

標準偏差の検定 (2 つ以上のサンプル)

概要

Minitab アシスタントには、独立したサンプルを比較して、その変動性が有意に異なるかどうかを判断するための、2 つの分析が含まれています。2 サンプル標準偏差検定では 2 つのサンプルの標準偏差を比較し、標準偏差検定では 3 つ以上のサンプルの標準偏差を比較します。本書では、 $k = 2$ の k サンプル計画は 2 サンプル計画、 $k > 2$ の k サンプル計画は多重サンプル計画を指します。一般的に、これら 2 種類の計画は個別に分析されます（「付録 A」を参照）。

標準偏差は分散の平方根であるため、標準偏差を比較する仮説検定は、分散を比較する仮説検定と同等です。多くの統計手法は、2 つ以上の母集団の分散を比較するために開発されました。これらの検定のうち、最も頑健で一般的に使用されるのが、Levene/Brown-Forsythe 検定です。ただし、Levene/Brown-Forsythe 検定の検出力の性能は、2 サンプル計画のタイプ I の誤りの特性より劣ります。Pan (1999) は、正規母集団を含む一部の母集団では、2 サンプル計画の検定の検出力の上限が、標準偏差間の差の大きさに関係なく、1 よりもかなり下になる場合があることを示しています。つまり、これらのタイプのデータでは、標準偏差間の差の大きさに関係なく、検定では差がないと結論付けられる可能性が高いということです。このことから、アシスタントでは、2 サンプル標準偏差検定に新しい Bonett 検定を使用します。多重サンプル計画を使用する標準偏差検定には、多重比較 (MC) 手順を使用します。

2 サンプルの等分散性を調べる Layard (1978) の検定の修正版である Bonett (2006) 検定は、小さなサンプルを使用して検定の性能を向上させます。Banga and Fox (2013A) は、Bonett の検定に関連付けられた信頼区間を導出し、それが Levene/Brown-Forsythe 検定に関連付けられた信頼区間と同程度に正確であり、ほとんどの分布では精度がより高いことを示しています。さらに、Banga and Fox (2013A) は、Bonett 検定が Levene/Brown-Forsythe 検定と同程度に頑健であり、ほとんどの分布ではより強力であると判断しました。

多重比較 (MC) 手順には、複数のサンプルの標準偏差 (または分散) の等質性 (同等性) の全体検定が含まれています。これは、標準偏差の各ペアの比較区間に基づきます。比較区間は、少なくとも 1 つのペアの信頼区間が重なり合っていない場合のみ、MC 検定が有意に

なるように導出されます。Banga and Fox (2013B) は、MC 検定のタイプ I およびタイプ II の誤りの特性は、ほとんどの分布で Levene/Brown-Forsythe 検定と類似していることを示しています。MC 検定の重要な利点の 1 つは、比較区間のグラフ表示です。これは、異なる標準偏差を持つサンプルを特定するための効果的な視覚ツールとなります。計画のサンプルが 2 つのみの場合、MC 検定は Bonett 検定と同等です。

本書では、異なるデータ分布およびサンプルサイズで Bonett 検定および MC 検定の妥当性を評価します。さらに、大サンプル近似法に基づく Bonett 検定で使用される検出力とサンプルサイズ分析について調査します。これらの因子に基づき、アシスタントでデータに対して自動的に実行され、レポートカードに表示される、次のチェックを作成しました。

- 異常なデータ
- 正規性
- 検定の妥当性
- サンプルサイズ (2 サンプル標準偏差検定のみ)

標準偏差の方法の検定

Bonett の検定および MC 検定の妥当性

等分散の検定の比較研究において、Conover et al. (1981) は、タイプ I およびタイプ II の過誤率に基づき、Levene/Brown-Forsythe 検定が最も優れた検定の 1 つであることを発見しました。それ以降、2 サンプルおよび多重サンプル計画の等分散の検定に、他の方法が提案されてきました (Pan 1999、Shoemaker 2003、Bonett 2006)。たとえば、Pan はその頑健さと解釈の単純さにもかかわらず、Levene/Brown-Forsythe 検定には、サンプルが正規母集団を含む一部の母集団からのものである場合、2 つの標準偏差間の重要な差を検出するのに十分な検出力がないことを示しています。この重要な制限があるため、アシスタントでは 2 サンプル標準偏差検定に Bonett 検定を使用します (「付録 A」または Banga and Fox 2013A を参照)。3 つ以上のサンプルを使用する標準偏差検定では、比較区間を使用する MC 手順が使用されます。この手順では、グラフが表示され、MC 検定が有意な場合に異なる標準偏差を持つサンプルを特定できます (「付録 A」または Banga and Fox 2013B を参照)。

目的

最初に、2 つの母集団の標準偏差を比較する場合の、Bonett 検定の性能を評価する必要性がありました。次に、3 つ以上の母集団の標準偏差を比較する場合の、MC 検定の性能を比較する必要があります。具体的に言えば、異なるタイプの分布から得たさまざまなサイズのサンプルで、これらの検定が実行される場合の妥当性を評価する必要性がありました。

方法

Bonett 検定および MC 検定に使用される統計手法は、「付録 A」で定義しています。検定の妥当性を評価するために、異なる条件下でタイプ I 過誤率が目標有意水準 (α) に近いかどうかを調べる必要性がありました。そのために、2 つの独立したサンプルの標準偏差を比較する場合の Bonett 検定の妥当性を評価するシミュレーションセットを実行しました。また、複数 (k 個) の独立したサンプル ($k > 2$) の標準偏差を比較する場合の MC 検定の妥当性を評価する別のシミュレーションセットを実行しました。

バランス型計画とアンバランス型計画の両方を使用して、複数の分布からさまざまなサイズの 10,000 ペアまたは複数 (k 個) のランダムサンプルを生成しました。次に、2 つのサンプルの標準偏差を比較するために両側 Bonett 検定を実行しました。また、目標有意水準を使用して、各実験の k サンプルの標準偏差を比較するために MC 検定を実行しました。反復数 10,000 のうち、検定で帰無仮説が棄却された回数を数え (実際に真の標準偏差が等しかった場合)、シミュレートした有意水準と呼ばれるこの比率と目標有意水準を比較しました。検定が正しく行われると、実際のタイプ I 過誤率を表すシミュレートした有意水準が、目標有意水準にかなり近くなります。2 サンプルおよび k サンプルのシミュレーションで使用された特定の方法についての詳細は、「付録 B」を参照してください。

結果

2 サンプル比較の場合、サンプルが中規模または大規模のとき、Bonett 検定のシミュレートしたタイプ I 過誤率は、分布、および計画がバランス型かアンバランス型かに関係なく、目標有意水準に近くなりました。ただし、極端に偏った母集団から小さなサンプルが抽出され

た場合、Bonett 検定は一般に保守的で、タイプ I 過誤率は目標有意水準（すなわち目標タイプ I 過誤率）より少し低くなりました。

多重サンプル比較の場合、サンプルが中規模または大規模のとき、MC 検定のシミュレートしたタイプ I 過誤率は、分布、および計画がバランス型かアンバランス型かに関係なく、目標有意水準に近くなりました。ただし、小さくて極端に偏ったサンプルの場合、検定は概して比較的保守的で、計画のサンプル数が多いと、タイプ I 過誤率は目標有意水準より高くなりました。

