

# 多重比較の方法

A GRAPHICAL MULTIPLE COMPARISONS PROCEDURE FOR SEVERAL STANDARD DEVIATIONS (複数の標準偏差のグラフによる多重比較手順)

Senin J. Banga and Gregory D. Fox  
2013年6月18日

## 要旨

$k$ 標準偏差の多重比較にグラフを用いる新しい手順を提供する。等分散性検定として、この新しい手順には、Levene (1960) の検定のBrown and Forsythe (1974) によるバージョン  $W_{50}$  と同様のタイプIおよびタイプIIの誤りの特性がある。しかし、多重比較検定に関連するグラフ表示は、異なる標準偏差を持つサンプルのスクリーニングに有用な視覚的ツールとなる。

索引用語: 等分散性、Leveneの検定、Brown-Forsytheの検定、Layardの検定、多重比較

## 1. はじめに

一般に検定  $W_{50}$  と呼ばれる、Leveneの検定 (1960) のBrown and Forsythe (1974) における修正は、おそらく分散の等質性 (同等性) の検定に最も広く使用されている手順の1つです。検定  $W_{50}$  が一般的に使用される理由は、頑健であり、漸近的分布に非依存な方法であるためです。また、他の等分散性検定と比べると、検定  $W_{50}$  は計算が簡単です (これらの検定の比較は、Conover et al. (1981) を参照)。さらに、検定  $W_{50}$  はSAS、Minitab、R、JMPなど、多くの統計ソフトウェアパッケージに含まれているため、簡単に使用できます。

ただし、一部の分布では、特に小さいサンプルの場合、検定  $W_{50}$  の検出力が非常に低くなる可能性があります。たとえば、Pan (1999) は、正規分布を含む一部の分布で、差の大きさに関係なく、検定  $W_{50}$  には2サンプルの標準偏差の比を検出する十分な検出力がない場合があることを示しています。同じ制限が多重サンプル計画にも適用されるのかどうかは、Panの分析からは明らかではありません。3つ以上のサンプルを使用する計画には2サンプル計画よりも多くのデータが含まれる可能性が高いため、そのような計画にこの制限は適用されないことが予想されます。検定  $W_{50}$  には、良好な大サンプル特性があることが知られています (Miller 1968、Brown and Forsythe 1974、Conover et al. 1981)。

有意検定  $W_{50}$  の後に、Bonferroniの多重性補正に基づく同時ペアワイズ比較手順を使用するのが一般的な方法となっています。ただし、Pan (1999) で指摘されているように、2サンプル計画での検定  $W_{50}$  の低検出力により、このような手法は失敗するか、誤った結果が出る可

能性があります。Bonferroniの補正は保守的であるため、特にペアワイズ比較の数が多い場合は、この補正を使用すると問題が悪化します。対照的に、一元配置分散分析の後に平均の比較に使用できる効果的な多重比較手順が数多くあります。例は、Tukey (1953)、Hochberg et al. (1982)、Stoline (1981) を参照してください。サンプル分散の比較には類似の事後分析が役に立ちます。

本書では、複数のサンプルの分散（または標準偏差）を比較するグラフを用いた方法を提案します。分析は、Hochberg et al. (1982) で説明された平均の不確実性区間と類似する、分散の「不確実性区間」に基づきます。最初に、多重ペアワイズ比較手順は、2サンプル計画の等分散性を調べるLayard (1973) の検定のBonett (2006) による修正バージョンに基づきます。ペアワイズ比較で使用される多重性補正は、Nakayama (2009) によって提案された、Tukey-Kramerの方法 (Tukey 1953, Kramer 1956) の大サンプルの一般化に基づきます。「多重比較区間」または「MC区間」と呼ばれる不確実性区間は、Hochberg et al. (1982) で説明された最良の近似手順を使用してペアワイズ比較手順から導出されます。その結果、MC検定では、少なくとも1つのペアのMC区間が重なり合っていないとき、かつそのときに限り、帰無仮説が棄却されます。重なり合っていないMC区間により、有意に異なる分散（または標準偏差）を持つサンプルを識別できます。

MC検定の小サンプル特性を評価するため、シミュレーションの分析を行います。シミュレーションの分析には、比較のために検定 $W_{50}$ も含めます。

## 2. グラフによる多重比較手順

$Y_{i1}, \dots, Y_{in_i}, \dots, Y_{k1}, \dots, Y_{kn_k}$  を  $k$  個の独立サンプルとします。各サンプルは、互いに独立で同じ分布に従い、平均は  $E(Y_{il}) = \mu_i$ 、分散は  $\text{Var}(Y_{il}) = \sigma_i^2 > 0$  です。さらに、サンプルは共通尖度  $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$  を持つ母集団から得ると仮定します。

