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 サンプル不良%検定

概要

1 サンプルの比率の検定は、比率が目標値と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、不良品の割合が目標不良%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、1 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、サンプルに含まれる不良品の数です。これは、二項の確率変数の観測値と仮定されます。アシスタントでは、直接法を使用して仮説検定の結果と信頼区間が計算されるため、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、1 サンプル不良%検定の検出力とサンプルサイズの分析は、近似に基づいており、その精度を評価する必要があります。

本書では、1 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する手法について調べ、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ます。

また、不良品の割合と目標不良%が異なるかどうかを検出するのに、サンプルサイズの大きさが十分であるかどうかを評価するガイドラインをどのように確立したのかについても説明します。アシスタントでは、サンプルサイズのチェックが自動的に行われ、レポートカードで結果が報告されます。

また、1 サンプル不良%検定は、その他の仮定にも依存します。詳細は、「付録 A」を参照してください。

1 サンプル不良%の方法

理論上の検出力関数の性能

アシスタントでは、直接（尤度比）法を使用して、1 つのベルヌイ母集団比率（不良%）の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が近似されます。

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 1 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した。そのために、この理論上の検出力関数に正確（尤度比）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した。

方法

検定統計量の p 値、および正確（尤度比）検定の信頼区間は、「付録 B」で定義します。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、正確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。

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を生成しました。各ベルヌイ母集団で、10,000 個のサンプル反復に対してそれぞれ正確検定を実行しました。各サンプルサイズで、サンプル 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 D」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正確（尤度比）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、正規近似検定の理論上の検出力関数を使用して、正確検定で実質的に重要な不良率の差を検出するのに十分な検出力を確保するために必要なサンプルサイズが推定されます。

データチェック

サンプルサイズ

一般的に、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した。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。

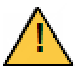
方法

1 サンプル不良%検定の検出力とサンプルサイズの分析は、正規近似を使用した理論上の検出力関数に基づいており、正確検定の実際の検出力の適切な推定が得られます（前述の「1 サンプル不良%の方法」セクションを参照）。目標不良%が与えられている場合、理論上の検出力関数はサンプルサイズおよび検出する必要がある差に依存します。

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、アシスタントでは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実質的な差を指定すると、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、1 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%と目標値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。検定で目標値との差は検出されませんでしたが、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≥ .90）。
	検出力は十分な可能性があります。検定で目標値との差は検出されませんでしたが、サンプルは少なくとも 80%～90%の確率で差を検出するのに十分な大きさです（.80 ≤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%～80%の確率で差を検出するのに十分な大きさです（.60 ≤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。

ステータス	状態
	検出力は十分ではありません。検定で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（検出力く .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。
	検定で目標値と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。

参考文献

Arnold, S.F.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：1 サンプル不良%のその他の仮定

1 サンプル不良%検定は次の仮定に基づきます。

- 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：正確（尤度比）検定

成功確率$p = \Pr(X_i = 1) = 1 - \Pr(X_i = 0)$を持つベルヌイ分布から得たランダムサンプル$X_1, \ldots, X_n$を観測するとします。

pに関する推論結果を導く直接法を以下に示します。

計算式 B1：正確検定と p 値

$H_A: p > p_0$、$H_A: p < p_0$、または$H_A: p \neq p_0$という対立仮説に対する帰無仮説$H_0: p = p_0$の検定について検討します。

$X = \sum_{i=1}^{n} X_i$とします。

Xは試行回数n、成功確率pを持つ二項確率変数です。

Xに基づく片側検定は、UMP（一様最強力）であり、尤度比検定です。両側検定の場合も、尤度比検定はXに基づき、検定統計量は

$$\Lambda(X) = \left(\frac{\hat{p}}{p_0}\right)^X \left(\frac{1-\hat{p}}{1-p_0}\right)^{n-X}$$

です（Arnold、1990 を参照）。

片側検定の p 値は、Xの正確な分布に基づいて直接求めることができます。両側検定では、p 値は帰無仮説のもとで、実際に観測された尤度比と少なくとも同等の尤度比（または対数尤度比）が観測される確率として計算されます。数値的な求根アルゴリズムは、一般にこの確率を計算するために使用されます。

