

本書は、Minitab 統計ソフトウェアのアシスタントで使用される方法およびデータチェックを開発するため、Minitab の統計専門家によって行われた調査に関する一連の文書群を構成する文書の 1 つです。

1 サンプル標準偏差検定

概要

1 サンプル標準偏差検定は、工程の変動性の推定、および目標値による変動性の比較に使用されます。一般に、変動性は分散または同等の標準偏差を使用して測定されます。

母集団の分散を評価するための多くの統計的手法が開発されており、それぞれに独自の長所と制限があります。分散の検定に使用される古典的なカイ二乗法が最も一般的ですが、この方法は正規性の仮定に大きく影響されるため、データが歪んでいる場合や裾が重い場合は結果が非常に不正確になる可能性があります。その他の方法も開発されていますが、それらの方法にも欠点があります。たとえば、一部の方法は対称な分布から生成された大きいサンプルまたはデータでのみ有効です（「付録 A」を参照）。

Minitab 15 では、Box (1953) によるサンプル分散の分布に対する尺度化されたカイ二乗近似に基づく、代替の大サンプル法を使用しています。「AdjDF (自由度調整済み) 法」と呼ばれるこの方法は、十分に大きいサンプルでは正規性の仮定による影響は少なく、その他の方法よりも信頼区間がより正確になることが示されています (Hummel, Banga, & Hettmansperger, 2005)。ただし、最近になって近似の信頼区間がより正確になるように修正された統計的手法が Bonett (2006) によって開発されました。

本書では、Bonett の方法の性能を評価します。さらに、サンプルサイズ計画において、Bonett の信頼区間に関連付けられた同等の検定手順の検出力関数を調査します。この結果に基づき、アシスタントでの 1 サンプル標準偏差検定には Bonett の方法を使用します。また、アシスタントのレポートカードで自動的に実行、表示される次のデータチェックを調査し、結果にどのような影響を及ぼすかも説明します。

- 異常なデータ
- 検定の妥当性
- サンプルサイズ

1 サンプル標準偏差の方法

Bonett の方法と AdjDF 法

Bonett の方法 (2006) が公表されるまで、母集団の分散を推測するための最も頑健な手順は AdjDF 法でした。Bonett によって公表された結果では、非正規母集団から中規模サイズのサンプルを抽出した場合、Bonett の方法は目標水準に近い安定した信頼水準を提供することが示されています。そのため、母集団の標準偏差または分散を推測するには Bonett の方法が効果的であると言えます。

目的

単一母集団の分散を推測する場合の Bonett の方法と AdjDF 法の性能を比較します。特に、さまざまなサイズのサンプルが非正規母集団から生成された場合に、分散 (または標準偏差) の信頼区間がより正確になるのはどちらの方法であるかを判断します。

Bonett の方法は信頼区間に直接適用されるため、信頼区間を比較します。Bonett の信頼区間に関連付けられた同等の仮説検定手順に基づいて調査することもできますが、Bonett の公表結果 (2006) と今回の結果を直接比較するため、仮説検定ではなく信頼区間を調査しました。

方法

AdjDF 法と Bonett の方法は、「付録 B」に正式に定義されています。各方法の信頼区間の精度を比較するため、次のシミュレーションを実行しました。まず、歪んでいて裾が重い分布、対称的で裾が重い分布、対称的で裾が軽い分布などの異なる特性を持つ分布から、さまざまなサイズのランダムサンプルを生成しました。サンプルサイズごとに各分布から 10,000 サンプル反復が抽出され、それぞれの方法を使用して分布の真の分散に対する両側 95% 信頼区間が計算されました。次に、シミュレートした被覆確率と呼ばれる、10,000 区間のうち真の分散が含まれる比率を計算しました。信頼区間が正確な場合、シミュレートした被覆確率は目標被覆確率である 0.95 に近くなります。さらに、各方法で信頼区間に関連付けられた平均幅を計算しました。2 つの方法の信頼区間でシミュレートした被覆確率がほぼ同じ場合、(平均で) より短い区間が得られる方法のほうが正確です。詳細は、「付録 C」を参照してください。

結果

通常、Bonett の方法は AdjDF 法よりも高い被覆確率と正確な信頼区間を得ることができません。その結果、Bonett の方法に基づく分散の統計検定の方がタイプ I 過誤率とタイプ II 過誤率が低くなります。そのため、アシスタントの 1 サンプル標準偏差検定は Bonett の方法に基づいています。

さらにこの結果には、Bonett の方法で裾が中程度から重い分布の場合に目標水準精度を達成するには、より大きいサンプルサイズが必要なことが示されています。

- 正規の裾または軽い裾を持つ分布の場合、サンプルサイズは 20 で十分です。
- やや重い裾を持つ分布の場合、サンプルサイズは 80 以上必要です。

- 重い裾を持つ分布の場合、サンプルサイズは 200 以上必要です。

そのため、データから妥当な 1 サンプル標準偏差検定または信頼区間の結果を確実に得るため、アシスタントにはサンプルサイズとデータ分布の裾の両方を同時に評価するデータチェックが含まれています（次の「検定の妥当性」のデータチェックを参照）。

理論的検出力の性能

Bonett の方法は、分散（または標準偏差）の信頼区間に直接適用されます。ただし、仮説検定と信頼区間の統計的関係を使用することで、Bonett の近似信頼区間に関連付けられた同等性検定が可能になります。この検定用の厳密な検出力関数は存在しないため、関数を導き出す必要があります。さらに、正規性の仮定に対する理論的検出力関数の感度を評価しません。

目的

Bonett の信頼区間に関連付けられた検定の理論的検出力関数を使用して、アシスタントの 1 サンプル標準偏差検定に必要な検出力とサンプルサイズを評価できるかどうかを判断します。これを行うには、正規データと非正規データが分析されるときに、この理論的検出力関数が検定の実際の検出力を正確に反映するかどうかを評価する必要があります。

方法

Bonett の方法を使用した検定の理論的検出力関数は、「付録 C」で導出されています。Bonett の方法を使用して実際の検出力水準（シミュレートした検出力水準）を推定するシミュレーションを実行しました。まず、前の分析で説明されているように、歪んでいて裾が重い分布、対称で裾が重い分布、および対称で裾が軽い分布から、さまざまなサイズのランダムサンプルを生成しました。各分布に対して 10,000 サンプル反復ごとに検定を実行し、各サンプルサイズに対して 10,000 サンプルのうち検定が有意である割合の差を検出する、検定のシミュレートした検出力を計算しました。さらに、比較するために検定の理論上の検出力関数を使用して、対応する検出力水準も計算しました。理論上の検出力関数が正規性に大きく影響されない限り、正規データと非正規データの理論上の検出力水準とシミュレートした検出力水準は近くなります。詳細は、「付録 D」を参照してください。