また、 $\bar{Y}_i$  および  $S_i$  をそれぞれサンプル  $i$  の平均および標準偏差とします。 $m_i$  を調整比率  $1/[2\sqrt{n_i - 4}]$  を使用したサンプル  $i$  の調整平均、 $\hat{\gamma}_{ij}$  を次のように与えられるサンプル  $(i, j)$  の併合尖度推定量とします。

$$\begin{aligned} \hat{\gamma}_{ij} &= (n_i + n_j) \frac{\sum_{l=1}^{n_i} (Y_{il} - m_i)^4 + \sum_{l=1}^{n_j} (Y_{jl} - m_j)^4}{\left[ \sum_{l=1}^{n_i} (Y_{il} - \bar{Y}_i)^2 + \sum_{l=1}^{n_j} (Y_{jl} - \bar{Y}_j)^2 \right]^2} \\ &= (n_i + n_j) \frac{\sum_{l=1}^{n_i} (Y_{il} - m_i)^4 + \sum_{l=1}^{n_j} (Y_{jl} - m_j)^4}{[(n_i - 1)S_i^2 + (n_j - 1)S_j^2]^2} \end{aligned}$$

$\hat{\gamma}_{ij}$  は Layard (1973) の併合尖度推定量と漸近的に等しく、サンプル平均  $\bar{Y}_i$  が調整平均  $m_i$  に置き換えられています。したがって、母集団分散が等しければ、 $\hat{\gamma}_{ij}$  は未知の共通尖度  $\gamma$  の一致推定量です。Bonett (2006) は、2サンプル問題でのLayardの検定の小サンプル性能を向上するため、Layardの併合尖度推定量の代わりにこの推定量を提案しています。本書全体を通して、Layardの検定のBonett (2006) による修正バージョンを単にBonettの検定と呼びます。

比較対象の独立したグループまたはサンプルが3つ以上あるとします ( $k > 2$ )。提案するグラフによる多重比較手順は、Bonettの検定に基づく多重ペアワイズ比較から導出されます。代替の手法は、検定  $W_{50}$  に基づくペアワイズ比較です。ただし、2サンプル計画の場合、正

規分布を含む一部の分布では、検定  $W_{50}$  の検出力性能に問題があります (Pan 1999)。さらに、Banga and Fox (2013) は、Bonnett の検定に基づく分散の比の信頼区間は、一般に検定  $W_{50}$  に基づくものより優れていることを示しています。

サンプルのペア  $(i, j)$  があると仮定すると、有意水準  $\alpha'$  を使用する両側 Bonnet 検定では、次のとき、かつそのときに限り、等分散性の帰無仮説が棄却されます。

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > z_{\alpha'/2} \sqrt{\frac{\hat{y}_{ij} - k_i}{n_i - 1} + \frac{\hat{y}_{ij} - k_j}{n_j - 1}}$$

ここで、 $z_{\alpha'/2}$  は標準正規分布の  $\alpha'/2 \times 100$  番目の上側百分位数点です。

$$k_i = \frac{n_i - 3}{n_i}, k_j = \frac{n_j - 3}{n_j}, c_i = \frac{n_i}{n_i - z_{\alpha/2}}, c_j = \frac{n_j}{n_j - z_{\alpha/2}}$$

複数のペアワイズ比較 (正確な比較数は  $k(k-1)/2$ ) があるため、多重性の調整が必要です。たとえば、目標全体またはファミリーワイズの有意水準  $\alpha$  が与えられると、Bonferroni の補正と呼ばれる一般的な手法では、 $k(k-1)/2$  個のペアワイズ比較それぞれに有意水準  $\alpha' = 2\alpha/(k(k-1))$  が選ばれます。ただし、Bonferroni の補正は、比較対象のサンプル数が増加するにつれ、ペアワイズ比較手順が次第に保守的になることで知られています。より適切な代替の手法は、Nakayama (2009) によって提案され、Tukey-Kramer の方法 (Tukey 1953, Kramer 1956) の大サンプルの近似に基づいています。具体的には、サンプルのあるペア  $(i, j)$  で次が真のとき、かつそのときに限り、多重ペアワイズ比較検定全体が有意になります。

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > \frac{q_{k,\alpha}}{\sqrt{2}} \sqrt{\frac{\hat{y}_{ij} - k_i}{n_i - 1} + \frac{\hat{y}_{ij} - k_j}{n_j - 1}}$$

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  $\alpha$  点です。つまり、 $q_{\alpha,k}$  は次を満たします。

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

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。Barnard (1978) は、正規範囲の分布関数を計算するために、16 点のガウス求積法に基づく単純な数値アルゴリズムを提供しています。

Hochberg et al. (1982) で提案されたように、前述した多重ペアワイズ比較手順を近似するグラフによる多重比較手順では、次のとき、かつそのときに限り、帰無仮説が棄却されません。

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

ここで  $V_i$  は次を最小化するために選択されます。

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

ここで

$$b_{ij} = \sqrt{\frac{\hat{y}_{ij} - k_i}{n_i - 1} + \frac{\hat{y}_{ij} - k_j}{n_j - 1}}$$

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。

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

したがって、この近似手順に基づく等分散性の検定では、次で与えられる区間ペアの少なくとも1つが重なり合っていないとき、かつそのときに限り、帰無仮説が棄却されることになります。

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

グラフによるMC手順では、重なり合わない区間を持つサンプルを視覚的に識別できるように、これらの区間がグラフに表示されます。さらに、分散（または標準偏差）の等質性の全体検定のp値も判断できます。p値を計算するアルゴリズムは、次のセクションで説明します。その前に、MC手順に関する簡単な事実をいくつか紹介します。