分析の結果は、Banga and Fox (2013A、2013B) の結果と一致しました。Bonett 検定および MC 検定は、最小サンプルのサイズが 20 以上の場合に正しく行われるという結論に達しました。したがって、アシスタントレポートカードの検定チェックの妥当性には、この最小サンプルのサイズの要件を使用します。

比較区間

2 つ以上の標準偏差を比較する検定が統計的に有意であり、少なくとも 1 つの標準偏差が別の標準偏差と異なることを示す場合、分析の次のステップは、どのサンプルが統計的に異なるのかを判断することです。この比較を行う直観的な方法は、各サンプルに関連付けられた信頼区間をグラフ化し、区間が重なり合っていないサンプルを特定することです。ただし、個別信頼区間は比較するためのものではないため、グラフから引き出す結論が検定結果と一致しない場合があります。

目的

等分散性の全体検定として、また全体検定が有意の場合に異なる分散を持つサンプルを特定するための方法として使用できる個別比較区間を計算する方法を開発する必要性がありました。MC 手順の重要な要件は、少なくとも 1 つのペアの信頼区間が重なり合っていない場合のみ、全体検定が有意になるということです。これは、少なくとも 2 つのサンプルの標準偏差が異なることを示します。

方法

複数の標準偏差を比較するために使用する MC 手順は、複数のペアワイズ比較から導出されます。サンプルの各ペアは、2 つの母集団標準偏差の同等性の Bonett (2006) 検定を使用して比較されます。ペアワイズ比較では、Nakayama (2009) に示す大サンプルの近似に基づいて多重性補正が使用されます。Bonferroni 補正はサンプル数が増大するにつれてより保守的になるため、一般的な Bonferroni 補正よりも大サンプルの近似の使用が好まれます。最後に、比較区間は、Hochberg et al. (1982) の最良の近似手順に基づくペアワイズ比較に由来します。詳細は、「付録 A」を参照してください。

結果

MC 手順は、少なくとも 2 つの比較区間が重なり合っていない場合のみ、標準偏差の同等性の全体検定が有意になるという要件を満たしています。逆に言えば、全体検定が有意な場合は、少なくとも 2 つの比較区間が重なり合っているということです。

比較区間は要約レポートの標準偏差比較管理チャートに表示されます。このグラフの横には、標準偏差の等質性の全体検定である MC 検定の p 値が表示されます。標準偏差検定が統計的に有意である場合、少なくとも他の 1 つの区間と重なり合っていない区間はすべて赤でマー

クされます。標準偏差検定が統計的に有意でない場合、赤でマークされる区間は1つもありません。

理論上の検出力の性能（2 サンプル計画のみ）

サンプルサイズを計画するには、Bonett 検定および MC 検定の理論上の検出力関数が必要です。2 サンプル計画の場合、検定の近似の理論上の検出力関数は、大サンプル理論の方法を使用して導出できます。この関数は大サンプル近似法に由来するため、正規分布および非正規分布から生成された小さいサンプルを使用して検定が行われる場合に、その特性を評価する必要があります。ただし、3 つ以上のグループの標準偏差を比較する場合、MC 検定の理論上の検出力関数を簡単に得ることはできません。

目的

アシスタントの 2 サンプル標準偏差検定の検出力およびサンプルサイズの要件を評価するために、大サンプルの近似に基づく理論上の検出力関数を使用できるかどうかを判断する必要があります。そのために、正規分布および非正規分布を含む複数の種類の分布で Bonett 検定が実行される場合に実現される実際の検出力が、近似された理論上の検出力関数で正確に反映されるかどうかを評価する必要があります。

方法

2 サンプル計画に対する Bonett 検定の近似された理論上の検出力関数は「付録 C」で得られます。

Bonett 検定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と呼ぶ）を推定するためにシミュレーションを行いました。最初に、正規分布および非正規分布を含む複数の分布から、さまざまなサイズのランダムサンプルのペアを生成しました。各分布で、サンプルペアの反復数 10,000 に Bonett 検定を実行しました。所与の差を、サンプルペア 10,000 個のうち検定が有意である割合として検出するために、検定のシミュレートした検出力を計算しました。比較のために、検定の近似された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。近似が有効に機能する場合、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

Bonett 検定の理論上の検出力関数とシミュレートした検出力関数は、ほとんどの分布において、小さいサンプルサイズの場合はほぼ等しく、最小サンプルサイズが 20 に達するとより近くなるのが、シミュレーションで示されました。軽い裾から適度の裾を持つ対称分布およびほぼ対称な分布では、理論上の検出力水準はシミュレートした（実際の）検出力水準より若干高くなります。ただし、歪んだ分布や裾の重い分布では、シミュレートした（実際の）検出力水準よりも小さくなります。詳細は、「付録 D」を参照してください。

この結果から、全般的に見て、理論上の検出力水準はサンプルサイズを計画するうえで良好な基準となることが示されました。

データチェック

異常なデータ

異常なデータとは、極端に大きいまたは小さいデータ値を指し、外れ値とも呼ばれています。異常なデータは、分析の結果に強い影響を与える可能性があり、特にサンプルが小さい場合には、統計的に有意な結果を検出する確率に影響することがあります。異常なデータは、データ収集での問題を示している可能性や、分析している工程の異常な動作に起因する場合があります。したがって、これらのデータ点は調査する価値があり、可能な場合には修正する必要があります。シミュレーションの分析によると、データに外れ値が含まれている場合、Bonett 検定および MC 検定は保守的になることがわかっています（「付録 B」を参照）。検定の実際の有意水準は、特に小さいサンプルで分析が行われる場合、目標水準よりも著しく小さくなります。

目的

分析の結果に影響する可能性がある、全体のサンプルに比べて非常に大きい、または非常に小さいデータ値をチェックする方法を開発する必要がありました。



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。四分位間範囲とは、2 つの四分位数間の差を指します。この方法は、特定の各外れ値を検出することが可能なため、複数の外れ値がある場合でも有効に機能します。

異常なデータを調べる場合、レポートカードには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。
	少なくとも 1 つのデータ点が異常です。結果に強い影響を与える可能性があります。

正規性

正規性の仮定のもとに導出される、等分散性のほとんどの検定とは違い、標準偏差の同等性を調べる Bonett 検定および MC 検定では、データの特定の分布について仮定しません。

目的

Bonett 検定および MC 検定は、大サンプル近似法に基づいていますが、小さいサンプルの正規および非正規データでも正しく行われることを確認する必要がありました。また、データの正規性が標準偏差検定の結果にどのように関係するのかをユーザーに説明する必要がありました。

方法


異なる条件下で検定の妥当性を評価するために、さまざまなサンプルサイズの正規および非正規データを使用した Bonett 検定および MC 検定のタイプ I 過誤率を調べるシミュレーションを行いました。詳細は、「標準偏差の方法の検定」セクションおよび「付録 B」を参照してください。

結果


十分に大きなサンプル（最小サンプルサイズ ≥ 20 ）の Bonett 検定または MC 検定では、データの分布がタイプ I の誤りの特性に大きな影響を与えないことが、シミュレーションで示されました。検定では、正規データと非正規データの両方で、タイプ I 過誤率は一貫して目標過誤率に近くなりました。