結果

このシミュレーションでは、正規の裾または軽い裾を持つ分布からサンプルが生成された場合、Bonett の方法を使用した検定の理論上の検出力とシミュレートした検出力はほぼ等しいことが示されています。一方、重い裾を持つ分布からサンプルが生成された場合、理論上の検出力関数は保守的で、目的の検出力を達成するために必要なサンプルサイズを高く推定してしまう可能性があります。そのため、検定の理論上の検出力関数では、分布に関係なく標準偏差の実質的に重要な差を検出するのに十分な大きさのサンプルサイズを確保します。ただし、重い裾を持つ分布からデータが生成された場合、推定されたサンプルサイズは実際に必要なサイズよりも大きくなり、項目のサンプリング時にコストが必要以上に高くなる可能性があります。

データチェック

異常なデータ

異常なデータとは、極端に大きいデータ値や極端に小さいデータ値（外れ値）のことです。異常なデータは、特にサンプルが小さい場合は分析結果に大きく影響し、統計的に有意な結果を得る確率に影響する可能性があります。異常なデータはデータ収集の問題を示していたり、分析している工程の異常動作が原因で発生する可能性があります。そのため、これらのデータ点は多くの場合調査する価値があり、可能な場合は修正する必要があります。

目的

全体のサンプルに対して非常に大きいか非常に小さいことが原因で、分析結果に影響を及ぼす可能性のあるデータ値をチェックする方法を開発します。



方法

Hoaglin, Iglewicz, and Tukey (1986) によって説明され、箱ひげ図の外れ値を特定するために使用される方法に基づき、異常なデータをチェックする方法を開発しました。

結果

アシスタントでは、分布の下位四分位数を下回る幅または上位四分位数を上回る幅が四分位間範囲の1.5倍より大きいデータ点は異常と識別されます。下位四分位数および上位四分位数とは、データの25番目および75番目の百分位数を指します。四分位間範囲は、2つの四分位数間の差です。複数の外れ値がある場合でもそれぞれの外れ値を検出できるため、この方法は有効に機能します。

異常なデータをチェックするときに、レポートカードに次のステータスインジケータが表示されます。

ステータス	状態
	異常なデータ点はありません。
	1つ以上のデータ点が異常で、結果に大きく影響する可能性があります。

検定の妥当性

上記の「1 サンプル標準偏差検定」セクションでは、一般に Bonett の方法は AdjDF 法よりも適切な結果が得られることを示しました。ただし、分布の裾が重くなるにつれ、Bonett の方法で正確な結果を得るには、より大きいサンプルサイズが必要になります。そのため、検定の妥当性を評価する方法は、サンプルサイズだけでなく親分布の裾の重さにも基づく必

要があります。Gel et al. (2007) は、重い裾を持つ分布からサンプルが生成されているかどうかを判断する検定を開発しました。「SJ 検定」と呼ばれるこの検定は、サンプル標準偏差 (S) と裾の推定 (J) の比に基づいています (詳細は「付録 E」を参照)。

目的

与えられたデータのサンプルに対してデータの裾の重さを評価することで、Bonett の方法の妥当性を評価する規則を定めます。



方法

裾が重い分布を識別する SJ 検定の検出力を調査するため、シミュレーションを実行しました。やや大きいサンプルで SJ 検定の検出力が高い場合、今回の目的である裾が重い分布と軽い分布の判別に SJ 検定を使用できます。詳細は、「付録 F」を参照してください。

結果

このシミュレーションで、サンプルが十分に大きい場合、裾が重い分布と軽い分布の判別に SJ 検定を使用できることがわかりました。サンプルサイズが中程度または大きい場合、小さい p 値は裾が重く、大きい p 値は裾が軽いことを示します。ただし、大きいサンプルでは小さいサンプルよりも p 値が小さくなる傾向にあるため、裾の重さの判別時にサンプルサイズも考慮する必要があります。そのため、SJ 検定のサンプルサイズと p 値の両方に基づき、サンプルごとの分布の裾を分類するアシスタントの規則セットを考案しました。軽い裾、中程度の裾、および重い裾を持つ分布に関連付けられた p 値とサンプルサイズの特定の範囲は、「付録 F」を参照してください。

これらの結果に基づき、サンプルデータの 1 サンプル標準偏差検定 (Bonett の方法) の妥当性を評価するときに、アシスタントレポートカードに次のステータスインジケータが表示されます。

ステータス	状態
	サンプルが重い裾を持つという証拠はありません。サンプルサイズは、この状態を確実にチェックするのに十分な大きさです。 または サンプルはやや重い裾または重い裾を持ちます。ただし、サンプルサイズは補正を行うために十分な大きさであるため、P 値は正確です。
	サンプルはやや重い裾または重い裾を持ちます。サンプルサイズは補正を行うために十分な大きさではありません。結果を解釈するときに注意が必要です。 または サンプルサイズは重い裾を確実にチェックするのに十分な大きさではありません。結果を解釈するときに注意が必要です。

サンプルサイズ

通常、統計的仮説検定は「差がない」という帰無仮説を棄却する証拠を集めるために実行されます。サンプルが小さすぎる場合、検定の検出力は実際に存在する差を検出するには不十

分で、タイプ II の誤りになる可能性があります。そのため、高確率で実質的に重要な差を検出するのに十分な大きさのサンプルサイズがあることが極めて重要です。

目的

データから帰無仮説を棄却する十分な証拠が得られない場合、検定で高確率で目的の実質的な差を検出するのに十分な大きさのサンプルサイズがあるかどうかを判断します。サンプルサイズ計画の目的は、高確率で重要な差を検出するのに十分な大きさのサンプルサイズを確保することですが、高確率で無意味な差が統計的に有意になってしまうほどサンプルサイズを大きくしないようにする必要があります。





方法


1 サンプル標準偏差検定の検出力とサンプルサイズの分析は、検定の理論上の検出力関数に基づいています。この検出力関数は、ほぼ正規の裾または軽い裾を持つデータで適切に推定できますが、重い裾を持つデータでは保守的な推定になる可能性があります（上記の「1 サンプル標準偏差の方法」内の「理論上の検出力の性能」セクションで要約されているシミュレーション結果を参照）。

結果

データから帰無仮説を棄却するために十分な証拠が得られない場合、アシスタントは正規近似検定の検出力関数を使用して、与えられたサンプルサイズで 80%の確率と 90%の確率で検出できる実質的な差を計算します。さらに、対象となる特定の実質的な差をユーザーが指定した場合、アシスタントは正規近似検定の検出力関数を使用して、その差を検出する確率が 80%と 90%になるサンプルサイズを計算します。

結果を解釈しやすくするため、検出力とサンプルサイズをチェックするときに、1 サンプル標準偏差検定のアシスタントレポートカードに次のステータスインジケータが表示されます。

