

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。

2 サンプル不良%検定

概要

2 サンプルの比率の検定は、2 つの比率が有意に異なるかどうかを判断するために使用されます。品質分析では、この検定は製品またはサービスが不良品または良品として特徴付けられる場合に、不良品の割合が 2 つの独立した工程から収集されたサンプルと有意に異なるかどうかを判断するために多く使用されます。

Minitab アシスタントには、2 サンプル不良%検定が含まれています。検定用に収集されるデータは、2 つの独立したサンプルそれぞれに含まれる不良品の数です。これは、二項確率変数の観測値とみなされます。アシスタントでは、厳密法を使用して仮説検定の結果が計算されるため、実際のタイプ I 過誤率は検定に指定された有意水準 (α) に近くなり、さらなる調査は必要ありません。ただし、検出力およびサンプルサイズの分析を行うため、不良%の差の信頼区間 (CI)、および正規近似検定の理論上の検出力関数の計算には正規近似法が使用されます。これらは近似法であるため、その精度を評価する必要があります。

本書では、近似の信頼区間が正確になる条件を調べます。また、近似法の理論上の検出力と正確検定の実際の検出力を比較し、2 サンプル不良%検定の検出力とサンプルサイズを評価するために使用される方法についても調べます。最後に、自動的に実行され、アシスタントレポートカードに表示される次のデータチェックについて調べ、分析結果にどのように影響するのかについて説明します。

- 信頼区間の妥当性
- サンプルサイズ

2 サンプル不良%検定は、その他の仮定に依存します。詳細は、「付録 A」を参照してください。

2 サンプル不良%検定の方法

信頼区間の精度

アシスタントでは、2つのサンプルの不良%が有意に異なるかどうかを評価するために、Fisher の正確検定が使用されますが、差の信頼区間は正規近似法に基づきます。ほとんどの統計学の教科書に記されている一般規則によれば、各サンプルの不良品の観測数と良品の観測数が少なくとも5であれば、この近似の信頼区間は正確です。

目的

正規近似に基づく信頼区間が正確になる条件を評価する必要がありました。具体的には、各サンプルの不良品数と良品数に関する一般規則が、近似の信頼区間の精度にどのように影響するのかを調べる必要がありました。

方法

2つの比率の差の信頼区間を計算するために使用した計算式および信頼区間の精度を保証する一般規則については、「付録D」で説明します。また、この調査の過程で開発した、あまり厳しくない、修正された規則についても説明します。

さまざまな条件のもとで、近似の信頼区間の精度を評価するシミュレーションを行いました。シミュレーションを行うために、複数のベルヌイ母集団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。各タイプのベルヌイ母集団で、10,000 ベルヌイサンプル反復の各ペアの2つの比率の差の近似の信頼区間を計算しました。次に、シミュレートした被覆確率と呼ばれる、2つの比率の真の差が含まれている10,000区間の比率を計算しました。近似区間が正確な場合、シミュレートした被覆確率は目標被覆確率の0.95に近くなります。各サンプルに必要な不良品と良品の最小数の元の規則と修正した規則に関連して近似区間の精度を評価するため、10,000サンプルペアで各規則が満たされた割合も計算しました。詳細は、「付録D」を参照してください。

結果

2つの比率の差の近似の信頼区間は、サンプルが十分に大きい場合、つまり、各サンプルの不良品の観測数と良品の観測数が少なくとも5のとき、一般に正確になります。したがって、レポートカードの信頼区間の妥当性チェックにこの規則を採用しました。この規則は、一般によく当てはまりますが、過度に保守的になるケースや、2つの比率が0または1に近いといくぶん緩和されたりする場合があります。詳細は、「データチェック」セクションおよび「付録D」を参照してください。

理論上の検出力関数の性能

アシスタントでは、2つのベルヌイ母集団比率（2つのサンプルの不良%）を比較するために、Fisher の検定を使用して仮説検定が実行されます。ただし、この正確検定の検出力関数は簡単には導出されないため、対応する正規近似検定の理論上の検出力関数を使用して検出力関数を近似する必要があります。

目的

正規近似検定に基づく理論上の検出力関数を、アシスタントの2サンプル不良%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使用することが適切かどうかを判断する必要があります。そのために、この理論上の検出力関数にFisherの正確検定の実際の検出力が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

p 値の計算を含め、Fisher の正確検定の手法については、「付録 B」で詳しく説明します。正規近似検定に基づく理論上の検出力関数については、「付録 C」で定義します。これらの定義に基づき、2つのサンプルの不良率の差の分析にFisherの正確検定が使用される場合の実際の検出力水準（シミュレートした検出力水準と呼ぶ）を推定するシミュレーションを行いました。

シミュレーションを行うために、複数のベルヌイ母集団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。ベルヌイ母集団の各カテゴリで、10,000 サンプル反復の各ペアにFisherの正確検定を実行しました。各サンプルサイズで、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所与の差を検出するために、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、正規近似検定に基づき、対応する理論上の検出力も計算しました。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 E」を参照してください。

結果

シミュレーションから、一般に、正規近似検定の理論上の検出力関数とFisherの正確検定のシミュレートした検出力関数はほぼ等しくなることがわかりました。したがって、アシスタントでは、Fisherの正確検定を実行するときに実質的に重要な差を検出するために必要なサンプルサイズの推定には、正規近似検定の理論上の検出力関数が使用されます。

データチェック

信頼区間の妥当性

2 サンプル不良%検定では、正確検定を使用して不良率の差が評価されるので、各サンプルの不良品数と良品数が精度に大きな影響を与えることはありません。ただし、不良%の差の信頼区間は、正規近似に基づきます。各サンプルの不良品と良品の数が増えると、近似の信頼区間の精度も高くなります（「付録 D」を参照）。

目的



不良%の差の近似の信頼区間が正確になるため、サンプルの不良品数と良品数が十分であるかどうかを判断する必要がありました。

方法

ほとんどの統計学の教科書に記されている一般規則を使用しました。各サンプルに不良品と良品がそれぞれ少なくとも 5 つ含まれている場合、2 サンプル不良%検定の近似の信頼区間は正確です。詳細は、前述の「2 サンプル不良%検定の方法」セクションを参照してください。