タイプ I 過誤率に関するこれらの結果に基づき、アシスタントではレポートカードに正規性についての情報が表示されます。

2 サンプル計画の場合は、次のインジケータが表示されます。

ステータス	状態
	この分析には、Bonett 検定が使用されます。サンプルが十分に大きいと、正規データと非正規データの検定は両方とも正しく行われます。

多重サンプル計画の場合は、次のインジケータが表示されます。

ステータス	状態
	この分析には、多重比較検定が使用されます。サンプルが十分に大きいと、正規データと非正規データの検定は両方とも正しく行われます。

検定の妥当性

「標準偏差の方法の検定」セクションで、2 サンプル比較と多重 (k) 比較の両方において、サンプルサイズが中程度または大きいとき、Bonett 検定および MC 検定のタイプ I 過誤率は、バランス型およびアンバランス型計画の正規および非正規データの目標過誤率に近くなることを示しました。ただし、サンプルが小さいと、一般的に Bonett 検定および MC 検定は正しく行われません。

目的



ユーザーのデータに基づき、2 サンプルと多重 (k) サンプルの標準偏差検定結果の妥当性を評価するために、ルールを適用する必要がありました。

方法

異なる異なる条件下で検定の妥当性を評価するために、「標準偏差の方法の検定」セクションで前述したように、さまざまなデータ分布、サンプル数、サンプルサイズを使用した Bonett 検定および MC 検定のタイプ I 過誤率を調べるシミュレーションを行いました。詳細は、「付録 B」を参照してください。

結果

Bonett 検定および MC 検定は、最小サンプルのサイズが 20 以上の場合に正しく行われます。したがって、標準偏差検定の妥当性を評価するために、レポートカードには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
	サンプルサイズが 20 以上なので p 値は正確です。
	一部のサンプルサイズが 20 未満なので p 値は正確でない可能性があります。サンプルサイズを少なくとも 20 に増やすことを検討してください。

サンプルサイズ (2 サンプル標準偏差検定のみ)

一般的に、統計的仮説検定は、「差なし」の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と、実際に存在する差を検出するのに検定の検出力が十分でない場合あり、その結果、タイプ II の誤りが生じます。したがって、サンプルサイズが実質的に重要な差を検出するのに十分な大きさであることを確認することが重要です。

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、検定のサンプルサイズが、高い確率で対象となる実質的な差を検出するのに十分な大きさかどうかを判断する必要があります。サンプルサイズの計画の目的は、サンプルサイズが、重要な差を高い確率で検出するのに十分な大きさであることを確認することですが、大きすぎて無意味な差が高い確率で統計的に有意になるようではいけません。






方法

2 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、Bonett 検定の検出力関数の近似に基づいており、通常は検定の実際の検出力関数の適切な推定が得られます（「方法」セクションの理論上の検出力関数の性能に要約されたシミュレーション結果を参照）。

結果

データから帰無仮説に反する十分な証拠が得られない場合、Bonett 検定の近似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%および 90%の確率で検出できる実質的な差が計算されます。さらに、ユーザーが対象の特定の实質的な差を指定すると、正規近似検定の検出力関数を使用して、80%および 90%の確率で差を検出できるサンプルサイズが計算されます。

結果の解釈に役立つよう、検出力とサンプルサイズをチェックすると、2 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は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
	<p>検定で標準偏差間の差が検出されます。検出力に問題はありません。</p> <p>または</p> <p>検出力は十分です。検定で標準偏差間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。</p>
	<p>検出力は十分な可能性があります。検定で標準偏差間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のに必要なサンプルサイズが報告されます。</p>
	<p>検出力は十分でない可能性があります。検定で標準偏差間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%および 90%の検出力を達成するのに必要なサンプルサイズが報告されます。</p>
	<p>検出力は十分ではありません。検定で標準偏差間の差は検出されず、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%および 90%の検出力を達成するのに必要なサンプルサイズが報告されます。</p>
	<p>検定で標準偏差間の差が検出されませんでした。ユーザーが検出する実質的な差を指定しなかったため、サンプルサイズと α に基づいて 80%および 90%の確率で検出できる差がレポートに示されます。</p>

参考文献

- Arnold, S.F. (1990). *Mathematical statistics*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Banga, S.J. and Fox, G.D. (2013A). On Bonett's Robust Confidence Interval for a Ratio of Standard Deviations. *White paper, Minitab Inc.*
- Banga, S.J. and Fox, G.D. (2013B) A graphical multiple comparison procedure for several standard deviations. *White paper, Minitab Inc.*
- Bonett, D.G. (2006). Robust confidence interval for a ratio of standard deviations. *Applied Psychological Measurements*, 30, 432-439.
- Brown, M.B., & Forsythe, A.B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69, 364-367.
- Conover, W.J., Johnson, M.E., & Johnson, M.M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, 23, 351-361. r
- Gastwirth, J. L.L. (1982). Statistical properties of a measure of tax assessment uniformity. *Journal of Statistical Planning and Inference*, 6, 1-12.
- Hochberg, Y., Weiss G., and Hart, S. (1982). On graphical procedures for multiple comparisons. *Journal of the American Statistical Association*, 77, 767-772.
- Layard, M.W.J. (1973). Robust large-sample tests for homogeneity of variances. *Journal of the American Statistical Association*, 68, 195-198.
- Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Probability and statistics* (278-292). Stanford University Press, Palo Alto, California.
- Nakayama, M.K. (2009). Asymptotically valid single-stage multiple-comparison procedures. *Journal of Statistical Planning and Inference*, 139, 1348-1356.
- Pan, G. (1999) On a Levene type test for equality of two variances. *Journal of Statistical Computation and Simulation*, 63, 59-71.
- Shoemaker, L. H. (2003). Fixing the F test for equal variances. *The American Statistician*, 57 (2), 105-114.

付録 A: Bonnett 検定および多重比較検定の方法

Bonnett 検定 (2 サンプル計画) または多重比較 (MC) 手順 (多重サンプル計画) を使用して標準偏差または分散について推測する場合の前提は、次のように説明できます。

$X_{11}, \dots, X_{1n_1}, \dots, X_{k1}, \dots, X_{kn_k}$ を k ($k \geq 2$) 個の独立したランダムサンプルとし、各サンプルは、 $i = 1, \dots, k$ について、それぞれ未知の平均 μ_i と分散 σ_i^2 を持つ分布から抽出されます。サンプルの親分布には、共通の有限尖度 $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$ があると仮定します。この仮定は理論的導出には極めて重要ですが、サンプルが十分に大きい最も実用的な応用では重要ではありません (Banga と Fox 2013A)。

方法 A1: 2 サンプルの等分散性の Bonnett 検定

Bonnett 検定は、2 サンプルの分散または標準偏差を比較する 2 サンプル計画にのみ適用されます。この検定は、2 サンプル計画の等分散性を調べる Layard (1978) 検定の修正版です。有意水準 α を持つ 2 サンプルの等分散性を調べる両側 Bonnett 検定では、次の場合にのみ、同等性の帰無仮説が棄却されます。

$$|\ln(c S_1^2 / S_2^2)| > z_{\alpha/2} \sqrt{\frac{\hat{\gamma}_P - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P - g_2}{n_2 - 1}}$$