#### 注釈

1. サンプルのペア $(i, j)$ に基づく併合尖度推定量 $\hat{\gamma}_{ij}$ は、 $k$ 個すべてのサンプルに基づき、全体の併合尖度推定量に置き換えることができました。この方法だと計算が少し簡略化されますが、ここに記載していないシミュレーション結果は、 $\hat{\gamma}_{ij}$ を使用するほうが良い結果が出ることを示しています。
2. サンプル $i$ に対応する区間は、サンプルの親母集団の標準偏差の信頼区間ではありません。Hochberg et al. (1982) では、このような区間を「不確実性区間」と呼んでいます。ここでは、「多重比較区間」または「MC区間」と呼んでいます。MC区間は、多重サンプル計画の標準偏差または分散を比較する場合にのみ有用です。
3. 本書で説明するMC区間は、3つ以上のサンプルの標準偏差を比較する場合にのみ使用されます。サンプルが2つしかない場合は、比較区間が作成されますが、検定結果と同じ情報が示されます。Banga and Fox (2013) で説明されているように、標準偏差の比の信頼区間を作成するほうが、より多くの情報を得られます。これは、Minitabの2サンプル分散コマンドで使用できます。

## 3. グラフによる多重比較方法のp値

グラフによる（MC）方法のp値を計算するアルゴリズムについて説明する前に、2サンプル計画でLayardの検定のBonett (2006) による修正に関連するp値を導出します。次に、2サンプル計画の結果を多重比較手順に適用する方法を示します。

### 3.1 2サンプル計画のp値

前述したように、2サンプル計画でのLayardの検定のBonett (2006) による調整では、次のとき、かつそのときに限り、等分散性の帰無仮説が棄却されます。

$$|\ln(c_1 S_1^2) - \ln(c_2 S_2^2)| > z_{\alpha/2} se$$

または同等の

$$|\ln(c_{\alpha/2} S_1^2 / S_2^2)| > z_{\alpha/2} se$$

ここで

$$se = \sqrt{\frac{\hat{y}_{12} - k_1}{n_1 - 1} + \frac{\hat{y}_{12} - k_2}{n_2 - 1}}$$

$$c_{\alpha/2} = \frac{c_1}{c_2} = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

Bonettは、小サンプルの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 $c_{\alpha/2}$ を小サンプル調整として導入しました。ただし、定数の効果は大サンプルのアンバランス型計画では無視できる程度で、バランス型計画ではまったく効果がありません。

したがって、計画がバランス型の場合、等分散性の両側検定のp値は単純に次のように計算されることになります。

$$P = 2 \Pr(Z > |Z_0|)$$

ここで

$$Z_0 = \frac{\ln(S_1^2) - \ln(S_2^2)}{se}$$

計画がアンバランス型の場合、 $P = 2 \min(\alpha_L, \alpha_U)$ で、ここで $\alpha_L$ は次の方程式の $\alpha$ の最小の解です。

$$\exp[\ln(c_\alpha S_1^2/S_2^2) - z_\alpha se] = 1 \quad (1)$$

$\alpha_U$ は次の方程式の $\alpha$ の最小の解です。

$$\exp[\ln(c_\alpha S_1^2/S_2^2) + z_\alpha se] = 1 \quad (2)$$

$\alpha_L$ と $\alpha_U$ を求めるアルゴリズムは、以下に示すとおりです。アルゴリズムの数学的な詳細は、「付録」セクションで後述します。

次を仮定します。

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

また、次を仮定します。

$$z_m = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

解 $\alpha_L$ および $\alpha_U$ は次の手順で計算されます。

ケース1:  $n_1 < n_2$

- 上記の結果で与えられたように $z_m$ を計算し、 $L(z_m, n_1, n_2, S_1, S_2)$ を評価します。
- $L(z_m) \leq 0$ の場合、区間 $(-\infty, z_m]$ 内の $L(z, n_1, n_2, S_1, S_2)$ の根 $z_L$ を求め、 $\alpha_L = \Pr(Z > z_L)$ を計算します。
- $L(z_m) > 0$ の場合、関数 $L(z, n_1, n_2, S_1, S_2)$ に根はありません。 $\alpha_L = 0.0$ に設定します。

ケース2:  $n_1 > n_2$

- $L(0, n_1, n_2, S_1, S_2) = \ln S_1^2/S_2^2$ を計算します。

- $L(0, n_1, n_2, S_1, S_2) \geq 0$ の場合、区間 $[0, n_2)$ 内の $L(z, n_1, n_2, S_1, S_2)$ の根 $z_0$ を求めます。それ以外の場合は、区間 $(-\infty, 0)$ 内の根 $z_L$ を求めます。
- $\alpha_L = \Pr(Z > z_L)$ を計算します。

$\alpha_U$ を計算するには、関数 $L(z, n_1, n_2, S_1, S_2)$ の代わりに、関数 $L(z, n_2, n_1, S_2, S_1)$ を使用して、単純に上記の手順を適用します。

## 3.2 グラフによる多重比較のp値

計画に $k$  ( $k > 2$ )個のサンプルがあると仮定し、 $P_{ij}$ をサンプルの任意のペア $(i, j)$ に関連する検定のp値とします。前述のとおり、多重比較検定では、 $k$ 個の比較区間の少なくとも1つのペアが重なり合っていないとき、かつそのときに限り、等分散性の帰無仮説が棄却されます。したがって、多重比較手順に関連する全体の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、次の式を使用して、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、 $V_i$ の定義は前述のとおりです。