計算式 B2：正確な信頼区間

pの正確な$100(1-\alpha)\%$両側信頼区間は

$$\frac{1}{1 + \dfrac{n-x+1}{x} F_{2(n-x+1),2x,\alpha/2}} \leq p \leq \frac{\dfrac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}{1 + \dfrac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}$$

です。ここで、xは観測された成功回数、$F_{\nu_1,\nu_2,\alpha}$は自由度ν_1およびν_2のF分布の上位α百分位数点です（Casella and Berger、1990 を参照）。$x = 0$のとき下側限界は 0、$x = n$のとき上側限界は 1 という規則を採用します。

付録C：理論上の検出力関数

正確検定の理論上の検出力関数は、複雑すぎて導出できません。したがって、正規近似に基づく検定の理論上の検出力関数を使用して、検定の検出力関数を近似します。この近似検定は、確率変数

$$Z = \frac{n^{1/2}(\hat{p} - p)}{(p(1-p))^{1/2}}$$

が標準正規分布として漸近的に分布されるという事実に基づきます。この検定の理論上の検出力関数は、十分に証明された既知の関数です。両側対立仮説の場合、検出力関数は次のように与えられます。

$$\pi(n,\delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha/2}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right) + \Phi\left(\frac{-\delta - z_{\alpha/2}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right)$$

ここで、$p = \delta + p_o$、$\Phi(.)$は標準正規分布の累積分布関数で、z_αは標準正規分布の上位百分位数点です。

片側対立仮説$H_A\colon p > p_0$の場合、検出力関数は次のように与えられることがあります。

$$\pi(n,\delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right)$$

片側対立仮説$H_A\colon p < p_0$に対して検定する場合、検出力関数は次のように与えられることもあります。

$$\pi(n,\delta) = \Phi\left(\frac{-\delta - z_{\alpha}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right)$$

付録 D：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 D1：正確検定を使用した実際の検出力の推定

実際の検出力水準の推定値（シミュレートした検出力水準と呼ぶ）と正規近似検定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、所与の成功確率pのベルヌイ母集団から、サイズnのサンプルを 10,000 個生成しました。成功確率については、（1）p値が 0.5 に近い（厳密には$p = 0.45$）中程度の成功確率、（2）p 値が 0 または 1 に近い（厳密には$p = 0.85$）小さいまたは大きい成功確率、という 2 つのケースを検討しました。これらの 2 つのケースを検討するのは、正規近似検定が導出される、二項分布への DeMoivre-Laplace 正規近似は、ベルヌイサンプルのサイズが 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るからです。ただし、成功確率がこれより小さいまたは大きい場合、近似が正確に行われるためにはより大きなベルヌイサンプルが必要となります。

各実験で、サンプルサイズを単一値のnに固定しました。ここで$n = 10, 15, 20, 30, \ldots, 100$です。すべての実験で、サンプルサイズを 100 まで増加させるときに、検出力が小さすぎたり、大きすぎたりしないように、検出対象の差$\delta = p - p_0$を 0.2 に固定しました。各シミュレーションの結果に基づいて検定の実際の検出力を推定するため、片側検定と両側検定の両方を使用して、10,000 個のサンプル反復のうち、$\alpha = 0.05$の目標有意水準で正確検定が有意だった割合を計算しました。最後に、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 1 に結果を示します。

表 1　両側および片側正確検定のシミュレートした検出力水準と近似の検出力水準 目標有意水準は$\alpha = 0.05$です。

n	両側検定				片側検定			
	$p = 0.45$		$p = 0.85$		$p = 0.45$		$p = 0.85$	
	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	シミュレートした検出力	近似の検出力
10	0.101	0.333	0.200	0.199	0.257	0.436	0.200	0.335
15	0.339	0.441	0.322	0.327	0.339	0.550	0.322	0.489
20	0.406	0.537	0.409	0.455	0.590	0.643	0.648	0.621
30	0.632	0.690	0.708	0.674	0.632	0.779	0.708	0.808
40	0.781	0.799	0.863	0.822	0.781	0.867	0.863	0.911