ステータス	状態
	検定で標準偏差と目標値の差が見つかったため、検出力は問題ではありません。 または 検出力が十分です。検定で標準偏差と目標値の差は検出されませんでした。サンプルは少なくとも 90%の確率で差を検出するのに十分な大きさです。
	検出力が十分な可能性があります。検定で標準偏差と目標値との差は検出されませんでした。サンプルは 80%~90%の確率で差を検出するのに十分な大きさです。90%の検出力を達成するために必要なサンプルサイズがレポートに表示されます。
	検出力が不十分な可能性があります。検定で標準偏差と目標値との差は検出されませんでした。サンプルは 60%~80%の確率で差を検出するのに十分な大きさです。80%の検出力と 90%の検出力を達成するために必要なサンプルサイズがレポートに表示されます。
	検出力が不十分です。検定で標準偏差と目標値との差は検出されませんでした。サンプルは少なくとも 60%の確率で差を検出するのに十分な大きさではありません。80%の検出力と 90%の検出力を達成するために必要なサンプルサイズがレポートに表示されます。

ステータス	状態
	検定で標準偏差と目標値との差が検出されませんでした。検出する実質的な差が指定されていません。データによっては、サンプルサイズと α に基づいて 80%と 90%の確率で検出できる差がレポートに表示される場合があります。

参考文献

- 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.

付録 A: 分散（または標準偏差）の 検定方法

次の表に、分散のさまざまな検定方法の長所と短所を要約します。

方法	コメント
従来のカイ二乗手順	正規性の仮定に大きく影響されます。正規性からの逸脱が小さい場合でも、サンプルの大きさに関係なく結果が不正確になる可能性があります。実際には、データが正規性から逸脱する場合、サンプルサイズを大きくすることで手順の精度が低下します。
サンプル分散の対数変換における漸近正規分布に基づく大サンプル法	一般に従来のカイ二乗法よりも優れていますが、正規性の仮定の影響を受けにくくするには、より大きいサンプルサイズが必要です。
片側（上裾）検定の Edgeworth 展開に基づく大サンプル法 Lee and Ping (1996) を参照。	許容できるタイプ I 過誤率を得ることができますが、対称な分布からデータを取得する必要があります。
尺度化されたカイ二乗分布によるサンプル分散の分布の近似に基づく大サンプル法 この方法は、「AdjDF（自由度調整済み）法」と呼ばれます。 Hummel, Banga, and Hettmansperger (2005) を参照。	サンプル分散の対数変換における漸近正規分布に基づく方法および信頼区間のノンパラメトリック ABC ブートストラップ近似法 (Efron and Tibshirani, 1993) よりも適切な被覆確率を得ることができます。 Minitab 15 の 1 サンプルの分散検定で使用されています。
サンプル分散の対数変換における Bonett の調整済み漸近分布 Bonett (2006) を参照。	やや大きいサンプルでも信頼区間の適切な被覆確率を得ることができます。ただし、裾が重い分布からデータが生成されている場合、はるかに大きいサンプルが必要です。 Minitab 16 の 1 サンプルの分散検定およびアシスタントの 1 サンプル標準偏差検定に使用されています。

付録 B: Bonnett の方法と AdjDF 法の定義

x_1, \dots, x_n を有限の 4 次モーメントを持つ母集団からのサイズ n の観測されたランダムサンプルとします。 \bar{x} と s をそれぞれ観測されたサンプル平均と標準偏差とします。 また、 γ と γ_e をそれぞれ母集団の尖度と過剰尖度とし、 $\gamma_e = \gamma - 3$ とします。 したがって、正規母集団では $\gamma = 3$ および $\gamma_e = 0$ になります。 さらに、 σ^2 を未知の母集団分散とします。 次のセクションで、 σ^2 に関する推測を行う AdjDF (自由度調整済み) 法と Bonnett の方法の 2 つの方法について説明します。

計算式 B1: AdjDF 法

AdjDF 法は、尺度化されたカイ二乗分布によるサンプル分散の分布の近似に基づきます (Box (1953) を参照)。 具体的には、サンプル分散の最初の 2 つのモーメントを尺度化されたカイ二乗分布のモーメントと照合し、未知の尺度と自由度を判断します。 この方法で、次の分散に対する近似両側 $(1 - \alpha)100\%$ 信頼区間が算出されます。

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

ここで

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

この過剰尖度の推定は、Minitab の [基本統計] コマンドで使用される推定と同じです。

計算式 B2: Bonnett の方法

Bonnett の方法はよく知られた従来の方法に依存し、中心極限定理および Cramer の δ の方法を使用して、サンプル分散の対数変換における漸近分布を取得します。 対数変換は、正規分布への収束を加速するために使用されます。 この方法では、分散に対する近似両側 $(1 - \alpha)100\%$ 信頼区間は次のように定義されます。

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

ここで、 z_{α} は標準正規分布の上側百分位数で、 se は次の計算式で表される対数変換済みサンプル分散の標準誤差の漸近推定です。

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

以前、Hummel et al. (2005) で、AdjDF 法はこの従来の方法よりも優れていることを実証するシミュレーションの分析が行われました。 ただし、Bonnett は従来の方法に 2 つの調

整を加えてその制限を克服しています。

1つ目の調整には、尖度の推定が含まれます。尖度を推定するため、Bonett は次の計算式を使用しています。

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

ここで、 m は調整比率が $1/2\sqrt{n-4}$ に等しい調整平均です。この尖度の推定は、裾が重い（対称または歪んでいる）分布の信頼水準の精度を上げる傾向があります。

2つ目の調整では、Bonett の方法はサンプル分散と標準誤差の定数乗数を実証的に判断します。この定数乗数はサンプルが小さい場合に裾の確率をほぼ均一にし、次の計算式で表されます。

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

これらの調整によって、Bonett の分散に対する近似両側 $(1 - \alpha)100\%$ 信頼区間が算出されます。

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

付録 C: Bonett の方法と AdjDF 法の精度

シミュレーション C1: 信頼区間の比較

AdjDF 法および Bonett の方法を使用して計算された分散の信頼区間の精度を比較します。いくつかの分布から異なるサイズ ($n = 20, 30, 40, 50, 60, 80, 100, 150, 200, 250, 300$) のランダムサンプルを生成し、それぞれの方法で信頼区間を計算しました。分布には次が含まれます。

- 標準正規分布 ($N(0, 1)$)
- 一様分布 ($U(0, 1)$)、および両方のパラメータが 3 に設定されたベータ分布 ($B(3, 3)$) を含む、対称的で裾が軽い分布
- 自由度が 5 と 10 の t 分布 ($t(5)$ 、 $t(10)$)、および位置が 0 で尺度が 1 のラプラス分布 ($Lp1$) を含む、対称的で裾が重い分布
- 尺度が 1 の指数分布 (Exp)、および自由度が 3、5、10 のカイ二乗分布 (Chi(3)、Chi(5)、Chi(10)) を含む、歪んでいて裾が重い分布
- パラメータがそれぞれ 8 と 1 に設定された、左方向の歪みがある裾の重いベータ分布 ($B(8, 1)$)