$n_i \neq n_j$ の場合、

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

ここで、 $\alpha_L = \Pr(Q > z_L \sqrt{2})$ 、 $\alpha_U = \Pr(Q > z_U \sqrt{2})$ 、 $z_L$ は関数 $L(z, n_i, n_j, S_i, S_j)$ の最小根、 $z_U$ は関数 $L(z, n_j, n_i, S_j, S_i)$ の最小根、および $Q$ はすでに定義済みの確率変数です。数量 $z_L$ と $z_U$ は、前述した2サンプル計画アルゴリズムをサンプルのペア $(i, j)$ に適用することにより求められます。

$n_i = n_j$ の場合、 $P_{ij} = \Pr(Q > |z_0| \sqrt{2})$ です。ここで

$$z_0 = \frac{\ln S_i^2 - \ln S_j^2}{V_i + V_j}$$

## 4. シミュレーションの分析と結果

等分散性を調べる全体検定としてMC検定の小サンプル性能を調べるため、2つの主要なシミュレーションの分析を行います。すべてのシミュレーションは、Mathematicaソフトウェアパッケージのバージョン8を使用して行われました。

### 分析1

最初の分析は、MC検定と検定 $W_{50}$ のタイプIの誤りの特定を評価および比較するために設計されています。3サンプル計画、4サンプル計画、6サンプル計画という3つの計画でさまざまな分布から生成されたサンプルを使用して、2つの検定の性能を比較します。各設計のサンプルサイズは10~50で、10単位で増分します。サンプルは次の親分布から抽出されます。

- 正規分布
- 一様分布、および $(3, 3)$ のパラメータを使用したベータ分布で表される、対称的で裾の軽い分布

- 自由度5のt分布 ( $t(5)$ )、およびラプラス分布で表される、対称的で裾の重い分布
- 指数分布、自由度1のカイ二乗分布 ( $\chi^2(1)$ )、および自由度5のカイ二乗分布 ( $\chi^2(5)$ ) で表される歪んだ裾の重い分布
- 観測値の90%は標準正規分布から、残りの10%は平均が0、標準偏差が3の正規分布から抽出される混合正規分布 (CN(0.9, 3))

各シミュレーションは、10,000個のサンプル反復で構成されます。目標とする名目 $\alpha$ 水準は、0.05です。シミュレーション誤差は約0.002です。表1に、各検定のシミュレートした有意水準を示します。

表1 シミュレートした有意水準の比較( $\alpha = 0.05$ )

説明	分布 [尖度]	$n_i$	$k = 3$		$k = 4$		$k = 6$	
			MC	$W_{50}$	MC	$W_{50}$	MC	$W_{50}$
正規	正規 [3.0]	10	.038	.033	.038	.031	.036	.029
		20	.039	.038	.040	.038	.041	.033
		30	.043	.041	.044	.038	.046	.039
		40	.046	.043	.046	.041	.048	.041
		50	.046	.046	.046	.044	.052	.047
対称的で裾が軽い	一様 [1.8]	10	.029	.029	.025	.024	.023	.020
		20	.028	.026	.030	.026	.028	.023
		30	.037	.035	.034	.032	.034	.030
		40	.038	.037	.037	.037	.035	.033
		50	.041	.041	.036	.036	.036	.036
	Beta(3, 3) [2.5]	10	.031	.032	.031	.029	.031	.025
		20	.035	.031	.036	.027	.037	.026
		30	.041	.035	.037	.034	.037	.032
		40	.040	.036	.039	.035	.040	.033
		50	.044	.039	.044	.037	.044	.035
対称的で裾が重い	ラプラス [6.0]	10	.056	.038	.063	.041	.071	.039
		20	.054	.044	.058	.043	.059	.041
		30	.051	.042	.053	.043	.052	.044
		40	.048	.045	.048	.045	.048	.046
		50	.045	.045	.051	.046	.049	.047
	$t(5)$	10	.042	.032	.044	.031	.042	.031

説明	分布 [尖度]	$n_i$	$k = 3$		$k = 4$		$k = 6$	
			MC	$W_{50}$	MC	$W_{50}$	MC	$W_{50}$
	[9.0]	20	.043	.039	.045	.038	.045	.040
		30	.039	.040	.040	.040	.041	.040
		40	.041	.042	.040	.041	.039	.038
		50	.040	.050	.039	.046	.038	.046
歪んでいて裾が重い	$\chi^2(5)$ [5.4]	10	.040	.039	.046	.040	.048	.039
		20	.040	.043	.040	.040	.042	.039
		30	.039	.047	.042	.044	.043	.042
		40	.040	.046	.041	.044	.039	.042
		50	.037	.047	.038	.047	.040	.048
	指数 [9.0]	10	.063	.051	.073	.049	.076	.048
		20	.051	.049	.053	.048	.057	.046
		30	.042	.048	.046	.051	.049	.049
		40	.034	.050	.038	.046	.037	.049
		50	.033	.045	.037	.047	.038	.046
	$\chi^2(1)$ [15.0]	10	.084	.048	.098	.050	.118	.050
		20	.053	.046	.060	.047	.068	.046
		30	.041	.041	.045	.045	.050	.047
		40	.044	.049	.046	.047	.045	.047
		50	.038	.050	.037	.049	.040	.049
混合正規	CN(0.9, 3) [8.3]	10	.020	.016	.018	.012	.016	.010
		20	.014	.015	.012	.013	.008	.007
		30	.012	.014	.010	.011	.007	.008
		40	.009	.017	.009	.014	.006	.008
		50	.009	.016	.007	.012	.006	.009