n	両側検定				片側検定			
	$p = 0.45$		$p = 0.85$		$p = 0.45$		$p = 0.85$	
	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	シミュレートした検出力	近似の検出力
50	0.877	0.872	0.874	0.910	0.877	0.921	0.933	0.961
60	0.878	0.920	0.942	0.957	0.922	0.954	0.969	0.984
70	0.925	0.951	0.972	0.981	0.953	0.973	0.987	0.994
80	0.954	0.971	0.986	0.992	0.986	0.985	0.993	0.998
90	0.971	0.982	0.993	0.996	0.991	0.991	0.996	0.999
100	0.989	0.990	0.998	0.999	0.994	0.995	0.999	1.000

結果は、シミュレートした検出力水準と近似の検出力水準は一般にかなり一致することを示しています。この一致は、次の図 1 と図 2 に示すように、結果を検出力曲線としてグラフ表示するとより明らかに見ることができます。

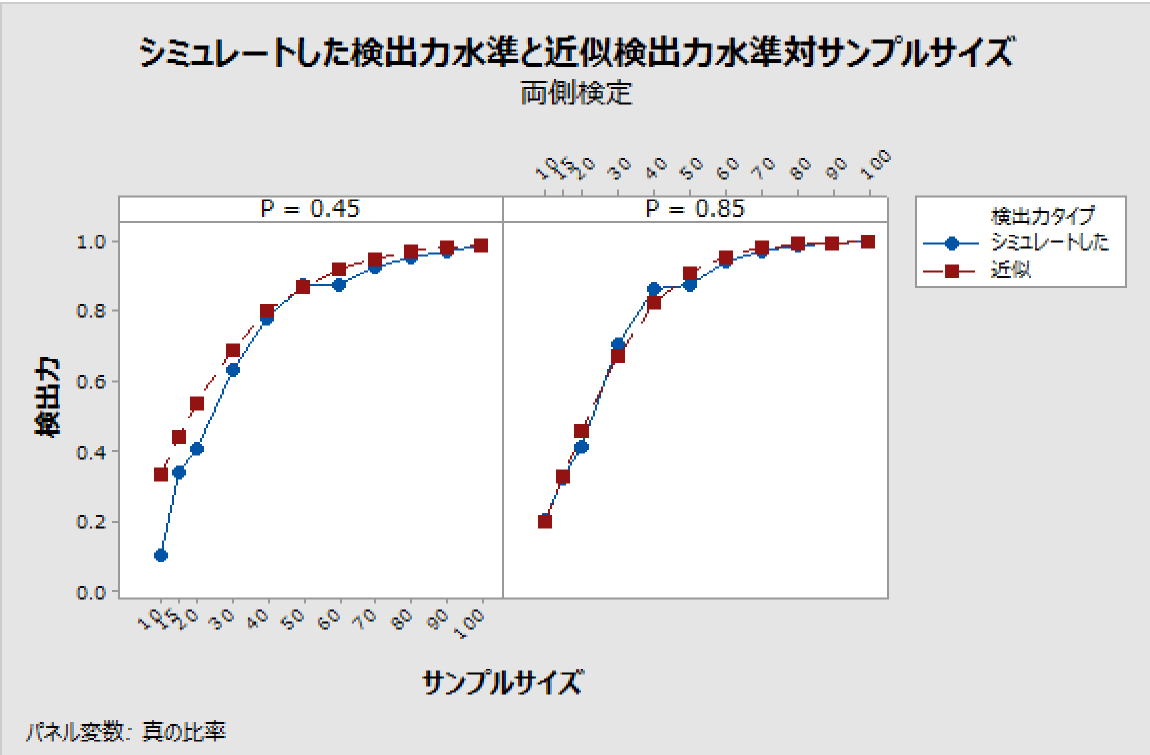


図 1　サンプルサイズと対比した両側正確検定のシミュレートした検出力水準と近似の検出力水準のプロット。