さらに、外れ値の直接的な影響を評価するため、混合正規分布から次のように定義されたサンプルを生成しました。

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

ここで、 p は混合パラメータで、 $1 - p$ は混合比率 (外れ値の比率) です。分析のため、2 つの混合正規母集団を選択しました。1 つ目は母集団の 10% が外れ値である $CN(0.9, 3)$ で、2 つ目は母集団の 20% が外れ値である $CN(0.8, 3)$ です。この 2 つの分布は対称的で、外れ値によって裾が長くなっています。

サンプルサイズごとに各分布から 10,000 サンプル反復が抽出され、それぞれの方法を使用して両側 95% 信頼区間が計算されました。両方の方法が同じサンプルに適用されるように、ランダムサンプルジェネレータがシードされました。これらの信頼区間に基づき、各方法でシミュレートした被覆確率 (CovP) と平均区間幅 (AveW) を計算しました。2 つの方法の信頼区間のシミュレートした被覆確率がほぼ同じ場合、(平均で) より短い区間が得られる方法のほうが正確です。95% の目標信頼水準を使用しているため、シミュレーションの誤差は $\sqrt{0.95(0.05)/10,000} = 0.2\%$ です。

シミュレーション結果を次の表 1 と表 2 に示します。

表1 AdjDF法とBonnettの方法を使用して計算された、分散に対する両側95%信頼区間のシミュレートした被覆確率 これらのサンプルは、軽い裾、正規の裾、ほぼ正規の裾、または重い裾を持つ対称な分布から生成されました。

分布		軽い裾、正規の裾、またはほぼ正規の裾を持つ対称分布				重い裾を持つ対称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(0.8, 3)	CN(0.9, 3)	t(5)
歪度		0	0	0	0	0	0	0	0
尖度(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
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

分布		軽い裾、正規の裾、またはほぼ正規の裾を持つ対称分布				重い裾を持つ対称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(0.8, 3)	CN(0.9, 3)	t(5)
歪度		0	0	0	0	0	0	0	0
尖度(γ_e)		-1.200	-0.667	0	1.000	3.000	4.544	5.333	6.000
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

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

分布		軽い裾、正規の裾、またはほぼ正規の裾を持つ対称分布				重い裾を持つ対称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(0.8, 3)	CN(0.9, 3)	t(5)
歪度		0	0	0	0	0	0	0	0
尖度(γ_e)		-1.200	-0.667	0	1.000	3.000	4.544	5.333	6.000
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

表2 AdjDF法とBonnettの方法を使用して計算された、分散に対する両側95%信頼区間のシミュレートした被覆確率 これらのサンプルは、ほぼ正規の裾、やや重い裾、または重い裾を持つ歪んだ分布から生成されました。

分布	ほぼ正規の裾またはやや重い裾を持つ歪んだ分布			重い裾を持つ歪んだ分布	
	Chi (10)	B(8, 1)	Chi (5)	Chi (3)	Exp
歪度	0.894	-1.423	1.265	1.633	2
尖度(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

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

分布	ほぼ正規の裾またはやや重い裾を持つ歪んだ分布			重い裾を持つ歪んだ分布	
	Chi (10)	B (8, 1)	Chi (5)	Chi (3)	Exp
歪度	0.894	-1.423	1.265	1.633	2
尖度(γ_e)	1.200	2.284	2.400	4.000	6

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	15.25	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	1.230

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

分布		ほぼ正規の裾またはやや重い裾を持つ歪んだ分布			重い裾を持つ歪んだ分布	
		Chi (10)	B (8, 1)	Chi (5)	Chi (3)	Exp
歪度		0.894	-1.423	1.265	1.633	2
尖度(γ_e)		1.200	2.284	2.400	4.000	6
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

この結果は、Bonett (2006) によって発表された内容と一致します。表 1 と表 2 に示されるように、Bonett の方法を使用して計算された信頼区間は、被覆確率が目標水準の 0.95 より近く、信頼区間がより狭いため、AdjDF 法を使用して計算された信頼区間よりも優れています。2つの方法の信頼区間のシミュレートした被覆確率がほぼ同じ場合、(平均で) より短い区間が得られる方法のほうが正確です。つまり、Bonett の方法に基づく分散の統計検定の方が適切に機能し、タイプ I 過誤率とタイプ II 過誤率が低くなります。サンプルサイズが大きい場合はこの 2つの方法の結果はほぼ同じになりますが、サンプルサイズが中程度以下の場合には Bonett の方法のほうが優れています。

Bonett の方法は一般的に AdjDF 法よりも適切に機能しますが、裾が重い (対称または歪んでいる) 分布の場合は、極端に大きいサンプル ($n > 100$) でも常に被覆確率は目標被覆水準の 0.95 よりも低くなります。次の図 1 にこのことが示されています。この図は、小、中