結果

「2 サンプル%不良検定の方法」セクションに要約されたシミュレーションが示すとおり、信頼区間の精度は各サンプルの不良品と良品の最小数によって異なります。したがって、2 つの不良率の差の信頼区間の精度を評価するために、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 |
|---|---|
|  | 両方のサンプルに少なくとも 5 つの不良品と 5 つの良品があります。差の信頼区間は正確です。 |
|  | 少なくとも 1 つのサンプルで、不良品数または良品数のいずれかが 5 未満です。差の信頼区間は正確でない可能性があります。 |

サンプルサイズ

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があり、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。

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 サンプル不良%検定の検出力とサンプルサイズの分析は、正規近似検定の理論上の検出力関数に基づいており、通常は Fisher の正確検定の実際の検出力の適切な推定が得られます（「2 サンプル不良%検定の方法」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）。理論上の検出力関数は、不良%および組み合わせたサンプルの全体的な不良%の目標差の関数として表現できます。

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、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%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 サンプル不良%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 |
|---|---|
|  | 検定で不良%の差が検出されます。検出力に問題はありません。 または 検出力は十分です。検定で不良%の差は検出されませんでした、サンプルは少なくとも 90%の確率で所与の差を検出するのに十分な大きさです（検出力 \geq 90）。 |
|  | 検出力は十分な可能性があります。検定で不良%の差は検出されませんでした、サンプルは 80～90%の確率で所与の差を検出するのに十分な大きさです（ $80 \leq$ 検出力 < 90 ）。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。 |
|  | 検出力は十分でない可能性があります。検定で不良%の差は検出されませんでした、サンプルは 60～80%の確率で所与の差を検出するのに十分な大きさです（ $60 \leq$ 検出力 < 80 ）。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。 |
|  | 検出力は十分ではありません。検定で不良%の差は検出されませんでした。サンプルは少なくとも 60%の確率で所与の差を検出するのに十分な大きさではありません（検出力 $< .60$ ）。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。 |
|  | 検定で不良%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%および 90%の確率で検出できる差がレポートに示されることがあります。 |

参考文献

Arnold, S.F. (1990). *Mathematical statistics*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Casella, G., & Berger, R.L. (1990). *Statistical inference*. Pacific Grove, CA: Wadsworth, Inc.

付録 A: 2 サンプル不良%検定のその他の仮定

2 サンプル不良%検定は次の仮定に基づきます。

- 各サンプルのデータは、 n 個の別個の項目で構成されており、各項目は不良品または良品のいずれかに分類されます。
- 項目が不良品となる確率は、サンプル内の各項目で同じです。
- 項目が不良品となる可能性は、別の項目が不良品か良品かによる影響を受けません。

この検定には生データではなく要約データが入力されるため、これらの仮定はアシスタントレポートカードのデータチェックでは検証できません。

付録 B: Fisher の正確検定

ベルヌイ母集団から得た 2 つの独立したランダムサンプル X_1, \dots, X_{n_1} および Y_1, \dots, Y_{n_2} を観測し、次のようになるものとします。

$$p_1 = \Pr(X_i = 1) = 1 - \Pr(X_i = 0) \text{ および } p_2 = \Pr(Y_j = 1) = 1 - \Pr(Y_j = 0)$$

次のセクションで、比率 $\delta = p_1 - p_2$ の差を推定するための手順を説明します。

計算式 B1: Fisher の正確検定と p 値

Fisher の正確検定の説明は、Arnold (1994) を参照してください。ここでは、検定について簡単に説明します。

V を 1 番目のサンプルの成功数、 $v = n_1 \hat{p}_1$ を実験が実行されるときに 1 番目のサンプルで観測される成功数とします。また、 W を 2 つのサンプルの成功の総数、 $w = n_1 \hat{p}_1 + n_2 \hat{p}_2$ を実験が実行されるときに観測される成功数とします。 \hat{p}_1 と \hat{p}_2 は、 p_1 と p_2 のサンプルの点推定です。

$\delta = p_1 - p_2 = 0$ という帰無仮説のもとで、 W を与えられた V の条件付き分布は、確率質量関数を使用した超幾何分布です。

$$f(v|w) = \frac{\binom{n_1}{v} \binom{n_2}{w-v}}{\binom{n_1+n_2}{w}}$$

$F(v|w)$ を分布の CDF とします。すると、片側検定および両側検定の p 値は次のようになります。

- $\delta < 0$ 、または同等の $p_1 < p_2$ に対して検定する場合
p 値は $F(v|w)$ のように計算されます。 v は V の観測値または 1 番目のサンプルで観測される成功数で、 w は W の観測値または両方のサンプルで観測される成功数です。
- $\delta > 0$ 、または同等の $p_1 > p_2$ に対して検定する場合
p 値は $1 - F(v-1|w)$ のように計算されます。 v は V の観測値または 1 番目のサンプルで観測される成功数で、 w は W の観測値または両方のサンプルで観測される成功数です。
- $\delta \neq 0$ 、または同等の $p_1 \neq p_2$ に対して検定する場合
p 値は次のアルゴリズムに従って計算されます。 m は前述した超幾何分布の最頻値です。
 - $v < m$ の場合、p 値は $1 - F(y-1|w) + F(v|w)$ のように計算されます。 v および w は先に定義したとおりで、 $y = \min\{k \geq m: f(k|w) \leq f(v|w)\}$ です。
 - $v = m$ の場合、p 値は 1.0 です。
 - $v > m$ の場合、p 値は $1 - F(v-1|w) + F(y|w)$ のように計算されます。 v および w は前述したとおりで、 $y = \max\{k \leq m: f(k|w) \leq f(v|w)\}$ です。

付録 C: 理論上の検出力関数

2つの比率（より具体的に言えば、2つの不良%）を比較するために、「付録 B」で説明した Fisher の正確検定を使用します。この検定の理論上の検出力関数は、導出するには複雑すぎるため、近似の検出力関数を使用します。より具体的に言えば、2つの比率の既知の正規近似検定の検出力関数を使用して、Fisher の正確検定の検出力を近似します。

両側検定の正規近似の検出力関数は、次のようになります。

$$\pi(n_1, n_2, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha/2}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right) + \Phi\left(\frac{-\delta - z_{\alpha/2}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right)$$

