

多重比较方法

多个标准差的图形化多重比较过程

Senin J. Banga and Gregory D. Fox
June 18, 2013

摘要

为 k 个标准差提供多重比较的全新图形化过程。作为一种方差同质性检验，此全新过程具有与 Levene (1960) 检验 W_{50} 的 Brown 和 Forsythe (1974) 版本类似的类型 I 和类型 II 误差属性。但是，与多重比较检验关联的图形显示可提供用于筛选具有不同标准差的样本的有用的可视化工具。

索引项: 方差同质性, Levene 检验, Brown-Forsythe 检验, Layard 检验, 多重比较

1. 简介

Levene 检验 (1960) 的 Brown 和 Forsythe (1974) 修改通常指检验 W_{50} ，它可能是使用最广泛的方差同质性（等同性）检验过程之一。检验 W_{50} 的广泛应用部分归因于其稳健性，并且不呈渐进式分布。与其他方差同质性检验相比，检验 W_{50} 还易于计算。（要进行此类比较，请参见 Conover 等人，(1981)）。此外，检验 W_{50} 还方便操作，可用于 SAS、Minitab、R 和 JMP 等诸多统计软件包。

但是，对于某些分布，检验 W_{50} 的功效可能非常低，特别是当样本不大时。例如，Pan (1999) 指出，对于正态分布等某些分布，检验 W_{50} 可能功效不足，无法检测到两个标准差之间的差值（与差值的量值无关）。Pan 分析不能确定相同的限制是否适用于多样本设计。有人可能认为此限制可能不适用于具有两个以上的样本的设计，因为此类设计包含的数据可能不止两个样本设计。我们已知 W_{50} 具有比较大的样本属性（Miller, 1968; Brown 和 Forsythe, 1974; Conover 等人, 1981）。

它已经成为遵照基于 Bonferroni 多重性校正的同步配对比较过程进行显著性检验 W_{50} 的常用做法。但是，正如 Pan (1999) 所指出的那样，此方法可能失败，或产生会引起误导的结果，原因是双样本设计中的检验 W_{50} 的功效较低。使用 Bonferroni 校正会使问题变得更严重，因为此校正比较保守，尤其是在配对比较的数量很大时。相比之下，许多高效的多重比较过程可用于根据单因子方差分析比较均值。例如，请参见 Tukey (1953)、Hochberg 等人 (1982) 和 Stoline (1981)。用于比较样本方差的类似因果分析很有用。

在本白皮书中，我们建议使用图形化方法来比较多个样本的方差（或标准差）。此分析基于方差的“不确定性区间”，该区间类似于 Hochberg 等人（1982）针对均值描述的不确定性区间。首先，多重配对比较过程基于双样本设计中方差等同性的 Layard（1973）检验的 Bonett（2006）修订版本。在配对比较中使用的多重性校正基于 Nakayama（2009）建议的 Tukey-Kramer 法（Tukey, 1953; Kramer, 1956）的大样本广义化。不确定性区间（我们称为“多重比较区间”或“MC 区间”）使用 Hochberg 等人（1982）描述的最佳近似过程，根据配对比较过程推导而出。当且仅当至少有一对 MC 区间不重叠时，最后的 MC 检验结果将否定原假设。非重叠 MC 区间可确定方差（或标准差）具有显著差异的样本。

我们将通过模拟研究来评估 MC 检验的小样本属性。为便于比较，我们还将模拟研究中增加检验 W_{50} 。

2. 图形化多重比较过程

设 $Y_{i1}, \dots, Y_{in_i}, \dots, Y_{k1}, \dots, Y_{kn_k}$ 为 k 独立样本，每个样本都各自独立，分布方式也相同，并且均值为 $E(Y_{il}) = \mu_i$ ，方差为 $\text{Var}(Y_{il}) = \sigma_i^2 > 0$ 。此外，假设样本源自常用峰度为 $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$ 的总体。

另外，设 \bar{Y}_i 和 S_i 分别为样本 i 的均值和标准差。设 m_i 为样本 i 的调整均值，调整比率为 $1/[2\sqrt{n_i-4}]$ ，设 $\hat{\gamma}_{ij}$ 为样本 (i, j) 的合并峰度估计量，按如下给定