結果は、両方の検定がほとんどの分布で良好な性能を示すことを表しています。シミュレートした有意水準のほとんどが目標の0.05に近くなっています。ただし、小さいサンプルが正規分布および対称的で裾の軽い分布から抽出される場合、シミュレートした有意水準は両方の検定で保守的（0.05未満）になる傾向があります。これらの分布では、MC検定のシミュレートした有意水準は、検定 $W_{50}$ よりも目標有意水準に近くなります。



小さいサンプルが裾の重い分布から抽出される場合、検定 $W_{50}$ は保守的に、MC検定はリベラルになる傾向があります。MC検定は、小さいサンプルが極端に歪んだ分布から抽出される場合、さらにリベラルになります。たとえば、サイズ10のサンプルが自由度1のカイ二乗分布から抽出される場合、MC検定のシミュレートした有意水準は3サンプル、4サンプル、および6サンプルの計画で、それぞれ0.084、0.098、0.118です。

どちらの検定も外れ値による影響を受けています。混合正規分布の有意水準は、サンプルサイズが50と大きい場合でも、極端に保守的です。

## 分析2

2つ目の分析では、4サンプル計画で2つの手順のタイプIIの誤りの特性（検出力）を評価および比較します。この分析では、分析1のサンプルサイズ20および $k = 4$ という条件で使用したのと同じサンプルを使用します。観測値は、1、2、3、または4倍にスケールされます。たとえば、1:1:4:4で表される条件では、サンプル1と2の観測値は分析1で使用されたものと同じです。サンプル3と4の観測値は4倍にスケールされます。

比較のために、1:1:1:1の条件を含めます。この条件の結果は、分析1のサイズ20および $k = 4$ というサンプルで報告されたものと同じであることに注目してください。サイズ20のサンプルを選んだのは、サイズ20のサンプルを使用した場合、どちらの検定でも、ほとんどの分布で目標水準に近い達成有意水準が得られることを、分析1の結果が示唆しているためです。

これらの実験のシミュレートした有意水準は、等分散性の帰無仮説の棄却につながるサンプル反復の比率として計算されます。

表2に結果を示します。

表2 シミュレートした検出力水準の比較( $\alpha = 0.05$ )

説明	分布	標準偏差比							
		1:1:1:1		1:1:2:2		1:2:3:4		1:1:4:4	
		MC	$W_{50}$	MC	$W_{50}$	MC	$W_{50}$	MC	$W_{50}$
	正規	.040	.038	.846	.853	.998	.994	1.000	1.000
対称的で裾が軽い	一様	.030	.026	.985	.962	1.000	.999	1.000	1.000
	Beta(3, 3)	.036	.027	.938	.916	1.000	.999	1.000	1.000
対称的で裾が重い	ラプラス	.058	.043	.597	.629	.931	.921	.996	.998
	$t(5)$	.045	.038	.657	.703	.952	.949	.997	.998
歪んでいて裾が重い	$\chi^2(5)$	.040	.040	.625	.704	.949	.949	.996	.999
	指数	.053	.048	.431	.507	.804	.779	.963	.978
	$\chi^2(1)$	.060	.047	.298	.291	.602	.504	.838	.824
混合	CN(0.9, 3)	.012	.013	.499	.612	.889	.917	.989	.998

この結果はMC検定と検定 $W_{50}$ のタイプIIの誤りの特性（検出力）は同様であることを示唆しています。一般に、両方の検定で達成されるシミュレートした検出力水準は、同程度のオーダーになります。2つの検定の検出力が0.1以上異なるケースは1つのみです。

MC検定のシミュレートした検出力水準は、軽い裾から中程度の裾を持つ対称的な分布からサンプルが抽出される場合、検定 $W_{50}$ より少し良くなります。一方、検定 $W_{50}$ は、裾の重い分布から抽出されるサンプルが場合、MC検定より少し強力になるように見えます。

## 5. 例

このセクションでは、グラフによるMC手順と検定 $W_{50}$ をOtt et al. (2010) の397ページから取得したデータセットに適用します。 データは次のように説明されています。

ある鑄造会社に、蠟型に流し込む前に原材料を加熱する窯がいくつかあります。これらの金属をほとんど変動がない精密な温度にまで加熱することが非常に重要です。3つの窯がランダムに選択され、連続する10回の加熱で、非常に正確に温度(°C)が記録されます。収集されたデータは、次のとおりです。

窯1	1670.87	1670.88	1671.51	1672.01	1669.63	1670.95	1668.70	1671.86	1669.12	1672.52
窯2	1669.16	1669.60	1669.76	1669.18	1671.92	1669.69	1669.45	1669.35	1671.89	1673.45
窯3	1673.08	1672.75	1675.14	1674.94	1671.33	1660.38	1679.94	1660.51	1668.78	1664.32