ここで $\delta = p_1 - p_2$ で、

$$se = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

および $p_c = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$ となります。

$p_1 > p_2$ に対して $p_1 = p_2$ を検定するとき、検出力関数は次のようになります。

$$\pi(n_1, n_2, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right)$$

$p_1 < p_2$ に対して $p_1 = p_2$ を検定するとき、検出力関数は次のようになります。

$$\pi(n_1, n_2, \delta) = \Phi\left(\frac{-\delta - z_{\alpha}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right)$$

付録 D: 近似の信頼区間

計算式 D1: 2 サンプルの比率の差の近似の信頼区間の計算

正規近似に基づく $\delta = p_1 - p_2$ の漸近の $100(1 - \alpha)\%$ 信頼区間は、次のようになります。

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2}$$

この近似の信頼区間の信頼性を評価する既知の一般規則は、 $n_1\hat{p}_1 \geq 5$ 、 $n_1(1 - \hat{p}_1) \geq 5$ 、 $n_2\hat{p}_2 \geq 5$ 、および $n_2(1 - \hat{p}_2) \geq 5$ です。つまり、信頼区間は各サンプルで観測される成功数と失敗数が少なくとも 5 であれば正確であるということです。

注: このセクションおよび後続のセクションでは、信頼区間の規則を各サンプルの成功数と失敗数に関してもっとも一般的な形式で表現します。成功は対象事象を指し、失敗は対象事象を補完するものです。したがって、2 サンプル不良%検定という特定の状況では、「成功」の数は不良品の数に相当し、「失敗」の数は良品の数に相当します。

計算式 D2: 近似の信頼区間の規則

正規近似に基づく信頼区間に使用される一般規則では、信頼区間は $n_1\hat{p}_1 \geq 5$ 、 $n_1(1 - \hat{p}_1) \geq 5$ 、 $n_2\hat{p}_2 \geq 5$ 、および $n_2(1 - \hat{p}_2) \geq 5$ の場合に正確になります。つまり、成功（良品）数と失敗（不良品）数が各サンプルにそれぞれ少なくとも 5 つある場合、区間の実際の信頼水準は、目標信頼水準と等しいか、ほぼ等しくなります。

実際には真の比率は未知であるため、この規則は、真の比率とは対照的に、成功および失敗の推定比率に関して表現されています。ただし、真の比率が仮定されるか、または既知である理論的な設定では、規則で真の比率に関して直接表現される場合があります。これらのケースでは、真の期待される成功数および失敗数、 n_1p_1 、 n_2p_2 、 $n_1(1 - p_1)$ 、および $n_2(1 - p_2)$ が、比率の差の信頼区間の実際の被覆確率にどのように影響するのかを直接評価できません。

実際の被覆確率は、成功確率 p_1 と p_2 を使用した 2 つのベルヌイ母集団からサイズ n_1 と n_2 のサンプルペアを多数抽出することによって評価できます。実際の被覆確率は、2 つの比率の真の差を含む信頼区間を出力するサンプルペアの相対的度数として計算されます。実際の被覆確率は、 $n_1p_1 \geq 5$ 、 $n_2p_2 \geq 5$ 、 $n_1(1 - p_1) \geq 5$ 、および $n_2(1 - p_2) \geq 5$ の場合に大数の強法則によって正確で、被覆確率は $n_1\hat{p}_1 \geq 5$ 、 $n_1(1 - \hat{p}_1) \geq 5$ 、 $n_2\hat{p}_2 \geq 5$ 、および $n_2(1 - \hat{p}_2) \geq 5$ の場合に正確になります。したがって、この規則が有効であれば、実際と目標の信頼水準が近くなる時、2 つのベルヌイ母集団から生成されたサンプルペアの大部分が、 $n_1\hat{p}_1 \geq 5$ 、 $n_1(1 - \hat{p}_1) \geq 5$ 、 $n_2\hat{p}_2 \geq 5$ 、および $n_2(1 - \hat{p}_2) \geq 5$ のようになることが期待されます。後述のシミュレーションでは、この規則を規則 1 と呼びます。

さらに、この調査の過程で、多くのケースでは、 $n_1p_1 \geq 5$ および $n_2p_2 \geq 5$ 、または $n_1(1 - p_1) \geq 5$ および $n_2(1 - p_2) \geq 5$ のいずれかの場合、区間のシミュレートした被覆確率は目標被覆に近くなることに気がきました。これにより、 $n_1\hat{p}_1 \geq 5$ および $n_2\hat{p}_2 \geq 5$ 、または $n_1(1 - \hat{p}_1) \geq 5$ および $n_2(1 - \hat{p}_2) \geq 5$ の場合、近似の信頼区間が正確になるという代替の緩和された規則が生じました。後述のシミュレーションでは、この修正された規則を規則 2 と呼びます。

シミュレーション D1: 近似の信頼区間の精度の評価

2つの比率の差について、近似の信頼区間が正確になる条件を評価するシミュレーションを行いました。特に、次の一般規則に関連して、区間の精度を調べました。

規則 1 (元) $n_1 p_1 \geq 5, n_2 p_2 \geq 5, n_1(1 - p_1) \geq 5, \text{および} n_2(1 - p_2) \geq 5$

規則 2 (修正) $n_1 \hat{p}_1 \geq 5 \text{および} n_2 \hat{p}_2 \geq 5 \text{または} n_1(1 - \hat{p}_1) \geq 5 \text{および} n_2(1 - \hat{p}_2) \geq 5$

各実験で、次の比率で定義されるベルヌイ母集団ペアから 10,000 サンプルペアを生成しました。

- A 比率: p_1 と p_2 の両方が 1.0 に近い (または 0 に近い)。シミュレーションでこのベルヌイ母集団のペアを表すのに、 $p_1 = 0.8$ と $p_2 = 0.9$ を使用しました。
- B 比率: p_1 と p_2 の両方が 0.5 に近い。シミュレーションでこのベルヌイ母集団のペアを表すのに、 $p_1 = 0.4$ と $p_2 = 0.55$ を使用しました。
- C 比率: p_1 は 0.5 に近く、 p_2 は 1.0 に近い。シミュレーションでこのベルヌイ母集団のペアを表すのに、 $p_1 = 0.4$ と $p_2 = 0.9$ を使用しました。

上記の比率の分類は、近似の信頼区間が導出される二項分布に対する DeMoivre-Laplace の正規近似に基づいています。正規近似は、ベルヌイサンプルが 10 より大きく、成功確率が 0.5 に近いときに正確であることが知られています。成功確率が 0 または 1 に近い場合は、一般により大きいベルヌイサンプルが必要です。