ここで、

S_i はサンプル i のサンプル標準偏差

$$g_i = (n_i - 3) / n_i, i = 1, 2$$

$z_{\alpha/2}$ は、標準正規分布の上位 $\alpha/2$ の百分位数を指す

$\hat{\gamma}_P$ は次のように与えられる併合尖度推定値

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

併合尖度推定の式では、 m_i はサンプル i の調整平均、調整比率は $1/[2(n_i - 4)^{1/2}]$ です。

上の式には、アンバランス型計画で等しくない裾部による誤差の確率の影響を削減するために、定数 c が小さなサンプル調整として含まれています。この定数は $c = c_1 / c_2$ として与えられます。

$$c_i = \frac{n_i}{n_i - z_{\alpha/2}}, i = 1, 2$$

計画がバランス型の場合 (つまり $n_1 = n_2$ のとき)、検定の p 値は次のように求めます。

$$P = 2 \Pr(Z > z)$$

ここで、 Z は標準正規分布に従う確率変数、 z は手元のデータに基づく次の統計量の観測値です。統計量は次のようになります。

$$Z = \frac{\ln(C S_1^2/S_2^2)}{se}$$

ここで

$$se = \sqrt{\frac{\hat{\gamma}_P - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P - g_2}{n_2 - 1}}$$

一方、計画がアンバランス型の場合、検定の p 値は次のように求めます。

$$P = 2\min(\alpha_L, \alpha_U)$$

ここで、 $\alpha_L = \Pr(Z > z_L)$ および $\alpha_U = \Pr(Z > z_U)$ となります。変数 z_L は、関数の最小根

$$L(z, S_1, S_2, n_1, n_2) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} - z se + \ln \frac{S_1^2}{S_2^2} - \ln \rho_0^2, z < \min(n_1, n_2)$$

および z_U は関数 $L(z, S_2, S_1, n_2, n_1)$ の最小根です。

方法 A2: 多重比較検定と比較区間

k ($k \geq 2$) 個の独立したグループまたはサンプルがあるとします。目的は、 k 区間のうち少なくとも 2 つの区間が重なり合っていない場合にのみ、標準偏差の同等性の検定が有意になるように母集団の標準偏差に対し k 区間のシステムを構築することでした。これらの区間は、比較区間と呼ばれます。この比較の方法は、Tukey-Kramer によって最初に開発され、後に Hochberg et al. (1982) で一般化された一元配置の分散分析モデルの平均の多重比較手順と似ています。

2 つの標準偏差の比較

2 サンプル計画の場合、Bonett 検定に関連付けられた標準偏差の比の信頼区間を直接計算し、標準偏差間の差のサイズを評価できます (Banga and Fox 2013A)。この手法は実際に、Minitab リリース 17 の [統計] > [基本統計] > [2 サンプルの分散] で使用されます。ただし、アシスタントでは、標準偏差の比の信頼区間より解釈しやすい比較区間を提供する必要がありました。そのために、「方法 A1」で説明されている Bonett の手順を使用して、2 つのサンプルの信頼区間を判断しました。

2 つのサンプルがあるとき、等分散性の Bonett 検定は、等分散性の Bonett 検定に関連付けられた次の許容区間に 0 が含まれていない場合にのみ、有意になります。

$$\ln(c_1 S_1^2) - \ln(c_2 S_2^2) \pm z_{\alpha/2} \sqrt{\frac{\hat{\gamma}_P - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P - g_2}{n_2 - 1}}$$

ここで、併合尖度推定値 $\hat{\gamma}_P$ と $g_i, i = 1, 2$ がそれぞれ与えられます。

この区間から、次の 2 つの信頼区間が重なり合っていない場合にのみ、分散または標準偏差の同等性の検定が有意であると推定します。これらの 2 つの信頼区間は、次のようになります。

$$\left[S_i \sqrt{C_i \exp(-z_{\alpha/2} V_i)}, S_i \sqrt{C_i \exp(z_{\alpha/2} V_i)} \right], i = 1, 2$$

ここで

$$V_i = \frac{\sqrt{\frac{\hat{Y}_P - g_i}{n_i - 1}}}{\sqrt{\frac{\hat{Y}_P - g_i}{n_i - 1} + \frac{\hat{Y}_P - g_j}{n_j - 1}}} \sqrt{\frac{\hat{Y}_P - g_i}{n_i - 1} + \frac{\hat{Y}_P - g_j}{n_j - 1}}, i = 1, 2; j = 1, 2; i \neq j$$

標準偏差の同等性の検定手順としてこれらの区間を使用することは、標準偏差の同等性の Bonnett 検定に相当します。具体的に言えば、標準偏差の同等性の Bonnett 検定が有意である場合にのみ、信頼区間は重なり合いません。ただし、これらの区間は標準偏差の信頼区間ではなく、標準偏差の多重比較にのみ適しています。同じ理由から、Hochberg et al. では平均を比較するための類似の区間を、不確実性区間と呼んでいます。Minitab では、これらの区間を比較区間と呼びます。

比較区間手順は標準偏差の同等性の Bonnett 検定に相当するため、比較区間に関連付けられた p 値は、前述した 2 サンプルの等分散性の Bonnett 検定の p 値と同一になります。

複数の標準偏差の比較

3 つ以上のグループまたはサンプルがある場合、 k 比較区間は、標準偏差の同等性の $k(k - 1)/2$ ペアワイズ同時検定で、全体有意水準 α を使用して推定されます。より具体的には、 X_{i1}, \dots, X_{in_i} と X_{j1}, \dots, X_{jn_j} を任意のサンプルペア (i, j) のサンプルデータとします。2 サンプルの場合と同様に、特定のサンプルペア (i, j) の標準偏差の同等性の検定は、次の区間に 0 が含まれていない場合にのみ、ある α' 水準で有意になります。

$$\ln(c_i S_i^2) - \ln(c_j S_j^2) \pm z_{\alpha' / 2} \sqrt{\frac{\hat{Y}_{ij} - g_i}{n_i - 1} + \frac{\hat{Y}_{ij} - g_j}{n_j - 1}}$$

上の式の \hat{Y}_{ij} は、サンプルペア (i, j) に基づく併合尖度推定値で、次のように与えられます。

$$\hat{Y}_{ij} = (n_i + n_j) \frac{\sum_{l=1}^{n_i} (X_{il} - m_i)^4 + \sum_{l=1}^{n_j} (X_{jl} - m_j)^4}{[(n_i - 1)S_i^2 + (n_j - 1)S_j^2]^2}$$

さらに、前述したように、 m_i は調整比率 $1/[2(n_i - 4)^{1/2}]$ と以下を使用した、サンプル i の調整平均です。

$$g_i = \frac{n_i - 3}{n_i}, g_j = \frac{n_j - 3}{n_j}, c_i = \frac{n_i}{n_i - z_{\alpha' / 2}}, c_j = \frac{n_j}{n_j - z_{\alpha' / 2}}$$

$k(k - 1)/2$ の同時ペアワイズ検定があるため、水準 α' は実際の全体過誤率が目標有意水準 α に近くなるように選択する必要があります。可能な調整の 1 つは、Bonferroni の近似に基づきます。ただし、Bonferroni 補正は計画のサンプル数が増えると、より保守的になることで知られています。より優れた手法は、Nakayama (2008) によって与えられた正規近似に基づきます。この手法を使用して、単に $z_{\alpha' / 2}$ を $q_{\alpha, k} / \sqrt{2}$ に置き換えます。ここで、 $q_{\alpha, k}$ は k 個の互いに独立で同じ分布に従う標準正規確率変数の範囲の上限 α 点です。つまり、

$$\Pr \left(\max_{1 \leq i < j \leq k} |Z_i - Z_j| \leq q_{\alpha, k} \right) = 1 - \alpha$$

ここで、 Z_1, \dots, Z_k は互いに独立で同じ分布に従う標準正規確率変数です。

さらに、Hochberg et al. (1982) と類似の方法を使用し、前述したペアワイズ手順を最良に近似する手順で、あるサンプルペア (i, j) について、標準偏差の同等性の帰無仮説が棄却されます。