程度、および大のサンプルサイズでの母集団の真の過剰尖度に対する、Bonett の方法でシミュレートした被覆確率をプロットしたものです。

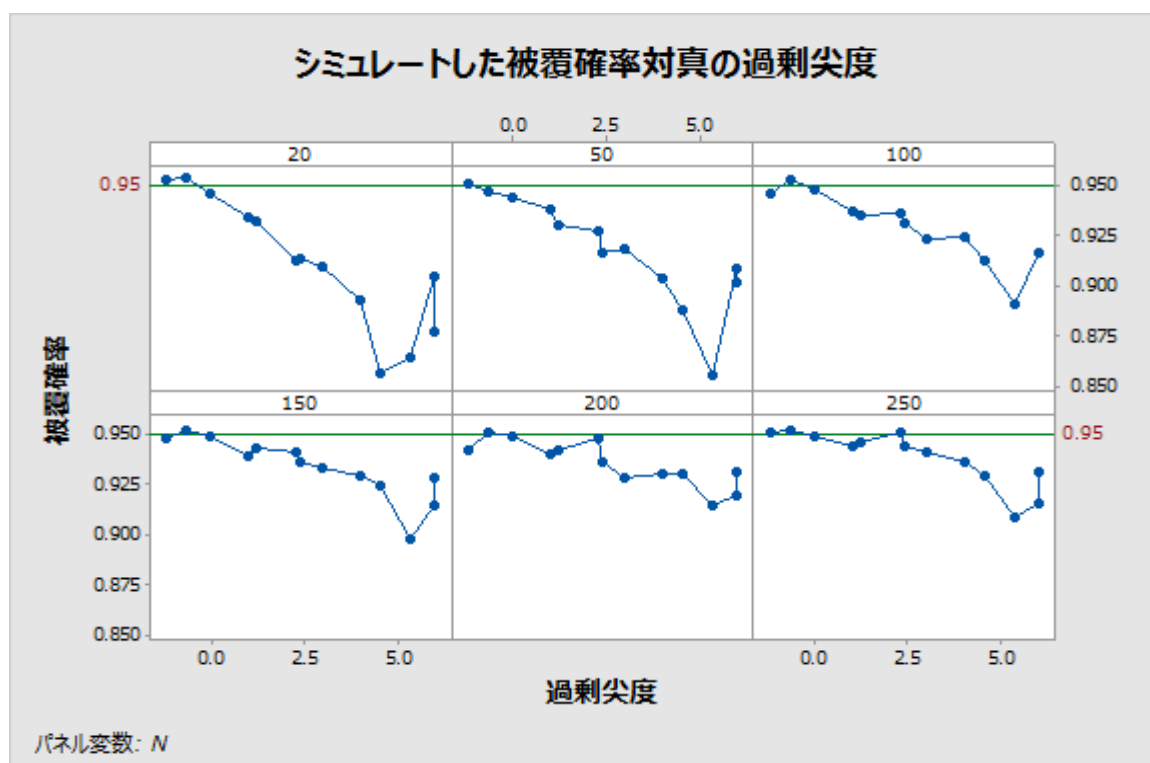


図1 さまざまなサンプルサイズでの各分布の過剰尖度に対する、Bonett の95%信頼区間でシミュレートした被覆確率

図1に示されるように、尖度が高くなるにつれ、シミュレートした被覆確率が目標水準に近づくために必要なサンプルサイズは大きくなります。前述のとおり、Bonett の方法でシミュレートした被覆確率は、裾が重い分布では低くなります。ただし、一様分布やベータ分布 Beta(3, 3) などの裾が軽い分布では、シミュレートした被覆確率は20程度のサンプルサイズで安定して目標水準を達成します。そのため、Bonett の方法の妥当性を判断する基準は、サンプルサイズ、およびサンプルが抽出された分布の裾の重さに基づきます。

この基準を決定する最初のステップとして、裾の重さに従って分布を次の3つのカテゴリに分類します。

- **裾が軽い分布または正規の裾の分布 (Lタイプ)**：これらの分布は、Bonett の信頼区間での被覆確率が安定して目標被覆水準に近くなります。これらの分布では、20程度のサンプルサイズで正確な結果を得ることができます。例として、一様分布、ベータ分布 Beta(3, 3)、正規分布、自由度10のt分布、自由度10のカイ二乗分布などがあります。
- **裾がやや重い分布 (Mタイプ)**：これらの分布の場合、Bonett の方法でシミュレートした被覆確率が目標被覆水準に近くなるには、80以上のサンプルサイズが必要です。例として、自由度が5のカイ二乗分布、ベータ分布 Beta(8, 1) などがあります。
- **裾が重い分布 (Hタイプ)**：これらの分布は、サンプルサイズが極端に大きい ($n \geq 200$) 場合以外は、Bonett の信頼区間での被覆確率は目標被覆水準よりもはるかに

低くなります。例として、自由度が5のt分布、ラプラス分布、自由度が3のカイ二乗分布、指数分布、 $CN(0.9, 3)$ と $CN(0.8, 3)$ の2つの混合正規分布などがあります。

Bonettの方法の妥当性を評価する一般規則では、サンプルデータがこの3つのどの分布タイプから取得されているかを特定する手順を開発する必要があります。検定データチェックの妥当性の一部として、この手順を開発しました。詳細は、「付録 E」を参照してください。

付録 D: 理論上の検出力

Bonett の方法に関連付けられた検定の理論上の検出力関数を導き出し、検定の理論上の検出力とシミュレートした検出力を比較するシミュレーションを実行しました。理論上の検出力曲線とシミュレートした検出力曲線が互いに近い場合、理論上の検出力関数に基づく検出力とサンプルサイズの分析結果は正確になります。

計算式 D1: Bonett の方法の理論上の検出力関数

前述のとおり、Bonett の方法はよく知られた従来の方法に基づき、中心極限定理および Cramer の δ の方法を使用して、対数変換済みサンプル分散の漸近分布を検出します。具体的に言うと、大きいサンプルで $\frac{\ln S^2 - \ln \sigma^2}{se}$ が標準正規分布として近似分布に従うことを立証します。分母の se は、大きいサンプルの対数変換済みサンプル分散の標準誤差で、次の計算式で表されます。

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

ここで、 γ は未知の母集団の尖度です。

Bonett の方法を使用した両側検定の近似 α 水準を含む近似検出力関数は、サンプルサイズ、比率 $\rho = \sigma/\sigma_0$ 、および母集団の尖度 γ の関数として次の計算式で表されます。

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

ここで、 σ_0 は未知の標準偏差の仮説値、 Φ は標準正規分布の CDF、 z_α は標準正規分布の上側 α 百分位点です。片側検出力関数をこれらの計算で取得することもできます。

分析のためサンプルサイズ計画を行う場合、真の尖度の代わりに尖度の推定値が使用される可能性があります。通常、この推定値は専門家の意見または以前の実験の結果に基づきます。この情報が得られない場合は、小規模の試験的調査を行うことで大規模の調査計画を立てることをお勧めします。試験的調査のサンプルを使用することで、尖度を次のように推定できます。

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

ここで、 m は調整比率が $1/2\sqrt{n-4}$ に等しい調整平均です。

シミュレーション D1: 実際の検出力と理論上の検出力の比較

Bonett の方法を使用して分散を検定するときに、推定された実際の検出力水準（シミュレートした検出力水準）と理論上の検出力水準（近似検出力水準）を比較するシミュレーションを計画しました。

この実験では、シミュレーション C1 で説明している各分布から、サイズが n ($n = 20, 30, 40, 50, \dots, 120$) ごとの 10,000 サンプル反復を生成しています（「付録 C」を参照）。各分布とサンプルサイズ n で、 α 水準 $\alpha = 0.05$ が有意である両側検定の 10,000 ランダムサンプル反復の割合として、シミュレートした検出力水準を計算しました。シミュレートした検出力を計算するとき、 $\rho = \sigma/\sigma_0 = 1.25$ を使用して比較的小さい検出力水準を取得しました。次に、比較するために理論上の検出力関数を使用して、対応する検出力水準を計算しました。

この結果を次の表 3 と表 4 に示し、図 2 ではグラフで表しています。

表 3 Bonett の方法に基づく分散の両側検定の ($\rho = \sigma/\sigma_0 = 1.25$ で評価される) シミュレートした検出力水準 (Simul.) と理論上 (正規近似) の検出力水準 (Approx.) の比較。このサンプルは、軽い裾、正規の裾、ほぼ正規の裾、または重い裾を持つ対称分布から生成されました。

n	検出力	軽い裾、正規の裾、またはほぼ正規の裾を持つ対称分布				重い裾を持つ対称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	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
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

n	検出力	軽い裾、正規の裾、またはほぼ正規の裾を持つ対称分布				重い裾を持つ対称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(.8, 3)	CN(.9, 3)	t(5)
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