両方のペアのサンプルサイズを単一値の n に固定しました。ここで $n = 10, 15, 20, 30, \dots, 100$ となります。どちらの規則も観測された成功数と失敗数に依存するため、分析は一般性の喪失がないバランス型計画 ($n_1 = n_2 = n$) に制限しました。これは、サンプルのサイズと成功の比率で制御可能です。

2 母集団の比率の差の信頼区間の実際の信頼水準 (シミュレートした信頼水準と呼ばれる) を推定するため、2つの比率の真の差が含まれる 10,000 区間の比率を計算しました。各実験の目標被覆確率は 0.95 でした。さらに、10,000 サンプルで 2つの規則が満たされた条件のパーセントを判定しました。

注: 一部の小さいサンプルでは、比率の差の推定された標準誤差が 0 でした。これらのサンプルは「退化」とみなし、実験から除外しました。その結果、サンプル反復数はいくつかのケースでは 10,000 より若干少なくなりました。

結果を表 1~表 11 に示し、グラフを図 1 に示します。

表 1 $n = 10$ の場合のシミュレートした被覆確率と規則 1 および規則 2 を満たすサンプルのパーセント。目標被覆確率は 0.95 です。

| n = 10 | | | | | | | |
|--------|--------|------|----------|------|----------------|----------------|------|
| カテゴリ | 比率 (p) | np | n(1 - p) | 被覆確率 | 規則 1 を満たすサンプル% | 規則 2 を満たすサンプル% | |
| A | p_1 | 0.80 | 8.00 | 2.00 | 0.907 | 0.0 | 99.1 |
| | p_2 | 0.90 | 9.00 | 1.00 | | | |

| n = 10 | | | | | | | |
|---------------|----------------------|--------|------|----------|-------|----------------|----------------|
| カテゴリ | | 比率 (p) | np | n(1 - p) | 被覆確率 | 規則 1 を満たすサンプル% | 規則 2 を満たすサンプル% |
| B | p₁ | 0.40 | 4.00 | 6.00 | 0.928 | 4.4 | 63.0 |
| | p₂ | 0.55 | 5.50 | 4.50 | | | |
| C | p₁ | 0.45 | 4.50 | 5.50 | 0.919 | 0.0 | 48.3 |
| | p₂ | 0.90 | 9.00 | 1.00 | | | |

表 2 n = 15 の場合のシミュレートした被覆確率と規則 1 および規則 2 を満たすサンプルのパーセント。目標被覆確率は 0.95 です。

| n = 15 | | | | | | | |
|---------------|----------------------|--------|-------|----------|-------|----------------|----------------|
| カテゴリ | | 比率 (p) | np | n(1 - p) | 被覆確率 | 規則 1 を満たすサンプル% | 規則 2 を満たすサンプル% |
| A | p₁ | 0.80 | 12.00 | 3.00 | 0.938 | 0.2 | 100.0 |
| | p₂ | 0.90 | 13.50 | 1.50 | | | |
| B | p₁ | 0.40 | 6.00 | 9.00 | 0.914 | 65.0 | 97.3 |
| | p₂ | 0.55 | 8.25 | 6.75 | | | |
| C | p₁ | 0.45 | 6.75 | 8.25 | 0.930 | 1.2 | 86.9 |
| | p₂ | 0.90 | 13.50 | 1.50 | | | |

表 3 n = 20 の場合のシミュレートした被覆確率と規則 1 および規則 2 を満たすサンプルのパーセント。目標被覆確率は 0.95 です。

| n = 20 | | | | | | | |
|---------------|----------------------|--------|-------|----------|-------|----------------|----------------|
| カテゴリ | | 比率 (p) | np | n(1 - p) | 被覆確率 | 規則 1 を満たすサンプル% | 規則 2 を満たすサンプル% |
| A | p₁ | 0.80 | 16.00 | 4.00 | 0.942 | 1.5 | 100.0 |
| | p₂ | 0.90 | 18.00 | 2.00 | | | |
| B | p₁ | 0.40 | 8.00 | 12.00 | 0.943 | 92.8 | 99.9 |
| | p₂ | 0.55 | 11.00 | 9.00 | | | |
| C | p₁ | 0.45 | 9.00 | 11.00 | 0.934 | 4.1 | 98.2 |
| | p₂ | 0.90 | 18.00 | 2.00 | | | |

表4 $n = 30$ の場合のシミュレートした被覆確率と規則1および規則2を満たすサンプルのパーセント。目標被覆確率は0.95です。

| $n = 30$ | | | | | | | |
|----------|-------|--------|-------|----------|-------|--------------|--------------|
| カテゴリ | | 比率 (p) | np | $n(1-p)$ | 被覆確率 | 規則1を満たすサンプル% | 規則2を満たすサンプル% |
| A | p_1 | 0.80 | 24.00 | 6.00 | 0.941 | 4.3 | 100.0 |
| | p_2 | 0.90 | 27.00 | 3.00 | | | |
| B | p_1 | 0.40 | 12.00 | 18.00 | 0.944 | 99.7 | 100.0 |
| | p_2 | 0.55 | 16.50 | 13.50 | | | |
| C | p_1 | 0.45 | 13.50 | 16.50 | 0.938 | 7.2 | 100.0 |
| | p_2 | 0.90 | 27.00 | 3.00 | | | |

表5 $n = 40$ の場合のシミュレートした被覆確率と規則1および規則2を満たすサンプルのパーセント。目標被覆確率は0.95です。

| $n = 40$ | | | | | | | |
|----------|-------|--------|-------|----------|-------|--------------|--------------|
| カテゴリ | | 比率 (p) | np | $n(1-p)$ | 被覆確率 | 規則1を満たすサンプル% | 規則2を満たすサンプル% |
| A | p_1 | 0.80 | 32.00 | 8.00 | 0.941 | 35.1 | 100.0 |
| | p_2 | 0.90 | 36.00 | 4.00 | | | |
| B | p_1 | 0.40 | 16.00 | 24.00 | 0.945 | 100.0 | 100.0 |
| | p_2 | 0.55 | 22.00 | 18.00 | | | |
| C | p_1 | 0.45 | 18.00 | 22.00 | 0.945 | 37.7 | 100.0 |
| | p_2 | 0.90 | 36.00 | 4.00 | | | |

