent distribution of the sample. An alternative approach is to assess the heaviness of the tails of the distribution based on both the strength of the rejection of the null hypothesis of the SJ test and the sample size. More specifically, smaller p-values indicate heavier tails and larger p-values indicate lighter tails. However, larger samples tend to have smaller p-values than smaller samples. Therefore, based on the simulated power levels, sample sizes, and average p-values in Table 3, we devise a general set of rules for evaluating the tails of a distribution for each sample using the SJ test.

For moderate to large sample sizes ($40 \leq n \leq 100$), if the p-value is between 0.01 and 0.05, we deem that there is mild evidence against the null hypothesis. That is, the distribution of the sample is classified as a moderately heavy-tailed (M-type) distribution. On the other hand, if the p-value is below 0.01, then there is strong evidence against the null hypothesis, and the parent distribution of the sample is classified as a distribution with heavy tails (H-type).

For large samples ($n > 100$), we categorize the parent distribution as an M-type distribution if the p-value falls between 0.005 and 0.01, and as an H-type distribution if the p-value is extremely small (below 0.005). Note that when the sample size is below 40, the power of the SJ test is generally too low for the distribution of the sample to be effectively determined.

The general classification rules for the validity of the 1-variance test using Bonett's method are summarized in Table 7 below.

Table 7 Classification rules for identifying the parent distribution of each sample (p is the p-value of the SJ test)

| Condition | Distribution type |
|---|---------------------|
| $n < 40$ | None is determined |
| $100 \geq n \geq 40$ and $p > 0.05$ | L-type distribution |
| $n > 100$ and $p > 0.01$ | L-type distribution |
| $40 \leq n \leq 100$ and $0.01 < p \leq 0.05$ | M-type distribution |
| $n > 100$ and $0.005 < p \leq 0.01$ | M-type distribution |
| $40 \leq n \leq 100$ and $p \leq 0.01$ | H-type distribution |
| $n > 100$ and $p \leq 0.005$ | H-type distribution |

As indicated earlier, based on the results of Tables 1 and 2 in simulation C1, the approximate minimum sample size required to achieve a minimum of 0.93 coverage probability when samples are generated from an L-type, an M-type, and an H-type distribution is 20, 80, and 200, respectively. However, because the power of the SJ test is low for small samples, the minimum sample size requirement for L-type distributions is set at 40.

Simulation F2: Verifying the rules for classifying distributions

We generated samples from some of the distributions described in simulation C1 and used the SJ test to determine the proportions of the samples that were classified in one of the three distribution groups: L-type, M-type, and H-type. The simulation results are shown in Table 8.

Table 8 Fraction of 10,000 samples of different sizes from various distributions that are identified as L-type, M-type, and H-type

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

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

The results in Table 8 show that when samples are from light-tailed (L-type) and heavy-tailed (H-type) distributions, a higher proportion of the samples are correctly classified. For example, when samples of size 40 were generated from the Beta(3,3) distribution, 99.6% of the samples were correctly classified as having lighter tails; when samples of size 90 were generated from the Laplace distribution, 76.0% were correctly classified as having heavy tails. As a result, warning messages in the Report Card regarding the validity of the test are not wrongly issued when samples are truly from lighter tailed distributions, and are correctly issued when the sample comes from a distribution with heavy tails and the minimum sample size requirement is not met. In addition, for samples from distributions with moderately heavy tails (M-type), such as the chi-square (5) distribution, a higher proportion of samples are misclassified as being light-tailed (L-type) when the samples are small (for a sample size of 40, 66% of the samples are misclassified as light-tailed distribution). Consequently, for these cases, warning messages in the Report Card may not be issued even though the parent distributions have moderately heavy tails. However, when the sample size is greater than 80, the misclassification as an L-type distribution has no effect because the minimum sample size requirement has already been met.

© 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.