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > q_{\alpha, k} (V_i + V_j) / \sqrt{2}$$

ここで、 V_i は数量

$$\sum_{i \neq j} \sum (V_i + V_j - b_{ij})^2$$

を次で最小化するために選択されます。

$$b_{ij} = \sqrt{\frac{\hat{Y}_{ij} - g_i}{n_i - 1} + \frac{\hat{Y}_{ij} - g_j}{n_j - 1}}$$

Hochberg et al. (1982) で示されたこの問題の解は、次を選択することです。

$$V_i = \frac{(k-1) \sum_{j \neq i} b_{ij} - \sum_{1 \leq j < l \leq k} b_{jl}}{(k-1)(k-2)}$$

その結果、次の k 区間の少なくとも 1 つのペアが重なり合っていない場合にのみ、近似手順に基づく検定が有意であるということになります。

$$\left[S_i \sqrt{C_i \exp(-q_{\alpha, k} V_i / \sqrt{2})}, S_i \sqrt{C_i \exp(q_{\alpha, k} V_i / \sqrt{2})} \right], i = 1, \dots, k$$

MC 検定に関連付けられた全体の p 値を計算するために、 P_{ij} を任意のサンプルペア (i, j) に関連付けられた p 値とします。その結果、多重比較検定に関連付けられた全体の p 値は、次のようになります。

$$P = \min\{P_{ij}, 1 \leq i < j \leq k\}$$

P_{ij} を計算するために、次の式を使用して、「方法 A1」で与えられた 2 サンプル計画のアルゴリズムを実行します。

$$se = V_i + V_j$$

ここで、 V_i は上で与えられたとおりです。

より具体的に言えば、 $n_i \neq n_j$ の場合、次のようになります。

$$P_{ij} = \min(\alpha_L, \alpha_U)$$

ここで、 $\alpha_L = \Pr(Q > z_L \sqrt{2})$ 、 $\alpha_U = \Pr(Q > z_U \sqrt{2})$ 、変数 z_L は、関数 $L(z, S_i, S_j, n_i, n_j)$ の最小根、変数 z_U は関数 $L(z, S_j, S_i, n_j, n_i)$ の最小根、および Q は前述の範囲分布を待つ確率変数です。

$n_i = n_j = n$ のとき、 $P_{ij} = \Pr(Q > |z_o| \sqrt{2})$ となります。ここで、

$$z_o = \frac{\ln S_i^2 - \ln S_j^2}{V_i + V_j}$$

付録 B: Bonett 検定および多重比較検定の妥当性

シミュレーション B1: Bonett 検定の妥当性 (2 サンプルモデル、バランス型およびアンバランス型計画)

異なる特性を持つ分布から、小中サイズのランダムサンプルのペアを生成しました。これらの分布には、以下が含まれます。

- 標準正規分布 ($N(0, 1)$)
- 対称的で裾の軽い分布。一様分布 ($U(0, 1)$) および両パラメータを 3 に設定したベータ分布 ($B(3, 3)$) を含む
- 対称的で裾の重い分布。自由度 5 と 10 の t 分布 ($t(5)$ 、 $t(10)$)、および位置 0、尺度 1 のラプラス分布 ($Lp1$) を含む
- 歪んだ裾の重い分布。尺度 1 の指数分布 (Exp)、および自由度 5 と 10 のカイ二乗分布 ($Chi(5)$ 、 $Chi(10)$) を含む
- 左方向に歪んだ裾の重い分布。具体的には、パラメータをそれぞれ 8 と 1 に設定したベータ分布 ($B(8, 1)$)

さらに、外れ値の直接効果を評価するために、次のように定義される混合正規分布からサンプルのペアを生成しました。

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

ここで、 p は混合パラメータ、 $1 - p$ は混入の比率 (外れ値の比率に等しい) です。この調査のために、2 つの混合正規母集団を選択しました。 $CN(0.9, 3)$ では母集団の 10% が外れ値、 $CN(0.8, 3)$ では母集団の 20% が外れ値になります。これら 2 つの分布は、対称的で、外れ値に起因する長い裾があります。

$\alpha = 0.05$ の目標有意水準を使用して、各分布の各サンプルペアで両側 Bonett 検定を実行しました。各ケースでシミュレートした有意水準は、サンプルペアの反復数 10,000 に基づいており、5% の目標有意水準を使用したため、シミュレーションの誤差は $\sqrt{0.95(0.05)/10,000} = 0.2\%$ でした。

下の表 1 に、シミュレーションの結果を要約します。

表 1 バランス型およびアンバランス型の 2 サンプル計画の両側 Bonett 検定のシミュレートした有意水準。目標有意水準は 0.05 です。

分布	n_1, n_2	シミュレートした水準	分布	n_1, n_2	シミュレートした水準
$N(0, 1)$	10, 10	0.038	Exp	10, 10	0.052

分布	n_1, n_2	シミュレートした水準	分布	n_1, n_2	シミュレートした水準
	20, 10	0.043		20, 10	0.051
	20, 20	0.045		20, 20	0.049
	30, 10	0.044		30, 10	0.044
	30, 20	0.046		30, 20	0.042
	25, 25	0.048		25, 25	0.043
	30, 30	0.048		30, 30	0.042
	40, 40	0.051		40, 40	0.042
	50, 50	0.047		50, 50	0.039
t (5)	10, 10	0.044	Chi (5)	10, 10	0.040
	20, 10	0.042		20, 10	0.043
	20, 20	0.046		20, 20	0.040
	30, 10	0.041		30, 10	0.039
	30, 20	0.046		30, 20	0.043
	25, 25	0.048		25, 25	0.042
	30, 30	0.043		30, 30	0.043
	40, 40	0.046		40, 40	0.040
	50, 50	0.050		50, 50	0.039

分布	n_1, n_2	シミュレートした水準	分布	n_1, n_2	シミュレートした水準
t(10)	10, 10	0.041	Chi(10)	10, 10	0.044
	20, 10	0.040		20, 10	0.042
	20, 20	0.045		20, 20	0.041
	30, 10	0.046		30, 10	0.043
	30, 20	0.045		30, 20	0.045
	25, 25	0.046		25, 25	0.046
	30, 30	0.048		30, 30	0.038
	40, 40	0.045		40, 40	0.042
	50, 50	0.051		50, 50	0.049
Lpl	10, 10	0.054	B(8, 1)	10, 10	0.053
	20, 10	0.056		20, 10	0.045
	20, 20	0.055		20, 20	0.048
	30, 10	0.057		30, 10	0.042
	30, 20	0.058		30, 20	0.047
	25, 25	0.057		25, 25	0.041
	30, 30	0.053		30, 30	0.040
	40, 40	0.047		40, 40	0.042
	50, 50	0.048		50, 50	0.038

分布	n_1, n_2	シミュレートした水準	分布	n_1, n_2	シミュレートした水準
B(3, 3)	10, 10	0.032	CN(0.9, 3)	10, 10	0.024
	20, 10	0.037		20, 10	0.022
	20, 20	0.042		20, 20	0.018
	30, 10	0.039		30, 10	0.019
	30, 20	0.038		30, 20	0.020
	25, 25	0.039		25, 25	0.019
	30, 30	0.041		30, 30	0.015
	40, 40	0.044		40, 40	0.020
	50, 50	0.046		50, 50	0.017
U(0, 1)	10, 10	0.030	CN(0.8, 3)	10, 10	0.022
	20, 10	0.032		20, 10	0.019
	20, 20	0.031		20, 20	0.020
	30, 10	0.034		30, 10	0.017
	30, 20	0.034		30, 20	0.020
	25, 25	0.034		25, 25	0.021
	30, 30	0.037		30, 30	0.017
	40, 40	0.043		40, 40	0.023
	50, 50	0.043		50, 50	0.020