表4 Bonettの方法に基づく分散の両側検定の ($\rho = \sigma/\sigma_0 = 1.25$ で評価される) シミュレートした検出力水準と理論上(正規近似)の検出力水準の比較。このサンプルは、ほぼ正規の裾、やや重い裾、または重い裾を持つ歪んだ分布から生成されました。

n	検出力	ほぼ正規の裾またはやや重い裾を持つ歪んだ分布			重い裾を持つ歪んだ分布	
		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
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

n	検出力	ほぼ正規の裾またはやや重い裾を持つ歪んだ分布			重い裾を持つ歪んだ分布	
		Chi (10)	B (8, 1)	Chi (5)	Chi (3)	Exp
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

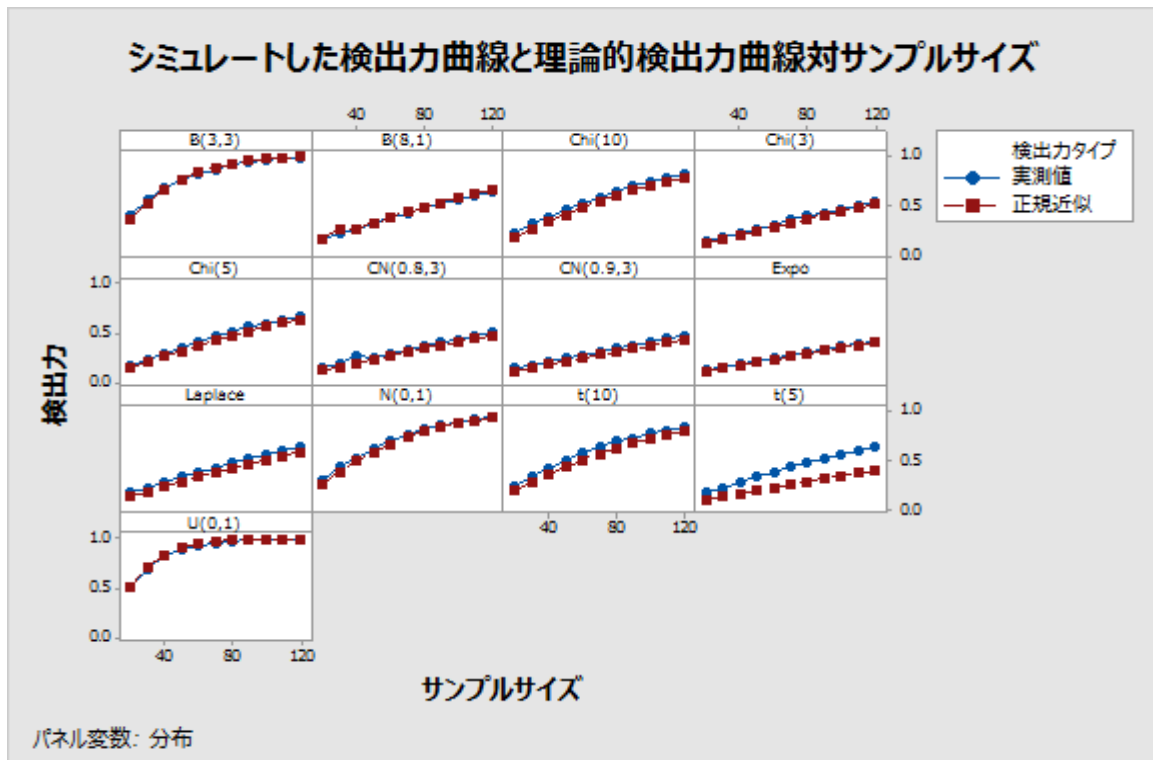


図2 さまざまな分布でのシミュレートした検出力曲線と理論的検出力曲線の比較

表3、表4、および図2の結果は、一様分布、ベータ分布 $Beta(3, 3)$ 、正規分布、自由度10の t 分布、自由度10のカイ二乗分布など、裾が軽い分布（「付録C」で定義されているLタイプの分布）からサンプルが生成されている場合、理論上の検出力の値とシミュレートした検出力水準はほとんど区別できません。

一方、裾が重い分布（Hタイプの分布）では、サンプルが小さい場合、シミュレートした検出力曲線は明らかに理論上の検出力曲線より上になります。これらの裾が重い分布には、自由度5の t 分布、ラプラス分布、自由度3のカイ二乗分布、指数分布、 $CN(0.9, 3)$ と $CN(0.8, 3)$ の2つの混合正規分布などがあります。そのため、裾が重い分布からサンプルが生成され、分析のためサンプルサイズ計画を行う場合、理論上の検出力関数によって推定されたサンプルサイズは、目標検出力を達成するために必要な実際のサンプルサイズよりも大きくなる可能性があります。

付録 E: 正規の裾と重い裾での SJ 検定

「付録 C」のシミュレーションの分析結果では、分布の裾が重くなるほど、Bonett の信頼区間のシミュレートした被覆確率が目標水準に近づくために必要なサンプルサイズは大きくなることが示されています。ただし、シミュレートした被覆確率への歪度による有意な効果は見られませんでした。

そのため、サンプルサイズ、およびサンプルが抽出された分布の裾の重さに基づき、Bonett の方法の妥当性を判断する基準を作成する必要がありました。Gel et al. (2007) は、分布が重い裾を持つという対立仮説に対する、分布が正規の裾を持つという帰無仮説を直接検定するための、適度に検出力の高い検定を提供しています。SJ 検定と呼ばれるこの検定は、次の統計量に基づきます。

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

ここで、 S はサンプル標準偏差、 \hat{j} は中央値 m からのサンプル平均絶対偏差の推定値で、次の計算式で表されます。

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

重い裾の対立仮説に対する近似サイズ α 検定は、次の場合に正規の裾の帰無仮説を棄却します。

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

ここで、 z_α は標準正規分布の上側 α 百分位数で、 $\sigma_R = (\pi - 3)/2$ です。

Gel et al. (2007) では、中程度のサンプルサイズの場合、標準正規分布の上側 α 百分位数を自由度が $(\sqrt{n} + 3)/2$ の t 分布の百分位数に置き換えることで、より適切な近似を得ることができることが示されています。そのため、検定データチェックの妥当性に対して SJ 検定を適用する場合、 z_α を $t_{d,\alpha}$ (自由度が $d = (\sqrt{n} + 3)/2$ の t 分布の上側 α 百分位数) に置き換えます。