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

请注意， $\hat{\gamma}_{ij}$ 渐进等效于 Layard 的（1973）合并峰度估计量，其中，样本均值 \bar{Y}_i 已被调整均值 m_i 替换。因此，只要总体方差相等， $\hat{\gamma}_{ij}$ 就是未知常用峰度 γ 的一致估计量。Bonett（2006）建议用此估计量代替 Layard 的合并峰度估计量，以提高双样本问题中 Layard 检验的小样本性能。在本白皮书中，Bonett（2006）的 Layard 检验的修订版本简称 Bonett 检验。

假设有两个以上的独立组或样本需要比较（ $k > 2$ ）。我们建议的图形化多重比较过程根据基于 Bonett 检验的多重配对比较推导而出。备选方法是基于检验 W_{50} 进行配对比较。但是，在双样本设计中，对于正态分布（Pan, 1999）等某些分布，检验 W_{50} 的功效性能可能存在问题。此外，Banga 和 Fox（2013）指出，基于 Bonett 检验的方差比率置信区间通常优于基于检验 W_{50} 的方差比率置信区间。

如果给定任何样本对 (i, j) ，当且仅当满足以下条件时，显著性水平为 α' 的双侧 Bonett 检验才会否定等同性原假设，

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

其中, $z_{\alpha'/2}$ 是标准正态分布的第 $\alpha'/2 \times 100$ 个上四分位点。

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

由于存在多重配对比较, 即精确的 $k(k-1)/2$ 比较, 因此需要进行多重性调整。例如, 如果给定了目标整体或整族显著性水平 α , 则可使用一种常用的方法, 即 Bonferroni 校正来选择每个 $k(k-1)/2$ 配对比较 $\alpha' = 2\alpha/(k(k-1))$ 的显著性水平。但是, 我们也知道, 随着要比较的样本数量增大, Bonferroni 校正会导致配对比较过程越来越保守。Nakayama (2009) 推荐了一种更好的备选方法, 并且该方法基于 Tukey-Kramer 法 (Tukey, 1953; Kramer, 1956) 的大样本近似之上。尤其是, 当且仅当对于某些样本对 (i, j) 以下条件为真时, 整体多重配对比较检验的效果显著:

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

其中, $q_{\alpha,k}$ 是一系列 k 独立和相同分布的标准正态随机变量的上限 α 点。即, $q_{\alpha,k}$ 满足

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

其中, Z_1, \dots, Z_k 是独立和相同分布的标准正态随机变量。Barnard (1978) 基于用于计算正常范围的分函数的 16 点 Gaussian 求积提供了一种简单的数值算法。

根据 Hochberg 等人 (1982) 的建议, 当且仅当满足以下条件时, 接近上述多重配对比较过程的图形化多重比较过程会否定原假设,

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

其中, V_i 可选择用于对以下方程求最小值:

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

其中

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

Hochberg 等人 (1982) 中所述的此问题的解决方法是选择

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

它遵循以下原则: 当且仅当以下给定的区间中至少有一对区间不重叠时, 基于此近似过程的方差同质性检验才会否定原假设。

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

图形化的 MC 过程包括在图形上显示这些区间, 以便能够直观地找出具有非重叠区间的样本。此外, 还可以确定方差 (或标准差) 同质性的整体检验的 p 值。我们在下一部分中提供用于计算 p 值的详细算法。不过, 我们首先会举几个有关 MC 过程的简单实例。

备注

1. 基于样本对 (i, j) 的合并峰度估计量 $\hat{\gamma}_{ij}$ 已被基于所有 k 样本的整体合并峰度估计量替换。虽然此方法可以在某种程度上简化计算，但此处未显示的模拟结果指出，使用 $\hat{\gamma}_{ij}$ 可以获得更好的结果。
2. 样本 i 对应的区间不是样本父总体的标准差的置信区间。Hochberg 等人 (1982) 将此类区间称作“不确定性区间”。我们将其称作“多重比较区间”或“MC 区间”。MC 区间仅用于比较多样本设计的标准差或方差。
3. 本白皮书中描述的 MC 区间只能用于比较两个以上的标准差。当只有两个样本时，可以构造比较区间，但它们传达的信息与检验结果提供的信息相同。相比之下，构造标准差比率的置信区间（如 Banga 和 Fox (2013) 描述的区间和 Minitab 双样本方差命令提供的区间）所提供的信息丰富得多。

3. 图形化多重比较方法的 P 值

在我们描述用于计算图形化 (MC) 方法的 p 值的算法时，我们首先推导出与双样本设计中 Bonett (2006) 的 Layard 检验的修改相关的 p 值。然后，我们介绍如何将双样本设计结果应用于多重比较过程。