表6 $n = 50$ の場合のシミュレートした被覆確率と規則1および規則2を満たすサンプルのパーセント。目標被覆確率は0.95です。

| $n = 50$ | | | | | | | |
|----------|-------|--------|-------|----------|-------|--------------|--------------|
| カテゴリ | | 比率 (p) | np | $n(1-p)$ | 被覆確率 | 規則1を満たすサンプル% | 規則2を満たすサンプル% |
| A | p_1 | 0.80 | 40.00 | 10.00 | 0.942 | 36.4 | 100.0 |
| | p_2 | 0.90 | 45.00 | 5.00 | | | |
| B | p_1 | 0.40 | 20.00 | 30.00 | 0.944 | 100.0 | 100.0 |
| | p_2 | 0.55 | 27.50 | 22.50 | | | |

| n = 50 | | | | | | | |
|--------|-------|--------|-------|----------|-------|----------------|----------------|
| カテゴリ | | 比率 (p) | np | n(1 - p) | 被覆確率 | 規則 1 を満たすサンプル% | 規則 2 を満たすサンプル% |
| C | p_1 | 0.45 | 22.50 | 27.50 | 0.935 | 38.3 | 100.0 |
| | | 0.90 | 45.00 | 5.00 | | | |
| | p_2 | | | | | | |

表 7 n = 60 の場合のシミュレートした被覆確率と規則 1 および規則 2 を満たすサンプルのパーセント。目標被覆確率は 0.95 です。

| n = 60 | | | | | | | |
|--------|-------|--------|-------|----------|-------|----------------|----------------|
| カテゴリ | | 比率 (p) | np | n(1 - p) | 被覆確率 | 規則 1 を満たすサンプル% | 規則 2 を満たすサンプル% |
| A | p_1 | 0.80 | 48.00 | 12.00 | 0.947 | 72.8 | 100.0 |
| | | 0.90 | 54.00 | 6.00 | | | |
| | p_2 | | | | | | |
| B | p_1 | 0.40 | 24.00 | 36.00 | 0.947 | 100.0 | 100.0 |
| | | 0.55 | 33.00 | 27.00 | | | |
| | p_2 | | | | | | |
| C | p_1 | 0.45 | 27.00 | 33.00 | 0.949 | 73.1 | 100.0 |
| | | 0.90 | 54.00 | 6.00 | | | |
| | p_2 | | | | | | |

表 8 n = 70 の場合のシミュレートした被覆確率と規則 1 および規則 2 を満たすサンプルのパーセント。目標被覆確率は 0.95 です。

| n = 70 | | | | | | | |
|--------|-------|--------|-------|----------|-------|----------------|----------------|
| カテゴリ | | 比率 (p) | np | n(1 - p) | 被覆確率 | 規則 1 を満たすサンプル% | 規則 2 を満たすサンプル% |
| A | p_1 | 0.80 | 56.00 | 14.00 | 0.939 | 71.7 | 100.0 |
| | | p_2 | 0.90 | 63.00 | | | |
| B | p_1 | 0.40 | 28.00 | 42.00 | 0.945 | 100.0 | 100.0 |
| | | p_2 | 0.55 | 38.50 | | | |
| C | p_1 | 0.45 | 31.50 | 38.50 | 0.944 | 71.8 | 100.0 |
| | | p_2 | 0.90 | 63.00 | | | |

表9 n = 80 の場合のシミュレートした被覆確率と規則1および規則2を満たすサンプルのパーセント。目標被覆確率は0.95です。

| n = 80 | | | | | | | |
|--------|--------|------|--------|-------|--------------|--------------|-------|
| カテゴリ | 比率 (p) | np | n(1-p) | 被覆確率 | 規則1を満たすサンプル% | 規則2を満たすサンプル% | |
| A | p_1 | 0.80 | 64.00 | 16.00 | 0.947 | 91.3 | 100.0 |
| | | 0.90 | 72.00 | 8.00 | | | |
| p_2 | | | | | | | |
| B | p_1 | 0.40 | 32.00 | 48.00 | 0.947 | 100.0 | 100.0 |
| | | 0.55 | 44.00 | 36.00 | | | |
| p_2 | | | | | | | |
| C | p_1 | 0.45 | 36.00 | 44.00 | 0.948 | 91.3 | 100.0 |
| | | 0.90 | 72.00 | 8.00 | | | |
| p_2 | | | | | | | |

表10 n = 90 の場合のシミュレートした被覆確率と規則1および規則2を満たすサンプルのパーセント。目標被覆確率は0.95です。

| n = 90 | | | | | | | |
|--------|--------|------|--------|-------|--------------|--------------|-------|
| カテゴリ | 比率 (p) | np | n(1-p) | 被覆確率 | 規則1を満たすサンプル% | 規則2を満たすサンプル% | |
| A | p_1 | 0.80 | 72.00 | 18.00 | 0.947 | 95.18 | 100.0 |
| | | 0.90 | 81.00 | 9.00 | | | |
| p_2 | | | | | | | |
| B | p_1 | 0.40 | 36.00 | 54.00 | 0.951 | 100.0 | 100.0 |
| | | 0.55 | 49.50 | 40.50 | | | |
| p_2 | | | | | | | |
| C | p_1 | 0.45 | 40.50 | 49.50 | 0.945 | 95.2 | 100.0 |
| | | 0.90 | 81.00 | 9.00 | | | |
| p_2 | | | | | | | |

表 11 $n = 100$ の場合のシミュレートした被覆確率と規則 1 および規則 2 を満たすサンプルのパーセント。目標被覆確率は 0.95 です。

| n = 100 | | | | | | | |
|---------|-------|--------|-------|--------|-------|----------------|----------------|
| カテゴリ | | 比率 (p) | np | n(1-p) | 被覆確率 | 規則 1 を満たすサンプル% | 規則 2 を満たすサンプル% |
| A | p_1 | 0.80 | 80.00 | 20.00 | 0.952 | 97.7 | 100.0 |
| | | 0.90 | 90.00 | 10.00 | | | |
| p_2 | | | | | | | |
| | | | | | | | |
| B | p_1 | 0.40 | 40.00 | 60.00 | 0.945 | 100.0 | 100.0 |
| | | 0.55 | 55.00 | 45.00 | | | |
| p_2 | | | | | | | |
| | | | | | | | |
| C | p_1 | 0.45 | 45.00 | 55.00 | 0.948 | 97.7 | 100.0 |
| | | 0.90 | 90.00 | 10.00 | | | |
| p_2 | | | | | | | |
| | | | | | | | |