付録 F: 検定の妥当性

シミュレーション F1: SJ 検定のシミュレートした検出力を使用した分布の分類の判別

SJ 検定の検出力を調査するシミュレーションを実行し、さまざまな分布からさまざまなサイズ ($n = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200$) のサンプルを生成しました。各分布はシミュレーション C1 と同様に、正規の裾、軽い裾、中程度の裾、または重い裾を持ちます（「付録 C」を参照）。サンプルサイズごとに、各分布から 10,000 サンプル反復が抽出されました。（親分布が正規の裾を持つという）帰無仮説が棄却された場合の比率として、SJ 検定のシミュレートした検出力を計算しました。さらに、平均 R 値 (AveR) と平均 p 値 (AvePV) も計算しました。

シミュレーション結果を次の表 5 と表 6 に示します。

表 5 SJ 検定のシミュレートした検出力水準。このサンプルは、軽い裾、正規の裾、ほぼ正規の裾、または重い裾を持つ対称な分布から生成されました。

分布		軽い裾、正規の裾、またはほぼ正規の裾を持つ対称な分布				重い裾を持つ対称な分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(.8, 3)	CN(.9, 3)	t(5)
n	真の R	0.921	0.965	1.0	1.032	1.128	1.152	1.118	1.085
10	検出力	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	検出力	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	検出力	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	検出力	0.002	0.011	0.065	0.153	0.500	0.550	0.397	0.293
	AveR	0.959	0.995	1.025	1.050	1.128	1.141	1.107	1.086
	AvePV	0.721	0.535	0.391	0.305	0.120	0.128	0.208	0.223
30	検出力	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

分布		軽い裾、正規の裾、またはほぼ正規の裾を持つ対称な分布				重い裾を持つ対称な分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(.8, 3)	CN(.9, 3)	t(5)
n	真の R	0.921	0.965	1.0	1.032	1.128	1.152	1.118	1.085
40	検出力	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	検出力	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	検出力	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	検出力	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	検出力	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	検出力	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	検出力	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	検出力	0.000	0.000	0.052	0.334	0.974	0.974	0.852	0.732
	AveR	0.929	0.971	1.005	1.035	1.128	1.149	1.114	1.083
	AvePV	0.982	0.809	0.452	0.206	0.006	0.007	0.041	0.064
140	検出力	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

分布		軽い裾、正規の裾、またはほぼ正規の裾を持つ対称な分布				重い裾を持つ対称な分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(.8, 3)	CN(.9, 3)	t(5)
<i>n</i>	真の R	0.921	0.965	1.0	1.032	1.128	1.152	1.118	1.085
160	検出力	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	検出力	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	検出力	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

表6 SJ 検定のシミュレートした検出力水準。このサンプルは、ほぼ正規の裾、やや重い裾、または重い裾を持つ歪んだ分布から生成されました。

分布		ほぼ正規の裾またはやや重い裾を持つ歪んだ分布			重い裾を持つ歪んだ分布	
		Chi(10)	B(8, 1)	Chi(5)	Chi(3)	Exp
<i>n</i>	真の R	1.028	1.075	1.059	1.098	1.151
10	検出力	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	検出力	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	検出力	0.152	0.295	0.223	0.328	0.449
	AveR	1.051	1.089	1.070	1.101	1.142
	AvePV	0.335	0.260	0.296	0.242	0.186
25	検出力	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

分布		ほぼ正規の裾またはやや重い裾を持つ歪んだ分布			重い裾を持つ歪んだ分布	
		Chi (10)	B (8, 1)	Chi (5)	Chi (3)	Exp
<i>n</i>	真の R	1. 028	1. 075	1. 059	1. 098	1. 151
30	検出力	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	検出力	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	検出力	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	検出力	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	検出力	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	検出力	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	検出力	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	検出力	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	検出力	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	検出力	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

分布		ほぼ正規の裾またはやや重い裾を持つ歪んだ分布			重い裾を持つ歪んだ分布	
		Chi (10)	B (8, 1)	Chi (5)	Chi (3)	Exp
<i>n</i>	真の R	1.028	1.075	1.059	1.098	1.151
160	検出力	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	検出力	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	検出力	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

表5と表6のシミュレーション結果は、Gel et al. (2007) で発表された内容と一致します。正規母集団からサンプルが抽出された場合、シミュレートした検出力水準（この場合は検定の実際の有意水準を表す）は、サンプルサイズが25の場合でも目標水準とそれほど離れていません。裾が重い分布からサンプルが抽出された場合、検定の検出力は小さいサンプルサイズでは低くなりますが、サンプルサイズが40に達すると40%以上高くなります。具体的に言うと、サンプルサイズ40での検出力は、自由度5のt分布では約40.1%、ラプラス分布では約66.5%、指数分布では約65.1%です。

裾が軽い分布（ベータ分布 Beta(3, 3)）と一様分布では、検定の検出力はサンプルサイズが小さい場合はほぼ0で、サンプルサイズが増えるにつれてさらに低下します。これらの分布に関する証拠は裾が重い分布の対立仮説ではなく裾が軽い分布の対立仮説を支持するため、この結果は驚くべきことではありません。

自由度10のt分布または自由度10のカイ二乗分布など、裾が若干重い分布からサンプルが抽出された場合、中程度から大きいサンプルサイズで検出力水準が低くなります。このシミュレーションの目的では、1サンプルの分散（標準偏差）の検定はこれらの分布で適切に機能し、これらの分布は裾が重いと見なされないことが望ましいため、これは実に良い結果です。ただし、サンプルサイズが大きくなるにつれて検定の検出力が高まるため、裾が若干重いこれらの分布は裾が重い分布として検出されます。

したがって、この検定で分布の裾の重さを評価する規則として、サンプルサイズも考慮に入れる必要があります。これを行う1つの方法は、裾の重さの測度に対する信頼区間を計算することですが、SJ統計量の分布はサンプルの母集団分布に大きく影響されます。別の方法として、SJ検定の帰無仮説を棄却する強度とサンプルサイズの両方に基づき、分布の裾の重さを評価する方法があります。具体的には、小さいp値は裾が重く、大きいp値は裾が軽いことを示します。ただし、大きいサンプルでは小さいサンプルよりもp値が小さくなる傾向にあります。そのため、表3のシミュレートした検出力水準、サンプルサイズ、および平均p値に基づき、SJ検定を使用して各サンプルの分布の裾を評価する規則セットを考案しました。

中程度から大きいサンプルサイズ ($40 \leq n \leq 100$) では、p 値が 0.01~0.05 の場合、帰無仮説を棄却する弱い証拠があると見なします。つまり、サンプルの分布は中程度に裾が重い (M タイプ) 分布に分類されます。一方、p 値が 0.01 未満の場合、帰無仮説を棄却する強力な証拠があると見なし、サンプルの母集団分布を裾が重い (H タイプ) 分布に分類します。