3.1 双样本设计中的 P 值

如前所述，当且仅当满足以下条件时，双样本设计中的 Bonett (2006) 的 Layard 检验调整将否定方差同质性的原假设，

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

或等效情况

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

其中

$$se = \sqrt{\frac{\hat{\gamma}_{12} - k_1}{n_1 - 1} + \frac{\hat{\gamma}_{12} - k_2}{n_2 - 1}}$$
$$c_{\alpha/2} = \frac{c_1}{c_2} = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

Bonett 引入常数 $c_{\alpha/2}$ 作为小样本调整，以降低小样本不平衡设计中不等尾误差概率的影响。但是，此常数的影响在大样本不平衡设计中可以忽略，并且，此常数对平衡设计不产生任何影响。

它遵循以下原则，如果设计是平衡设计，则方差同质性的双侧检验的 p 值可以简单地计算为

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

其中

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

如果此设计是不平衡设计，则 $P = 2 \min(\alpha_L, \alpha_U)$ ，其中， α_L 是此方程中 α 的最小解。

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

并且, α_U 是此方程中 α 的最小解。

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

用于查找 α_L 和 α_U 的算法按如下给定。此算法的数学理论将在后文的附录部分详述。

设

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

另设

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

解 α_L 和 α_U 按以下步骤计算得到:

实例 1: $n_1 < n_2$

- 计算在上述结果中给定的 z_m , 然后评估 $L(z_m, n_1, n_2, S_1, S_2)$ 。
- 如果 $L(z_m) \leq 0$, 则在区间 $(-\infty, z_m]$ 中找到 $L(z, n_1, n_2, S_1, S_2)$ 的根 z_L , 并计算 $\alpha_L = \Pr(Z > z_L)$ 。
- 如果 $L(z_m) > 0$, 则函数 $L(z, n_1, n_2, S_1, S_2)$ 没有根。设 $\alpha_L = 0.0$ 。

实例 2: $n_1 > n_2$

- 计算 $L(0, n_1, n_2, S_1, S_2) = \ln S_1^2/S_2^2$ 。
- 如果 $L(0, n_1, n_2, S_1, S_2) \geq 0$, 则在区间 $[0, n_2)$ 中找到 $L(z, n_1, n_2, S_1, S_2)$ 的根 z_o , 否则, 在区间 $(-\infty, 0)$ 中找到根 z_L 。
- 计算 $\alpha_L = \Pr(Z > z_L)$ 。

要计算 α_U , 我们只要使用函数 $L(z, n_2, n_1, S_2, S_1)$ 而不是函数 $L(z, n_1, n_2, S_1, S_2)$ 来应用上述步骤。

3.2 图形化多重比较的 P 值

假设设计中有 $k (k > 2)$ 个样本, 我们设 P_{ij} 为与任何样本对 (i, j) 关联的检验的 p 值。我们回想一下, 当且仅当至少有一对 k 比较区间不重叠时, 多重比较检验就会否定方差同质性的原假设。它遵循以下原则: 与多重比较过程关联的整体 p 值是

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

为计算 P_{ij} , 我们通过使用以下方法, 来执行双样本设计的算法:

$$se = V_i + V_j$$

其中, V_i 按以前那样定义。

如果 $n_i \neq n_j$, 则

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

其中, $\alpha_L = \Pr(Q > z_L\sqrt{2})$ 、 $\alpha_U = \Pr(Q > z_U\sqrt{2})$ 、 z_L 是函数 $L(z, n_i, n_j, S_i, S_j)$ 的最小根, z_U 是函数 $L(z, n_j, n_i, S_j, S_i)$ 的最小根, Q 是以前定义的随机变量。可以通过将以前描述的双样本设计算法应用到样本对 (i, j) 来找到数量 z_L 和 z_U 。

如果 $n_i = n_j$, 则 $P_{ij} = \Pr(Q > |z_o|\sqrt{2})$, 其中

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

4. 模拟研究和结果

我们进行了两项主要的模拟研究, 以调查 MC 检验 (作为方差同质性的整体检验) 的小样本性能。所有的模拟都采用 Mathematica 软件包版本 8 来进行。

研究 1

第一项研究旨在评估和比较 MC 检验和检验 W_{50} 的类型 I 误差属性。我们采用从以下三种不同设计中的各种分布生成的样本比较了这两种检验的性能: 3 样本设计、4 样本设计和 6 样本设计。在每种设计中, 样本数量从 10 到 50 不等, 增量为 10。样本抽取自以下父分布:

- 正态分布
- 对称轻尾分布, 用均匀分布和 Beta 分布 (这两个分布的参数均为 (3, 3)) 表示
- 对称重尾分布, 用自由度为 5 ($t(5)$) 的 t 分布和 Laplace 分布表示
- 偏斜重尾分布, 用指数分布、自由度为 1 的卡方分布 ($\chi^2(1)$) 和自由度为 5 的卡方分布 ($\chi^2(5)$) 表示
- 90% 的观测值抽取自标准正态分布, 剩余的 10% 抽取自均值为 0 且标准差为 3 的正态分布的污染正态分布 (CN(0.9, 3))。

每项模拟研究都有 10,000 个取样仿行。目标名义 α 水平为 0.05。模拟误差大约 0.002。表 1 中报告了每个检验的模拟显著性水平。

表 1 模拟显著性水平 ($\alpha = 0.05$) 比较

说明	分布 [峰度]	n_i	$k = 3$		$k = 4$		$k = 6$	
			MC	W_{50}	MC	W_{50}	MC	W_{50}
正态	正态 [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

说明	分布 [峰度]	n_i	$k = 3$		$k = 4$		$k = 6$	
			MC	W_{50}	MC	W_{50}	MC	W_{50}
对称轻尾	均匀 [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
对称重尾	Laplace [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
	$t(5)$ [9.0]	10	.042	.032	.044	.031	.042	.031
		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
偏态重尾	$\chi^2(5)$ [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

说明	分布 [峰度]	n_i	$k = 3$		$k = 4$		$k = 6$	
			MC	W_{50}	MC	W_{50}	MC	W_{50}
	$\chi^2(1)$ [15.0]	50	.033	.045	.037	.047	.038	.046
		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
污染正态	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

结果表明，对于大多数分布，这两个检验的执行效果不错。大多数模拟显著性水平均接近目标值 0.05。但是，当小样本抽取自对称轻尾正态分布时，这两个检验的模拟显著性水平通常比较保守（低于 0.05）。对于这些分布，MC 检验的模拟显著性水平比检验 W_{50} 的显著性水平更接近目标显著性水平。

当小样本抽取自重尾分布时，检验 W_{50} 通常比较保守，MC 检验通常比较宽松。当小样本抽取自极度偏态分布时，MC 检验会更宽松。例如，当从自由度为 1 的卡方分布中抽取 10 个样本后，对于 3 样本设计、4 样本设计和 6 样本设计，MC 检验的模拟显著性水平分别为 0.084、0.098 和 0.118。

这两个检验均受异常值影响。污染正态分布的显著性水平极度保守，即使在样本数量大到 50 时也是如此。

研究 2

第二项研究评估和比较 4 样本设计中两个过程的类型 II 误差属性（功效）。我们对此研究使用的样本与研究 1 中数量为 20 并在 $k = 4$ 条件下使用的样本相同。观测值按因子 1、2、3 或 4 进行缩放。例如，在表示为 1:1:4:4 的条件下，样本 1 和 2 的观测值与研究 1 中使用的观测值相同。样本 3 和 4 中的观测值按因子 4 进行缩放。

我们按 1:1:1:1 条件进行比较。请注意，此条件的结果与研究 1 中报告的数量为 20 和 $k = 4$ 的样本的结果相同。我们选择了数量为 20 的样本，因为研究 1 的结果表明，对于这两个检验，数量为 20 的样本可实现接近于大多数分布的目标显著性水平的显著性水平。

这些试验中的模拟功效水平计算为样本仿行的比率，从而会导致否定方差同质性的原假设。

表 2 中报告了这些结果。

表 2 模拟功效水平 ($\alpha = 0.05$) 比较

说明	分布	标准差比率							
		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}
	正态	.040	.038	.846	.853	.998	.994	1.00	1.00
对称轻尾	均匀	.030	.026	.985	.962	1.00	.999	1.00	1.00
	Beta(3, 3)	.036	.027	.938	.916	1.00	.999	1.00	1.00
对称重尾	Laplace	.058	.043	.597	.629	.931	.921	.996	.998
	$t(5)$.045	.038	.657	.703	.952	.949	.997	.998
偏态重尾	$\chi^2(5)$.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
污染	CN(0.9, 3)	.012	.013	.499	.612	.889	.917	.989	.998