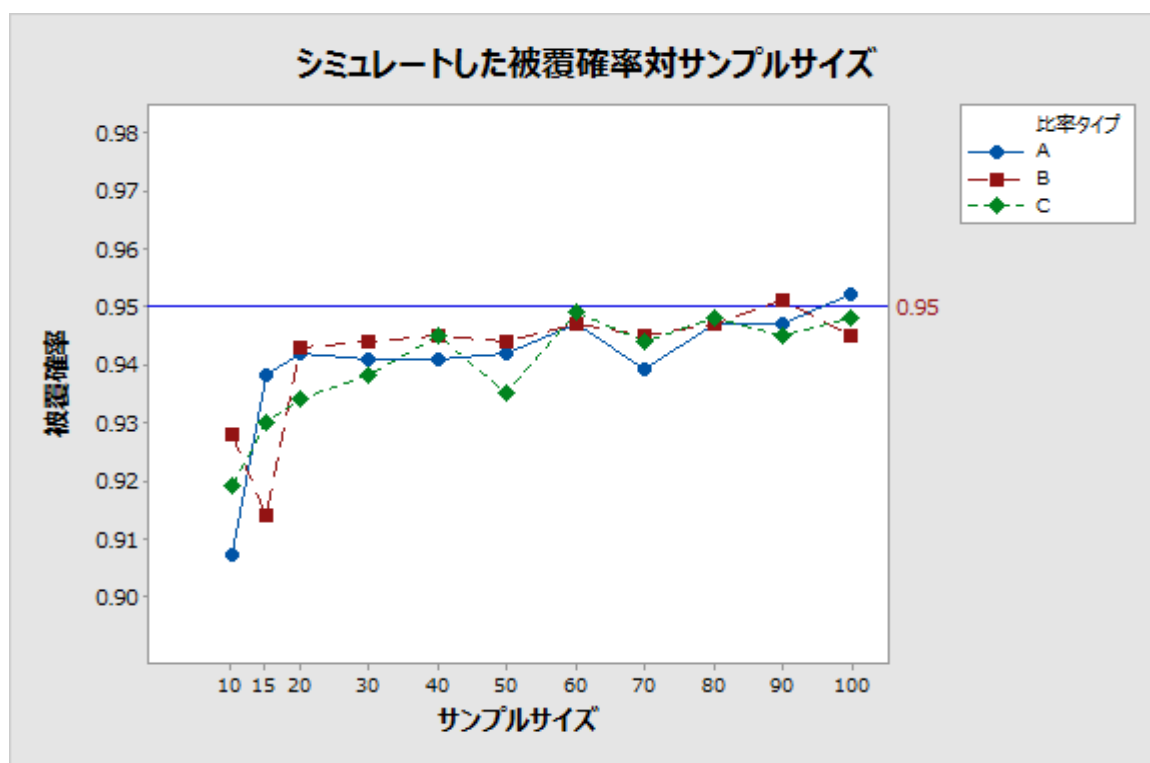


図 1 ベルヌイ母集団の各カテゴリのサンプルサイズに対してプロットされたシミュレートした被覆確率。

表 1～表 11 の結果および図 1 は、カテゴリ B (両方の比率が 0.5 に近い) ベルヌイ母集団から生成されたサンプルでは一般に、より安定した、目標被覆の 0.95 に近いシミュレートした被覆確率が得られます。このカテゴリでは、サンプルが小さい場合でも、両方の母集団の期待される成功数と失敗数は他のカテゴリよりも多くなります。

一方、カテゴリ A (両方の比率が 1.0 に近い) またはカテゴリ C (一方の比率が 1.0 に近く、他方は 0 に近い) のベルヌイ母集団ペアから生成されたサンプルの場合、シミュレートした被覆確率はより小さいサンプルでは目標から外れます。ただし、期待される成功数 (np) または期待される失敗数 ($n(1-p)$) が十分に大きい場合は除きます。

たとえば、 $n = 15$ の場合に、カテゴリ A のベルヌイ母集団から生成されたサンプルについて考えてみます。各母集団の期待される成功数は 12.0 と 13.5、期待される失敗数は 3.0 と 1.5' です。両方の母集団の期待される失敗数は 5 未満ですが、シミュレートした被覆確率は約 0.94 です。このような結果により、規則 2 が作成されました。規則 2 では、各サンプルの期待される成功数と期待される失敗数のいずれかが 5 以上となる必要があります。

規則 1 と規則 2 でどれぐらい効果的に信頼区間の近似を評価できるのかをより完全に評価するために、規則 1 を満たすサンプルのパーセントおよび規則 2 を満たすサンプルのパーセントを、実験のシミュレートした被覆確率に対してプロットしました。図 2 に、このプロットを示します。