表1で示したように、軽い裾から適度の裾を持つ、対称分布またはほぼ対称な分布では、サンプルサイズが小さくなると、Bonett 検定のシミュレートした有意水準は目標有意水準（0.05）より低くなります。一方、小さいサンプルを大きく歪んだ分布から得る場合、シミュレートした水準は目標水準より少し大きくなる傾向があります。

サンプルサイズが中規模または大きい場合、シミュレートした有意水準はすべての分布で目標水準に近くなります。実際、指数分布やベータ分布 Beta(8, 1)などの大きく歪んだ分布でも、検定はある程度正常に行われました。

さらに、外れ値は大きなサンプルよりも小さなサンプルにより大きく影響を与えるように見えます。混合正規母集団のシミュレートした有意水準は、2つのサンプルの最小サイズが20に達したときに、約0.020で安定しました。

2つのサンプルの最小サイズが20のとき、水平な一様分布と混合正規分布を除き、シミュレートした有意水準は一貫して区間[0.038, 0.058]内に収まります。シミュレートした有意

水準の 0.040 は、目標水準の 0.05 に対して少し保守的ですが、このタイプ I 過誤率は、大抵の実用目的においては許容範囲内です。したがって、2つのサンプルの最小サイズが 20 以上のとき、Bonett 検定は妥当であると結論付けます。

シミュレーション B2: MC 検定の妥当性 (多重サンプルモデル)

第 I 部: バランス型計画

バランス型計画を使用した多重サンプルモデルでの MC 検定の性能を調べるためにシミュレーションを行いました。前述のシミュレーション B1 で前述した分布のセットを使用して、同じ分布から k 個のサンプルを同じサイズで生成しました。計画のサンプル数を $k = 3$ 、 $k = 4$ 、 $k = 6$ とし、各実験の k 個のサンプルのサイズを 10、15、20、25、50、100 に固定しました。

$\alpha = 0.05$ の目標有意水準を使用して、各計画の同じサンプルで両側 Bonett 検定を実行しました。各ケースでシミュレートした有意水準は、サンプルペアの反復数 10,000 に基づいており、5% の目標有意水準を使用したため、シミュレーションの誤差は $\sqrt{0.95(0.05)/10,000} = 0.2\%$ でした。

下の表 2a および 2b に、シミュレーションの結果を要約します。

表 2a バランス型多重サンプル計画における両側多重比較検定のシミュレートした有意水準。検定の目標有意水準は 0.05 です。

分布	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	シミュレートした水準	n_i	シミュレートした水準	n_i	シミュレートした水準
N(0, 1)	10	0.038	10	0.038	10	0.036
	15	0.040	15	0.041	15	0.039
	20	0.039	20	0.040	20	0.041
	25	0.045	25	0.047	25	0.047
	50	0.046	50	0.046	50	0.052
	100	0.049	100	0.049	100	0.052
t(5)	10	0.042	10	0.044	10	0.042
	15	0.041	15	0.044	15	0.046
	20	0.043	20	0.045	20	0.045
	25	0.046	25	0.048	25	0.046

分布	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	シミュレートした水準	n_i	シミュレートした水準	n_i	シミュレートした水準
	50	0.040	50	0.039	50	0.038
	100	0.038	100	0.040	100	0.040
T(10)	10	0.033	10	0.037	10	0.038
	15	0.040	15	0.042	15	0.041
	20	0.042	20	0.043	20	0.043
	25	0.041	25	0.042	25	0.045
	50	0.047	50	0.044	50	0.047
	100	0.048	100	0.046	100	0.047
Lp1	10	0.056	10	0.063	10	0.071
	15	0.056	15	0.061	15	0.063
	20	0.054	20	0.058	20	0.059
	25	0.051	25	0.056	25	0.58
	50	0.045	50	0.051	50	0.049
	100	0.044	100	0.047	100	0.050
B(3, 3)	10	0.031	10	0.031	10	0.031
	15	0.037	15	0.036	15	0.034
	20	0.035	20	0.036	20	0.037
	25	0.039	25	0.038	25	0.040
	50	0.044	50	0.044	50	0.044
	100	0.044	100	0.046	100	0.043
U(0, 1)	10	0.029	10	0.025	10	0.023
	15	0.026	15	0.027	15	0.026
	20	0.028	20	0.030	20	0.028
	25	0.034	25	0.033	25	0.032
	50	0.041	50	0.036	50	0.036

分布	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	シミュレートした水準	n_i	シミュレートした水準	n_i	シミュレートした水準
Exp	100	0.048	100	0.047	100	0.045
	10	0.063	10	0.073	10	0.076
	15	0.056	15	0.058	15	0.064
	20	0.051	20	0.053	20	0.057
	25	0.043	25	0.045	25	0.050
	50	0.033	50	0.037	50	0.038
	100	0.033	100	0.035	100	0.035

表 2b バランス型多重サンプル計画における両側多重比較検定のシミュレートした有意水準。検定の目標有意水準は 0.05 です。

分布	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	シミュレートした水準	n_i	シミュレートした水準	n_i	シミュレートした水準
Chi (5)	10	0.040	10	0.046	10	0.048
	15	0.043	15	0.046	15	0.049
	20	0.040	20	0.040	20	0.042
	25	0.040	25	0.045	25	0.042
	50	0.037	50	0.038	50	0.040
	100	0.036	100	0.037	100	0.038
Chi (10)	10	0.042	10	0.045	10	0.045
	15	0.038	15	0.044	15	0.047
	20	0.036	20	0.039	20	0.040
	25	0.043	25	0.044	25	0.045
	50	0.041	50	0.040	50	0.042
	100	0.038	100	0.040	100	0.042

分布	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	シミュレートした水準	n_i	シミュレートした水準	n_i	シミュレートした水準
B(8, 1)	10	0.058	10	0.060	10	0.066
	15	0.057	15	0.061	15	0.064
	20	0.049	20	0.051	20	0.055
	25	0.044	25	0.046	25	0.050
	50	0.037	50	0.037	50	0.039
	100	0.037	100	0.038	100	0.039
CN(0.9, 3)	10	0.020	10	0.018	10	0.016
	15	0.022	15	0.020	15	0.017
	20	0.014	20	0.012	20	0.008
	25	0.011	25	0.011	25	0.008
	50	0.009	50	0.007	50	0.006
	100	0.010	100	0.008	100	0.008
CN(0.8, 3)	10	0.017	10	0.015	10	0.011
	15	0.013	15	0.011	15	0.008
	20	0.012	20	0.012	20	0.009
	25	0.013	25	0.010	25	0.009
	50	0.011	50	0.011	50	0.009
	100	0.014	100	0.012	100	0.010

表 2a および 2b で示したように、サンプルサイズが小さい場合、バランス型計画の対称分布およびほぼ対称な分布の MC 検定は一般に保守的です。一方、指数分布やベータ分布 Beta(8, 1) などの大きく歪んだ分布から得た小さいサンプルの場合、検定はリベラルです。ただし、サンプルサイズが増加するにつれ、シミュレートした有意水準が目標有意水準 (0.05) に近づきます。さらに、中規模のサンプルでは、サンプル数が検定の性能に大きな影響を与えるようには見えません。ただし、データに外れ値が混入していると、検定の性能に顕著な影響があります。データに外れ値がある場合、検定は一貫して過剰に保守的になります。