図1に、各窯の温度の箱ひげ図を示します。この箱ひげ図は、記録された温度には外れ値がないこと、および窯3の温度の変動性は窯1または窯2とは異なることを示唆しています。

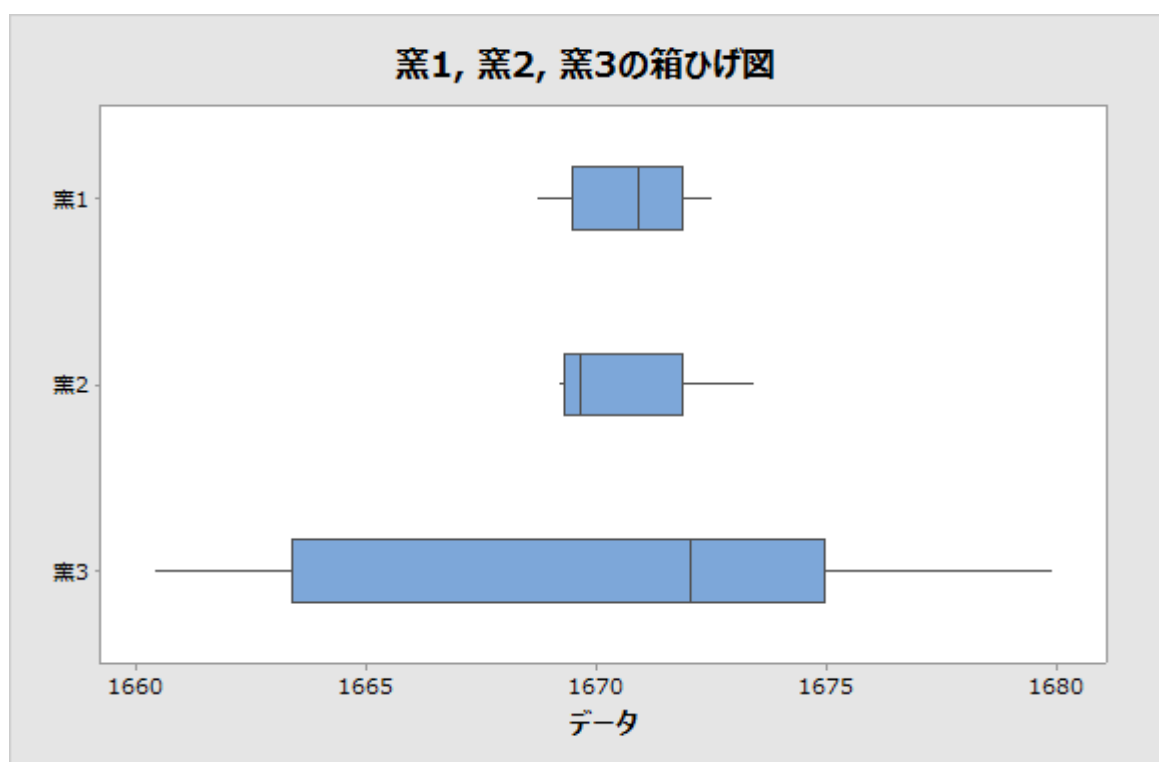


図1 窯の温度の箱ひげ図 (°C)

図2は、同じデータのMC区間、および全体MC検定と検定 $W_{50}$ （凡例にはLeveneの検定と表記）の結果を示しています。両方の検定の有意なp値は、温度の変動性が3つの窯で異なることを示しています。重なり合っていないMC区間から、窯1の変動性が窯2または窯1とは異なることが確認できます。窯1、窯2、窯3のMC区間は、それぞれ(0.896, 2.378)、(1.072, 2.760)、(4.366, 12.787)です。