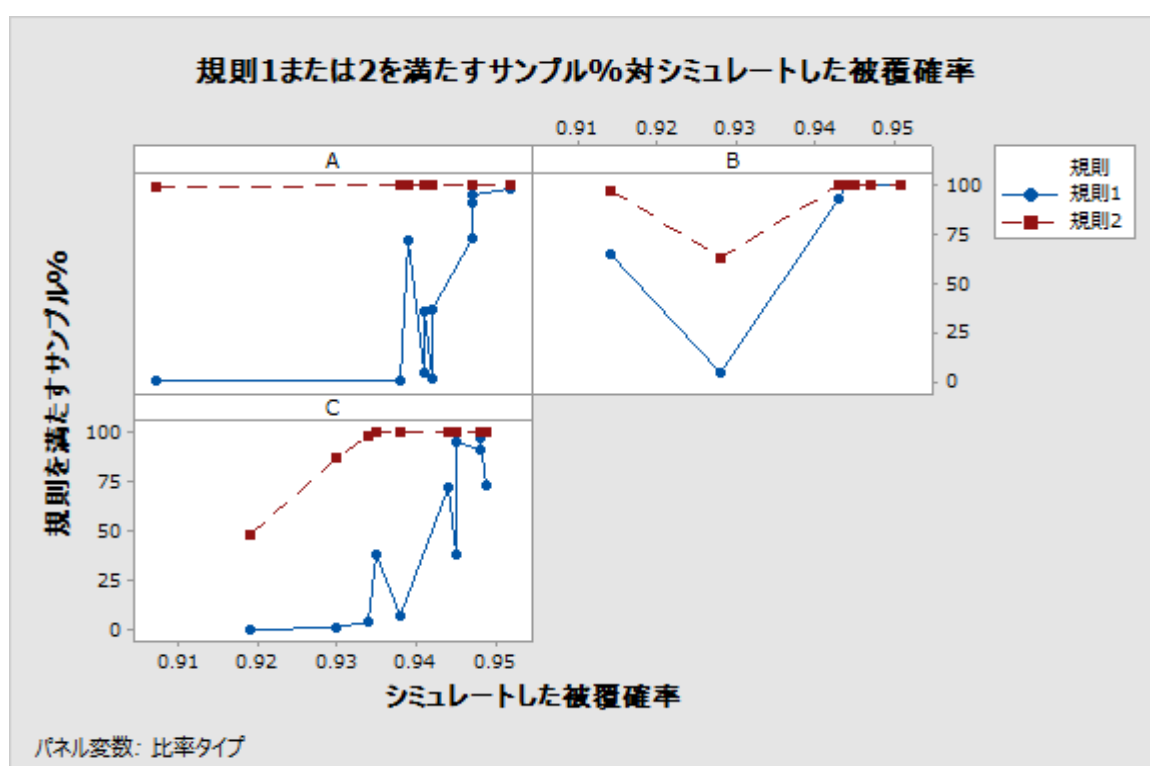


図 2 ベルヌイ母集団の各カテゴリで、シミュレートした被覆確率に対してプロットされた規則 1 と規則 2 を満たすサンプルのパーセント。

プロットは、シミュレートした被覆確率が目標被覆率の 0.95 に近づくにつれ、一般に各規則の要件を満たすサンプルのパーセントは 100% に近づきます。カテゴリ A および C のベルヌイ母集団から生成されたサンプルの場合、シミュレートした被覆確率が目標に近い場合でも、規則を満たすサンプルのパーセントが極端に低いことから明らかなように、規則 1 はサンプルが小さい場合に厳しくなります。たとえば、 $n = 20$ で、サンプルがカテゴリ A のベルヌイ母集団から生成される場合、シミュレートした被覆確率は 0.942 です (表 3 を参照)。ただし、規則を満たすサンプルの比率はほぼ 0 (0.015) です (図 2 を参照)。したがって、これらのケースでは、規則が保守的すぎる可能性があります。

一方、規則 2 は、カテゴリ A のベルヌイ母集団から生成された小さいサンプルでは、それほど厳しくありません。たとえば、表 1 で示したように、 $n = 10$ で、サンプルがカテゴリ A のベルヌイ母集団から生成される場合、シミュレートした被覆確率は 0.907 で、規則を満たすサンプルは 99.1% です。

結論として、規則 1 はサンプルが小さいときに過度に保守的になる傾向があります。規則 2 はそれほど保守的ではなく、サンプルサイズが小さいときに好まれる場合があります。ただし、規則 1 は既知の広く受け入れられている規則です。規則 2 は期待できる可能性を示していますが、前述したとおり、一部のケースではリベラルすぎる場合があります。1 つの可能性として、2 つの規則を組み合わせて、各規則の長所を生かすということがありますが、この方法を適用するためにはさらに調査が必要です。

付録 E: 実際の検出力と理論上の検出力の比較

シミュレーション E1: Fisher の正確検定を使用した実際の検出力の推定

Fisher の正確検定の推定される実際の検出力水準（シミュレートした検出力水準と呼ぶ）と、正規近似の検出力関数に基づく理論上の検出力水準（近似の検出力水準と呼ぶ）を比較するシミュレーションを設計しました。各実験で、ベルヌイ母集団ペアから 10,000 サンプルペアを生成しました。各サンプルペアで、比率の差が $p_1 - p_2 = -0.20$ となるように比率が選択されました。

- A 比率: p_1 と p_2 の両方が 1.0 に近い（または 0 に近い）。シミュレーションでこのベルヌイ母集団のペアを表すのに、 $p_1 = 0.70$ と $p_2 = 0.90$ を使用しました。
- B 比率: p_1 と p_2 の両方が 0.5 に近い。シミュレーションでこのベルヌイ母集団のペアを表すのに、 $p_1 = 0.40$ と $p_2 = 0.60$ を使用しました。
- C 比率: p_1 は 0.5 に近く、 p_2 は 1.0 に近い。シミュレーションでこのベルヌイ母集団のペアを表すのに、 $p_1 = 0.55$ と $p_2 = 0.75$ を使用しました。

両方のペアのサンプルサイズを単一値の n に固定しました。ここで $n = 10, 15, 20, 30, \dots, 100$ となります。通常 2 つのサンプルのサイズは同じと仮定されるため、分析はバランス型計画 ($n_1 = n_2 = n$) に制限しました。特定の検出力を使用して実質的に重要な差を検出するのに必要な共通のサンプルサイズを計算しました。

各シミュレーションの結果に基づいて Fisher の正確検定の実際の検出力を推定するために、目標有意水準の $\alpha = 0.05$ で両側検定が有意だった 10,000 サンプルペアの割合を計算しました。次に、比較のために、正規近似検定に基づき、対応する理論上の検出力水準を計算しました。次の表 12 に結果を示します。

表 12 ベルヌイ母集団の 3 つのカテゴリの近似の検出力水準と比較した Fisher の正確検定のシミュレートした検出力水準。目標有意水準は $\alpha = 0.05$ です。

| n | A 比率 | | B 比率 | | C 比率 | |
|----|------------------------------|------------|------------------------------|------------|------------------------------|------------|
| | $p_1 = 0.70$ $p_2 = 0.90$ | | $p_1 = 0.40$ $p_2 = 0.60$ | | $p_1 = 0.55$ $p_2 = 0.75$ | |
| | シミュレートした 検出力 | 近似の 検出力 | シミュレートした 検出力 | 近似の 検出力 | シミュレートした 検出力 | 近似の 検出力 |
| 10 | 0.063 | 0.193 | 0.056 | 0.140 | 0.056 | 0.149 |
| 15 | 0.151 | 0.271 | 0.097 | 0.190 | 0.101 | 0.204 |
| 20 | 0.244 | 0.348 | 0.146 | 0.240 | 0.183 | 0.259 |