パート II: アンバランス型計画

アンバランス型計画での MC 検定の性能を調べるためにシミュレーションを行いました。シミュレーション B1 で前述した分布のセットを使用して、同じ分布から3個のサンプルを同じサイズで生成しました。1 番目の一連の実験では、最初の 2 つのサンプルのサイズは $n_1 = n_2 = 10$ 、3 番目のサンプルのサイズは $n_3 = 15, 20, 25, 50, 100$ でした。2 番目の一連の実験では、最初の 2 つのサンプルのサイズは $n_1 = n_2 = 15$ 、3 番目のサンプルのサイズは $n_3 = 20, 25, 30, 50, 100$ でした。3 番目の一連の実験では、最小サンプルサイズを 20 に設定し、最初の 2 つのサンプルのサイズを $n_1 = n_2 = 20$ 、3 番目のサンプルのサイズを $n_3 = 25, 30, 40, 50, 100$ としました。

$\alpha = 0.05$ の目標有意水準を使用して、各分布から得た同じ 3 つのサンプルで両側 MC 検定を実行しました。各ケースでシミュレートした有意水準は、サンプルペアの反復数 10,000 に基づいており、5% の目標有意水準を使用したため、シミュレーションの誤差は $\sqrt{0.95(0.05)/10,000} = 0.2\%$ でした。

下の表 3a および 3b に、シミュレーションの結果を要約します。

表 3b アンバランス型多重サンプル計画における多重比較検定のシミュレートした有意水準。検定の目標有意水準は 0.05 です。

分布	$n_1 = n_2 = 10$		$n_1 = n_2 = 15$		$n_1 = n_2 = 20$	
	n_3	シミュレートした水準	n_3	シミュレートした水準	n_3	シミュレートした水準
N(0, 1)	15	0.032	20	0.040	25	0.045
	20	0.037	25	0.039	30	0.041
	25	0.038	30	0.037	40	0.043
	50	0.041	50	0.044	50	0.041
	100	0.042	100	0.042	100	0.044
t(5)	15	0.040	20	0.042	25	0.043
	20	0.036	25	0.040	30	0.037
	25	0.044	30	0.036	40	0.038
	50	0.033	50	0.036	50	0.035
	100	0.032	100	0.031	100	0.032
t(10)	15	0.039	20	0.042	25	0.042
	20	0.038	25	0.041	30	0.040
	25	0.040	30	0.041	40	0.041
	50	0.037	50	0.043	50	0.042

分布	$n_1 = n_2 = 10$		$n_1 = n_2 = 15$		$n_1 = n_2 = 20$	
	n_3	シミュレートした水準	n_3	シミュレートした水準	n_3	シミュレートした水準
Lpl	100	0.036	100	0.039	100	0.040
	15	0.059	20	0.060	25	0.054
	20	0.057	25	0.054	30	0.051
	25	0.056	30	0.051	40	0.050
	50	0.049	50	0.051	50	0.050
B(3, 3)	100	0.048	100	0.047	100	0.046
	15	0.034	20	0.033	25	0.037
	20	0.031	25	0.035	30	0.039
	25	0.031	30	0.034	40	0.039
	50	0.036	50	0.039	50	0.038
U(0, 1)	100	0.035	100	0.039	100	0.039
	15	0.027	20	0.030	25	0.032
	20	0.030	25	0.030	30	0.031
	25	0.028	30	0.032	40	0.036
	50	0.039	50	0.034	50	0.037
Exp	100	0.042	100	0.038	100	0.042
	15	0.061	20	0.053	25	0.042
	20	0.060	25	0.052	30	0.047
	25	0.054	30	0.049	40	0.043
	50	0.050	50	0.046	50	0.041
	100	0.044	100	0.040	100	0.040

表 3b 多重サンプル、アンバランス型計画における MC 検定のシミュレートした有意水準。
 検定の目標有意水準は 0.05 です。

分布	$n_1 = n_2 = 10$		$n_1 = n_2 = 15$		$n_1 = n_2 = 20$	
	n_3	シミュレートした水準	n_3	シミュレートした水準	n_3	シミュレートした水準
Chi (5)	15	0.047	20	0.045	25	0.041
	20	0.043	25	0.042	30	0.039
	25	0.043	30	0.039	40	0.040
	50	0.039	50	0.037	50	0.040
	100	0.034	100	0.035	100	0.034
Chi (10)	15	0.043	20	0.042	25	0.042
	20	0.039	25	0.038	30	0.041
	25	0.040	30	0.041	40	0.038
	50	0.038	50	0.041	50	0.042
	100	0.035	100	0.034	100	0.035
B(8, 1)	15	0.056	20	0.052	25	0.048
	20	0.054	25	0.046	30	0.044
	25	0.050	30	0.047	40	0.046
	50	0.046	50	0.043	50	0.043
	100	0.043	100	0.042	100	0.044
CN(0.9, 3)	15	0.017	20	0.020	25	0.017
	20	0.020	25	0.019	30	0.012
	25	0.017	30	0.016	40	0.013
	50	0.019	50	0.016	50	0.012
	100	0.014	100	0.016	100	0.010
CN(0.8, 3)	15	0.012	20	0.013	25	0.013
	20	0.016	25	0.012	30	0.012
	25	0.014	30	0.010	40	0.010
	50	0.015	50	0.010	50	0.013

分布	$n_1 = n_2 = 10$		$n_1 = n_2 = 15$		$n_1 = n_2 = 20$	
	n_3	シミュレートした水準	n_3	シミュレートした水準	n_3	シミュレートした水準
	100	0.012	100	0.011	100	0.010

図 3a および 3b に示したシミュレートした有意水準は、前述のバランス型計画を使用した複数のサンプルに対するものと一致しています。したがって、MC 検定の性能は、アンバランス型計画による影響を受けているようには見えません。さらに、最小サンプルサイズが少なくとも 20 のとき、シミュレートした有意水準は、データに混入がある場合を除き、目標水準に近くなります。

結論として、最小サンプルが少なくとも 20 のとき、多重 (k) サンプルの MC 検定は、バランス型計画とアンバランス型計画の両方で正しく行われると言えます。ただし、より小さいサンプルの場合、検定は対称的およびほぼ対称的なデータでは保守的、大きく歪んだデータではリベラルになります。

付録 C: 理論上の検出力関数

MC 検定の正確な理論上の検出力関数はありません。ただし、2 サンプル計画の場合は、大サンプル理論方法に基づく近似の検出力関数を得られます。多重サンプル計画の場合、同様の近似を導出するには、さらなる研究が必要です。

ただし、2 サンプル計画の場合、Bonett 検定の理論上の検出力関数は、大サンプル理論の方法を使用して得られます。より具体的に言えば、下の式で与えられる検定統計量 T は、自由度 1 のカイ二乗分布として漸近的に分布します。

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

この T の式では、 $\hat{\rho} = S_1/S_2$ 、 $\rho = \sigma_1/\sigma_2$ 、 $g_i = (n_i - 3)/n_i$ で、 γ は 2 つの母集団の未知の共通尖度です。