大きいサンプルサイズ ($n > 100$) では、p 値が 0.005~0.01 の場合は親分布を M タイプ分布、p 値が極端に小さい (0.005 未満) 場合は親分布を H タイプ分布に分類します。サンプルサイズが 40 未満の場合、一般的に SJ 検定の検出力はサンプルの分布を効果的に判別するには低すぎると見なされます。

Bonett の方法を使用した 1 サンプルの分散検定の妥当性に対する一般的な分類規則は、次の表 7 に要約されています。

表 7 各サンプルの母集団分布を識別するための分類規則 (p は SJ 検定の p 値)

状態	分布タイプ
$n < 40$	判別不能
$100 \geq n \geq 40$ および $p > 0.05$	L タイプ分布
$n > 100$ および $p > 0.01$	L タイプ分布
$40 \leq n \leq 100$ および $0.01 < p \leq 0.05$	M タイプ分布
$n > 100$ および $0.005 < p \leq 0.01$	M タイプ分布
$40 \leq n \leq 100$ および $p \leq 0.01$	H タイプ分布
$n > 100$ および $p \leq 0.005$	H タイプ分布

先に示したように、シミュレーション C1 の表 1 と表 2 の結果に基づき、L タイプ、M タイプ、H タイプ分布からのサンプルの生成時に、0.93 以上の被覆確率を達成するのに必要なおおよその最小サンプルサイズは、それぞれ 20、80、200 です。ただし、小さいサンプルでの SJ 検定の検出力は低いため、L タイプ分布の最小サンプルサイズ要件は 40 に設定されています。

シミュレーション F2: 分布を分類する規則の確認

シミュレーション C1 で説明されている分布のいくつかからサンプルを生成し、SJ 検定を使用して、L タイプ、M タイプ、H タイプの 3 つの分布グループそれぞれに分類されるサンプルの比率を調査しました。このシミュレーション結果を表 8 に示します。

表8 さまざまな分布から生成された異なるサイズの 10,000 サンプルでの L タイプ、M タイプ、および H タイプとして識別される割合

n	分布	L タイプ				M タイプ	H タイプ	
		B(3, 3)	N(0, 1)	t(10)	Chi(10)	Chi(5)	Lp1	Exp
40	L タイプの%	99.6	94.0	81.5	80.3	66.6	33.0	34.4
	M タイプの%	0.4	5.5	14.0	14.0	20.0	31.9	22.9
	H タイプの%	0.0	0.5	4.5	5.7	13.4	35.1	42.8
50	L タイプの%	99.7	94.4	78.7	79.1	64.0	25.1	28.0
	M タイプの%	0.3	5.1	15.6	14.2	20.0	29.9	20.7
	H タイプの%	0.0	0.5	5.7	6.7	16.0	45.0	51.3
60	L タイプの%	99.7	94.5	77.3	77.3	59.1	18.5	22.6
	M タイプの%	0.3	5.1	16.4	15.0	22.0	27.4	19.2
	H タイプの%	0.0	0.5	6.3	7.7	18.9	54.1	58.2
70	L タイプの%	99.8	94.4	74.5	75.2	55.9	14.0	18.1
	M タイプの%	0.2	5.0	18.1	16.0	22.2	24.0	17.5
	H タイプの%	0.0	0.6	7.4	8.8	21.9	62.0	64.4
80	L タイプの%	99.9	94.3	74.1	74.4	53.0	10.0	13.9
	M タイプの%	0.1	5.1	17.8	16.7	22.8	21.0	15.5
	H タイプの%	0.0	0.6	8.2	8.9	24.2	69.0	70.6
90	L タイプの%	99.9	94.4	71.2	72.1	49.5	7.5	11.1
	M タイプの%	0.1	5.0	19.1	17.2	22.6	16.5	13.7
	H タイプの%	0.0	0.6	9.7	10.7	27.9	76.0	75.3
100	L タイプの%	99.9	94.5	70.8	70.3	47.3	4.8	8.9
	M タイプの%	0.1	4.9	19.5	17.9	22.7	14.3	11.8
	H タイプの%	0.0	0.6	9.7	11.8	30.0	80.9	79.4
120	L タイプの%	100.0	99.4	87.4	87.2	64.8	12.0	14.4
	M タイプの%	0.0	0.4	5.0	4.5	7.9	7.8	5.6
	H タイプの%	0.0	0.2	7.6	8.4	27.4	80.4	80.0
140	L タイプの%	100.0	99.3	86.0	85.1	60.5	7.0	9.9
	M タイプの%	0.0	0.5	5.2	5.0	8.6	5.6	4.1
	H タイプの%	0.0	0.2	8.8	9.9	30.9	87.4	86.0
160	L タイプの%	100.0	99.4	83.4	83.0	55.6	4.0	6.9
	M タイプの%	0.0	0.5	6.3	5.8	9.5	3.5	3.0
	H タイプの%	0.0	0.1	10.4	11.2	34.9	92.5	90.1

n	分布	L タイプ				M タイプ	H タイプ	
		B(3, 3)	N(0, 1)	t(10)	Chi(10)	Chi(5)	Lp1	Exp
180	L タイプの%	100.0	99.3	81.1	81.7	51.0	2.5	4.6
	M タイプの%	0.0	0.5	6.8	5.9	9.4	1.9	2.2
	H タイプの%	0.0	0.2	12.1	12.4	39.6	95.6	93.2
200	L タイプの%	100.0	99.5	79.0	80.5	47.2	1.3	3.0
	M タイプの%	0.0	0.4	7.6	6.1	9.4	1.6	1.7
	H タイプの%	0.0	0.1	13.4	13.4	43.4	97.1	95.3

表 8 の結果では、裾が軽い (L タイプ) 分布と裾が重い (H タイプ) 分布からサンプルが生成された場合、サンプルは高い比率で正しく分類されることが示されています。たとえば、サイズ 40 のサンプルがベータ分布 Beta(3, 3) から生成された場合、99.6% のサンプルが軽い裾を持つと正しく分類され、サイズ 90 のサンプルがラプラス分布から生成された場合、76.0% が重い裾を持つと正しく分類されています。その結果、真に裾が軽い分布からサンプルが生成された場合、検定の妥当性に関するレポートカードの警告メッセージが誤って発行されることはなく、裾が重い分布からサンプルが生成されて最小サンプルサイズ要件が満たされていない場合は、警告メッセージが正しく発行されます。さらに、カイ二乗分布 Chi(5) などの裾がやや重い (M タイプ) 分布からサンプルが生成され、サンプルが小さい場合、サンプルは高い比率で裾が軽い (L タイプ) と誤って分類されます (サンプルサイズ 40 では、66% のサンプルが裾が軽い分布と誤って分類されています)。したがって、このような場合は親分布がやや重い裾を持っていても、レポートカードで警告メッセージは発行されません。ただし、サンプルサイズが 80 より大きい場合、最小サンプルサイズ要件はすでに満たされているため、L タイプ分布として誤って分類されることによる影響はありません。

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