| n | A 比率 | | B 比率 | | C 比率 | |
|-----|------------------------------|------------|------------------------------|------------|------------------------------|------------|
| | $p_1 = 0.70$ $p_2 = 0.90$ | | $p_1 = 0.40$ $p_2 = 0.60$ | | $p_1 = 0.55$ $p_2 = 0.75$ | |
| | シミュレートした 検出力 | 近似の 検出力 | シミュレートした 検出力 | 近似の 検出力 | シミュレートした 検出力 | 近似の 検出力 |
| 30 | 0.370 | 0.490 | 0.256 | 0.338 | 0.272 | 0.366 |
| 40 | 0.534 | 0.612 | 0.371 | 0.431 | 0.381 | 0.466 |
| 50 | 0.641 | 0.711 | 0.477 | 0.516 | 0.491 | 0.556 |
| 60 | 0.726 | 0.789 | 0.536 | 0.593 | 0.560 | 0.635 |
| 70 | 0.814 | 0.849 | 0.610 | 0.661 | 0.649 | 0.703 |
| 80 | 0.870 | 0.893 | 0.660 | 0.720 | 0.716 | 0.760 |
| 90 | 0.907 | 0.925 | 0.716 | 0.770 | 0.772 | 0.808 |
| 100 | 0.939 | 0.948 | 0.792 | 0.812 | 0.812 | 0.848 |

表 12 の結果は、ベルヌイ母集団の 3 つすべてのカテゴリ (A、B、C) のシミュレートした検出力よりも近似の検出力が高くなる傾向にあることを示しています。たとえば、カテゴリ A の比率では、0.91 の近似の検出力水準で -0.20 の絶対差を検出するのに必要な実際のサンプルサイズは約 90 です。対照的に、近似の理論上の検出力関数に基づく対応するサンプルサイズの推定値は、約 85 です。したがって、近似の検出力関数に基づくサンプルサイズの推定値は、与えられた検出力水準を達成するのに必要な実際のサンプルサイズより、一般にやや小さくなります。

この関係は、次の図3に示すように、結果を検出力曲線として示すとより明確に観察できます。

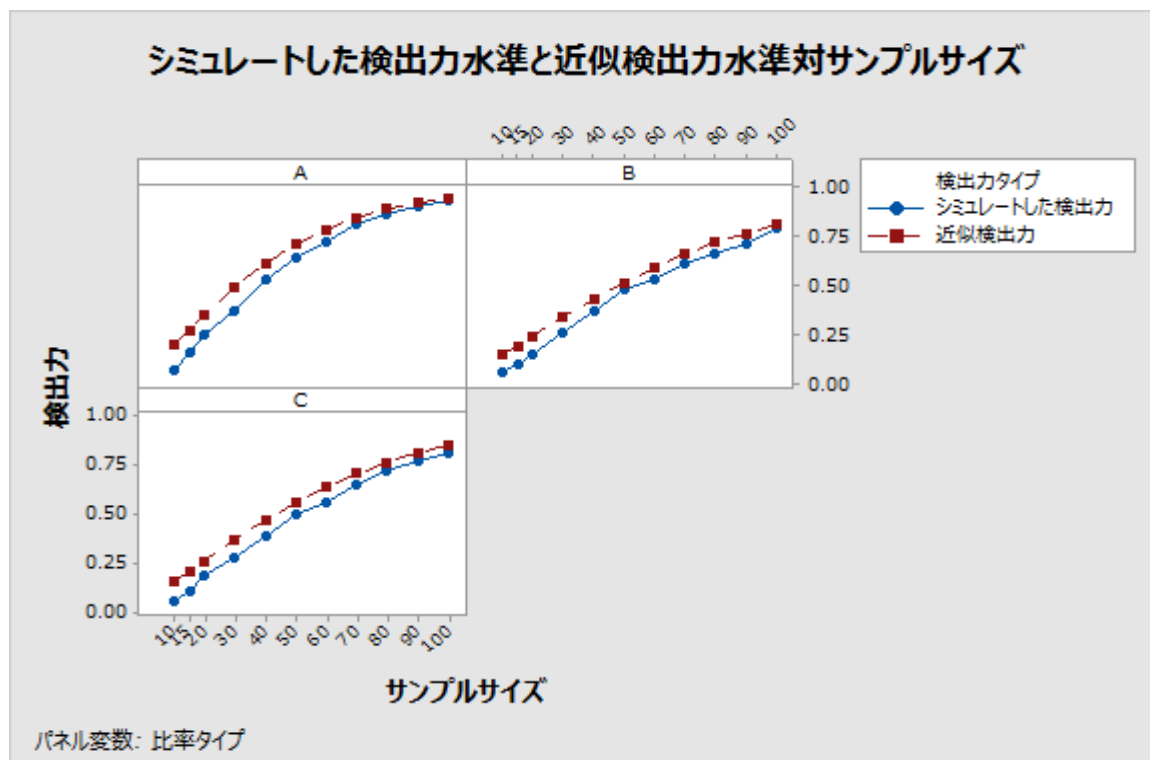


図3 2つの比率を比較する両側検定のシミュレートした検出力水準と近似検出力水準のプロット。検出力水準は、ベルヌイ母集団のカテゴリごとに別々のパネルでサンプルサイズに対してプロットされています。

ベルヌイ母集団の3つすべてのカテゴリ (A、B、C) で、シミュレートした検出力曲線は近似検出力曲線より低くなっていますが、曲線間の差のサイズは、サンプルが抽出されるベルヌイ母集団の真の比率によって異なることに注目してください。たとえば、2つの比率が0.5に近いとき (カテゴリ B)、一般に2つの検出力水準は近くなります。ただし、母集団カテゴリ A と C に関連付けられた比率では、2つの検出力曲線の格差は小さいサンプルでより顕著になります。

これらの結果から、一般に、正規近似検定の理論上の検出力関数と Fisher の正確検定のシミュレートした検出力関数はほぼ等しくなることがわかりました。したがって、アシスタントでは、Fisher の正確検定を実行する前に、正規近似検定の理論上の検出力関数を使用してサンプルサイズを推定します。ただし、近似の検出力関数を使用して計算されるサンプルサイズは、2つの比率の差 (不良%) を検出する与えられた検出力を達成するのに必要な実際のサンプルサイズより少し小さい場合があります。

© 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.