近似の有意水準 α で等分散性を調べる両側 Bonett 検定の理論上の検出力関数は、次のように与えられます。

$$\pi(n_1, n_2, \rho) = 1 - \Phi\left(z_{\alpha/2} - \frac{\ln \rho^2}{se}\right) + \Phi\left(-z_{\alpha/2} - \frac{\ln \rho^2}{se}\right)$$

ここで

$$se = \sqrt{\frac{\gamma - g_1}{n_1 - 1} + \frac{\gamma - g_2}{n_2 - 1}}$$

片側検定の場合、 $\sigma_1 > \sigma_2$ に対して検定するとき、近似の検出力関数は次のとおりです。

$$\pi(n_1, n_2, \rho) = 1 - \Phi\left(z_{\alpha/2} - \frac{\ln \rho^2}{se}\right)$$

$\sigma_1 < \sigma_2$ に対して検定するとき、近似の検出力関数は次のとおりです。

$$\pi(n_1, n_2, \rho) = \Phi\left(-z_{\alpha/2} - \frac{\ln \rho^2}{se}\right)$$

データ分析のサンプルサイズの計画段階時には、母集団の共通尖度 γ は不明です。したがって実験者は通常、 γ の計画値を得るには、専門家の意見や以前の実験結果に頼らざるを得ません。この情報が手に入らない場合は、主要な研究の計画を立てるために、小規模な試験的調査を行うことをお勧めします。試験的調査のサンプルを使用し、 γ の計画値は、次の式により与えられる併合尖度として求められます。

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

アシスタントメニューでは、 γ の計画推定値は、ユーザーの手元のデータに基づいて、過去にさかのぼって求められます。

付録 D: 理論上の検出力とシミュレートした検出力の比較

シミュレーション D1: Bonett 検定のシミュレートした (実際の) 検出力

Bonett 検定のシミュレートした検出力水準と、付録 C で得た近似の検出力関数に基づく検出力水準を比較するためにシミュレーションを行いました。

前述した分布のそれぞれに、10,000 個のサンプルペアを生成しました（「シミュレーション B1」を参照）。一般に、シミュレーション B1 での結果によれば、選択したサンプルサイズは、検定のシミュレートした有意水準が目標有意水準に適度に近くなるのに十分な大きさでした。

標準偏差の比 $\rho = \sigma_1/\sigma_2 = 1/2$ でシミュレートした検出力水準を評価するために、すべてのサンプルペアの 2 番目のサンプルに定数 2 を掛けました。その結果、所与の分布と所与のサンプルサイズ n_1 および n_2 について、シミュレートした検出力水準は、サンプルペアの反復数 10,000 で両側 Bonett 検定が有意だった割合として計算されました。検定の目標有意水準は $\alpha = 0.05$ に固定されました。比較のために、付録 C で得た近似の検出力関数に基づき、対応する理論上の検出力水準を計算しました。

下の表 4 に結果を示します。

表 4 シミュレートした検出力水準と両側 Bonett 検定の近似の検出力水準の比較 目標有意水準は 0.05 です。

分布	n_1, n_2	近似の検出力	シミュレートした検出力	分布	n_1, n_2	近似の検出力	シミュレートした検出力
N(0, 1)	20, 10	0.627	0.527	Exp	20, 10	0.222	0.227
	20, 20	0.83	0.765		20, 20	0.322	0.368
	20, 30	0.896	0.846		20, 30	0.377	0.434
	20, 40	0.925	0.886		20, 40	0.412	0.475
	30, 15	0.825	0.771		30, 15	0.32	0.307
	30, 30	0.954	0.925		30, 30	0.458	0.50
	30, 45	0.98	0.97		30, 45	0.531	0.579
	30, 60	0.989	0.984		30, 60	0.575	0.622
t(5)	20, 10	0.222	0.379	Chi(5)	20, 10	0.355	0.347
	20, 20	0.322	0.569		20, 20	0.517	0.53
	20, 30	0.377	0.637		20, 30	0.597	0.616
	20, 40	0.412	0.69		20, 40	0.644	0.661
	30, 15	0.32	0.545		30, 15	0.513	0.51
	30, 30	0.458	0.733		30, 30	0.701	0.711
	30, 45	0.531	0.795		30, 45	0.781	0.793
	30, 60	0.575	0.828		30, 60	0.823	0.833
t(10)	20, 10	0.476	0.45	Chi(10)	20, 10	0.454	0.414
	20, 20	0.673	0.673		20, 20	0.646	0.631
	20, 30	0.756	0.749		20, 30	0.73	0.717
	20, 40	0.80	0.803		20, 40	0.776	0.771
	30, 15	0.668	0.659		30, 15	0.641	0.618
	30, 30	0.85	0.852		30, 30	0.828	0.819
	30, 45	0.91	0.911		30, 45	0.892	0.882
	30, 60	0.936	0.937		30, 60	0.921	0.912
Lp1	20, 10	0.321	0.33	B(8, 1)	20, 10	0.363	0.278
	20, 20	0.469	0.519		20, 20	0.528	0.463

分布	n_1, n_2	近似の検出力	シミュレートした検出力	分布	n_1, n_2	近似の検出力	シミュレートした検出力
	20, 30	0.545	0.585		20, 30	0.609	0.549
	20, 40	0.59	0.632		20, 40	0.655	0.60
	30, 15	0.466	0.475		30, 15	0.524	0.419
	30, 30	0.647	0.673		30, 30	0.713	0.634
	30, 45	0.729	0.758		30, 45	0.792	0.737
	30, 60	0.773	0.80		30, 60	0.833	0.777
B(3, 3)	20, 10	0.777	0.628	CN(0.9, 3)	20, 10	0.238	0.284
	20, 20	0.939	0.869		20, 20	0.346	0.452
	20, 30	0.973	0.936		20, 30	0.405	0.517
	20, 40	0.984	0.964		20, 40	0.442	0.561
	30, 15	0.935	0.871		30, 15	0.343	0.374
	30, 30	0.993	0.98		30, 30	0.491	0.598
	30, 45	0.998	0.995		30, 45	0.567	0.70
	30, 60	0.999	0.999		30, 60	0.612	0.719
U(0, 1)	20, 10	0.916	0.74	CN(0.8, 3)	20, 10	0.26	0.223
	20, 20	0.992	0.95		20, 20	0.379	0.396
	20, 30	0.998	0.985		20, 30	0.444	0.467
	20, 40	0.999	0.995		20, 40	0.484	0.52
	30, 15	0.991	0.941		30, 15	0.376	0.354
	30, 30	1.0	0.996		30, 30	0.535	0.549
	30, 45	1.0	1.0		30, 45	0.614	0.65
	30, 60	1.0	1.0		30, 60	0.661	0.706

結果は、一般に、近似の検出力水準とシミュレートした検出力水準はお互いに近くなることを示しています。これらの水準は、サンプルサイズが増加すると、より近づきます。軽い裾から適度の裾を持つ対称分布およびほぼ対称な分布では、近似の検出力水準は通常、シミュレートした検出力水準より少し大きくなります。ただし、裾の重い対称分布や大きく歪んだ分布では、シミュレートした検出力水準よりも少し小さくなります。この2つの検出力関数

の差は、サンプルが自由度 5 の t 分布から生成される場合を除けば、通常は重要ではありません。

全般的に、最小サンプルサイズが 20 に達すると、近似の検出力水準とシミュレートした検出力水準は著しく近くなります。したがって、サンプルサイズの計画は、近似の検出力関数に基づくことができます。

© 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.