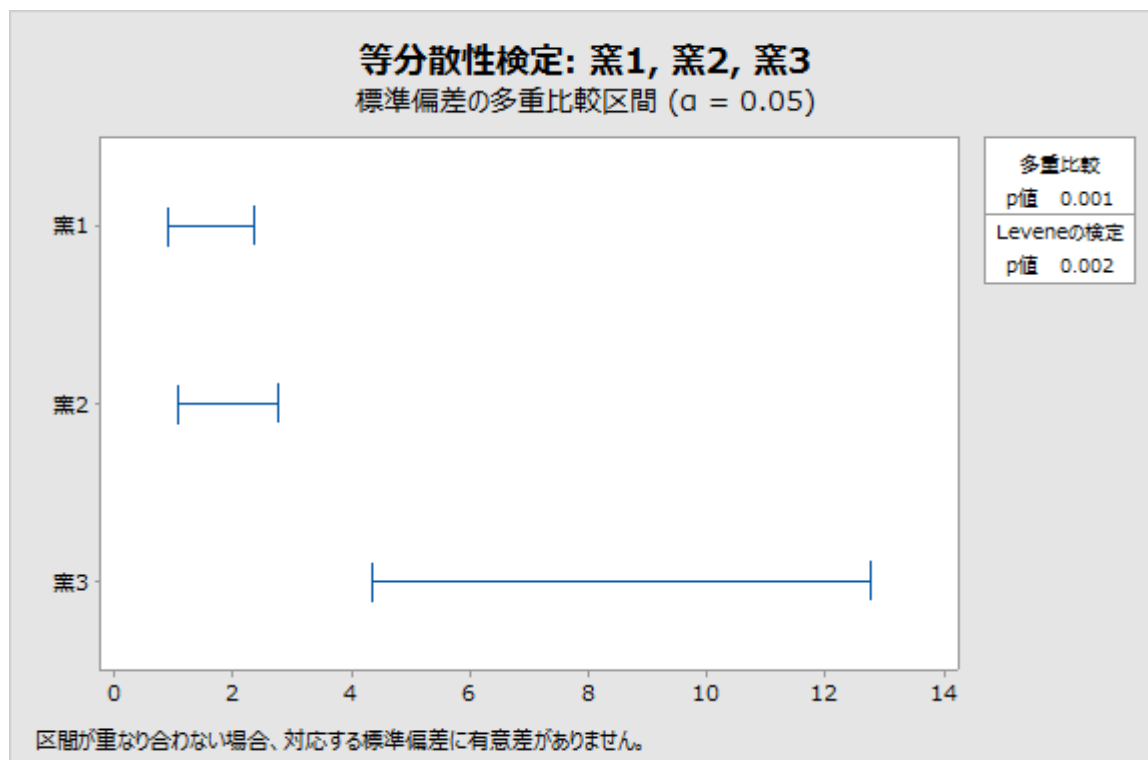


図2 MC検定と検定 $W_{50}$ （Leveneの検定）のMC区間とp値

## 6. 結論

全体的に、シミュレーションの結果は、複数の小さいサンプルを使用する計画の場合、MC検定の性能は検定 $W_{50}$ の性能と同様であることを示しています。MC検定は軽い裾から中程度の裾を持つ対称的な分布またはほぼ対称的な分布により適しているのに対し、検定 $W_{50}$ は大きく歪んだ分布や裾の重い分布からデータが抽出される場合に適していると考えられます。MC手順の明らかな利点の1つは、標準偏差の等質性の全体検定が有意である場合に、異なる標準偏差または分散を持つサンプルのスクリーニングに使用できる強力な視覚的ツールになるということです。グラフによるMC手順はMinitabのリリース17で使用できます。

## 7. 付録

2サンプル計画でのLayardの検定のBonett（2006）による調整では、次のとき、かつそのときに限り、等分散性の帰無仮説が棄却されます。

$$|\ln(c_1 S_1^2) - \ln(c_2 S_2^2)| > z_{\alpha/2} se$$

または同等の

$$|\ln(c_{\alpha/2} S_1^2 / S_2^2)| > z_{\alpha/2} se$$

ここで

$$se = \sqrt{\frac{\hat{\gamma}_{12} - k_1}{n_1 - 1} + \frac{\hat{\gamma}_{12} - k_2}{n_2 - 1}}$$

$$c_{\alpha/2} = \frac{c_1}{c_2} = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

したがって、計画がバランス型の場合、 $c_{\alpha/2} = 1$ で、検定のp値は単純に次のように求められます。

$$P = 2 \Pr(Z > |Z_0|)$$

ここで

$$Z_0 = \frac{\ln(S_1^2) - \ln(S_2^2)}{se}$$

計画がアンバランス型の場合、 $P = 2 \min(\alpha_L, \alpha_U)$ で、ここで

$\alpha_L$ は次の方程式の $\alpha$ の最小の解です。

$$\exp[\ln(c_\alpha S_1^2/S_2^2) - z_\alpha se] = 1 \quad (1)$$

また $\alpha_U$ は次の方程式の最小の解 $\alpha$ です。

$$\exp[\ln(c_\alpha S_1^2/S_2^2) + z_\alpha se] = 1 \quad (2)$$

これらの方程式を $\alpha$ について解く方法は、最初に $z \equiv z_\alpha$ について方程式を解き、次に $\alpha = \Pr(Z > z)$ を求めます。ここで、確率変数 $Z$ は、標準正規分布に従います。これらの方程式を解く方法を説明する前に、方程式(1)は方程式 $L(z) = 0$ として、書き換えることができます。ここで

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

同様に、方程式(2)は方程式 $U(z) = 0$ と同等です。ここで

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

$L(z, n_2, n_1, S_2, S_1) = -U(z, n_1, n_2, S_1, S_2)$ であることに留意します。その結果、2つの関数の一方のみの根を求める必要があります。

方程式(1)または(2)を解くアルゴリズムは、次の結果から導出されます。

## 結果

$n_1, n_2, S_1$ と $S_2$ が任意で、固定されているとします。アンバランス型計画では、関数 $L(z, n_1, n_2, S_1, S_2)$ の根の数は多くても2つです。

4.  $n_1 < n_2$ の場合、 $L(z, n_1, n_2, S_1, S_2)$ は凸です。これは $L(-\infty, n_1, n_2, S_1, S_2) = L(n_1, n_1, n_2, S_1, S_2) = +\infty$ を満たし、次で最小値に達します。

$$z_m = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

したがって、 $L(z_m, n_1, n_2, S_1, S_2) \leq 0$  の場合、根は2つあります。1つは区間 $(-\infty, z_m]$ 、もう1つは区間 $[z_m, n_1)$ にあります。一方、 $L(z_m, n_1, n_2, S_1, S_2) > 0$  の場合、関数  $L(z, n_1, n_2, S_1, S_2)$  に根はありません。