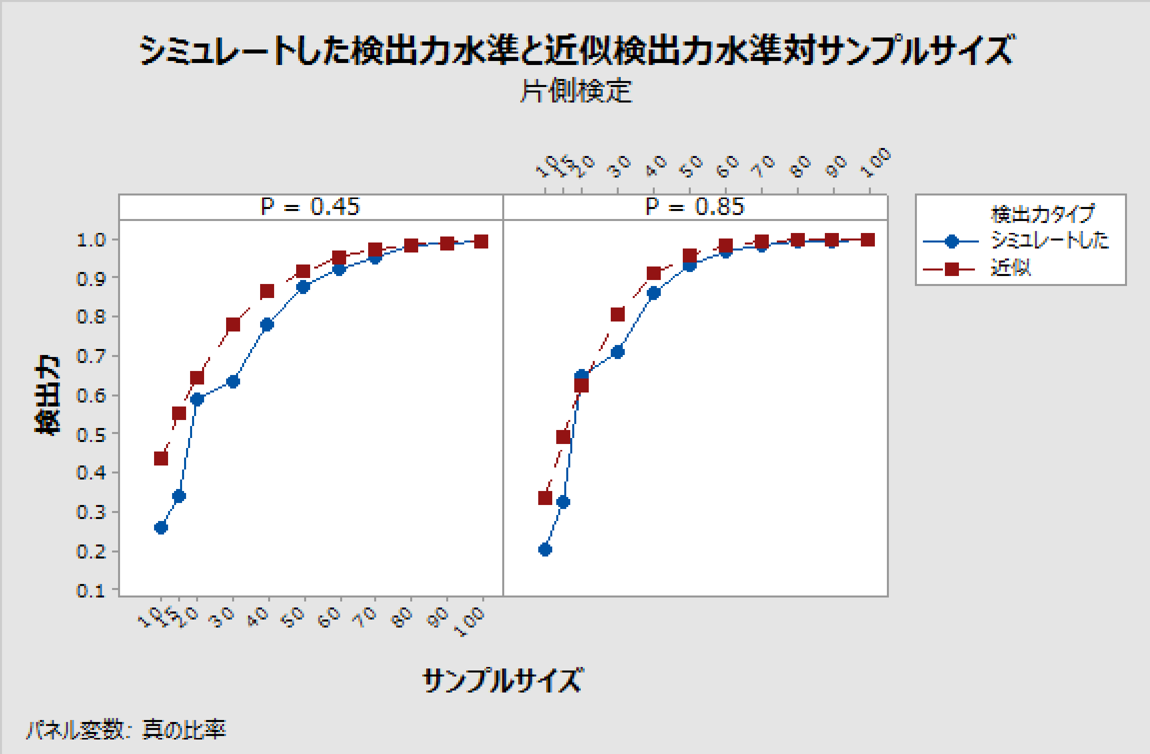


図 2　サンプルサイズと対比した片側正確検定のシミュレートした検出力水準と近似の検出力水準のプロット。

図 1 と図 2 の各パネルに示された 2 つの検出力曲線は、サンプルサイズが小さいいくつかの場合を除き、互いに近くなっています。曲線の近さは、正確検定が実際に適用される場合、近似の検出力関数はシミュレートした検出力とよく一致することを示しています。したがって、サンプルサイズの推定に近似の検出力関数を使用することは適切だと言えます。

また、図 1 と図 2 は、理論上の（近似）検出力曲線が一般にシミュレートした検出力曲線よりも高くなることも示しています。近似の検出力曲線がより高くなるのは、理論上の検出力水準が、目標有意水準の厳密値（0.05）を仮定して計算されるためです。それに比べ、正確検定は、特に小さいサンプルでは保守的になる傾向があるため、実際の有意水準は目標水準より小さくなります。その結果、シミュレートした検出力水準は、サンプルサイズが小さい場合には、より小さくなる傾向があります。

結論として、シミュレーションは、正規近似検定の理論上の検出力関数は正確（尤度比）検定の検出力をよく近似することを示しています。その結果、正規近似検定の理論上の検出力関数は、正確検定で実質的に重要な差を検出するのに十分な検出力を確保するために必要なサンプルサイズを推定するうえで妥当な基準となります。