5.  $n_1 > n_2$  の場合、 $L(z, n_1, n_2, S_1, S_2)$  は $+\infty$ から $-\infty$ へ単調に減少するため、一意な根があります。 $L(0, n_1, n_2, S_1, S_2) = \ln S_1^2/S_2^2 \geq 0$  の場合、根は区間 $[0, n_2)$ にあり、それ以外の場合は、根は区間 $(-\infty, 0)$ にあります。

## 証明

以下では、 $L(z) \equiv L(z, n_1, n_2, S_1, S_2)$  とします。

最初に、 $n_1 < n_2$  の場合、 $L(z)$  が凸で、次で最小値に達することを証明します。

$$z_m = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

すでに定義したように、

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

です。次に、 $\lim_{z \rightarrow -\infty} L(z) = +\infty$  および

$$\lim_{z \rightarrow \min(n_1, n_2)} L(z) = \begin{cases} +\infty & (n_1 < n_2) \\ -\infty & (n_2 < n_1) \end{cases}$$

です。また、 $L(z)$  の導関数は

$$-\frac{(n_1 - z)(n_2 - z)}{se} L'(z) = z^2 - (n_1 + n_2)z + n_1 n_2 + \frac{n_1 - n_2}{se}$$

を満たすという点に留意します。

$$Q(z) = -\frac{(n_1 - z)(n_2 - z)}{se} L'(z)$$

とします。 $n_1 < n_2$  の場合、2次式 $Q(z)$ には次のように与えられる2つの根があります。

$$z_1 = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

および

$$z_2 = \frac{n_1 + n_2 + \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

$Q(n_1) = \frac{n_1 - n_2}{se} < 0$  の場合、 $(-\infty, z_1)$  の $z$ は $Q(z) > 0$ 、 $(z_1, n_1)$  の $z$ は $Q(z) < 0$  となるような  $z_1 < n_1 = \min(n_1, n_2) < z_2$  が得られます。したがって、 $(-\infty, z_1)$  の $z$ は $L'(z) < 0$ 、 $(z_1, n_1)$  の $z$ は $L'(z) > 0$  ということになります。ゆえに、 $L(z)$  は領域 $(-\infty, \min(n_1, n_2))$  上で凸で、 $z_1 \equiv z_m$  で最小値に達します。

$n_1 > n_2$  の場合、2つのケースがあります。 $n_1 - n_2 > 4/se$  であるケースと

$0 < n_1 - n_2 < 4/se$ であるケースです。1番目のケースでは、 $z_1$ と $z_2$ は、 $n_2 = \min(n_1, n_2) < z_1 < z_2$ となるような $Q(z)$ の根です ( $n_2 - \frac{z_1+z_2}{2} = \frac{n_2-n_1}{2} < 0$ であるため)。ゆえに、領域 $(-\infty, \min(n_1, n_2))$ の $z$ は $Q(z) > 0$ です。2番目のケースでは、 $Q(z)$ には領域上で $Q(z) > 0$ となるような根はありません。

したがって、 $n_1 > n_2$ の場合、 $L(z)$ が $+\infty$ から $-\infty$ へ単調に減少するような $L'(z) < 0$ です。

## 8. 参考文献

- Banga, S. J. and Fox, G. D. (2013). On Bonett's Robust Confidence Interval for a Ratio of Standard Deviations. In press.
- Barnard, J. (1978). Probability Integral of the Normal Range. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 27, 197-198.
- Bonett, D. G. (2006). Robust Confidence Interval for a Ratio of Standard Deviations. *Applied Psychological Measurements*, 30, 432-439.
- Brown, M. B., and 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. and 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.
- Hochberg, Y., Weiss, G., and Hart S. (1982). On Graphical Procedures for Multiple Comparisons. *Journal of the American Statistical Association*, 77, 767-772.
- Kramer, C. Y. (1956). Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications. *Biometrics*, 12, 307-310.
- 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., *Contributions to Probability and Statistics*, Palo Alto, CA: Stanford University Press, 278-292.
- Miller, R. G. (1968). Jackknifing Variances. *Annals of Mathematical Statistics*, 39, 567-582.
- Nakayama, M. K. (2009). Asymptotically Valid Single-Stage Multiple-Comparison Procedures. *Journal of Statistical Planning and Inference*, 139, 1348-1356.
- Ott, R. L. and Longnecker, M. (2010). *An introduction to Statistical Methods and Data Analysis, sixth edition*, Brooks/Cole, Cengage Learning.
- Pan, G. (1999). On a Levene Type Test for Equality of Two Variances. *Journal of Statistical Computation and Simulation*, 63, 59-71.
- Stoline, M. R. (1981). The Status of Multiple of Comparisons: Simultaneous Estimation of All Pairwise Comparisons in One-Way ANOVA Designs. *The American Statistician*, 35, 134-141.
- Tukey, J. W. (1953). *The Problem of Multiple Comparisons*. Mimeographed monograph.

Wolfram, S. (1999). *The Mathematica Book*, 4th ed. Wolfram Media/Cambridge University Press.

© 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](http://www.minitab.com). All other marks referenced remain the property of their respective owners.