结果表明，MC 检验和检验 W_{50} 的类型 II 误差属性（功效）类似。通常，这两个检验实现的模拟功效水平的量值顺序相同。只有在一种情况下，这两个检验的功效差值超过 0.1。

当样本抽取自轻尾到中尾的对称分布时，MC 检验的模拟功效水平略优于检验 W_{50} 的功效水平。另一方面，当样本抽取自重尾分布时，检验 W_{50} 的功效略高。

5. 示例

在本部分，我们将图形化 MC 过程和检验 W_{50} 应用到从 Ott 等人 (2010) (第 397 页) 获取的数据集中。这些数据如下所述：

铸造公司有几个炉子，用于在将原材料倒入蜡模之前对原材料进行加热。非常重要的一点是，必须将这些金属加热到几乎不变的某个精确温度。随机选择三个炉子，并准确记录 10 次连续加热的温度 (°C)。收集的数据如下所示：

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

图 1 显示了每个炉子的温度箱线图。箱线图表明，记录的温度中没有异常值，并且炉子 3 的温度变化与炉子 1 或炉子 2 的温度变化不同。

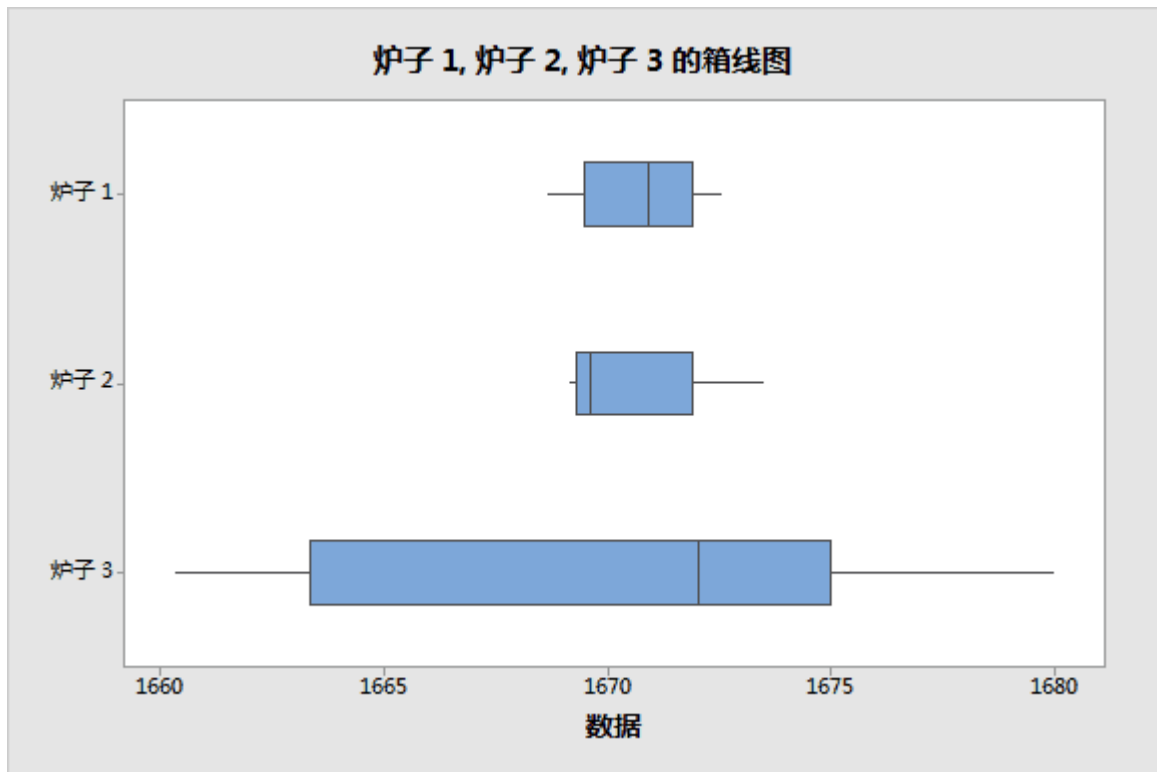


图 1 炉子温度的箱线图 (°C)

图 2 显示了相同数据的 MC 区间以及整个 MC 检验和检验 W_{50} 的结果，可参考 Levene 检验的图例。这两个检验的显著性 p 值指示，这三个炉子之间的温度变化各不相同。非重叠 MC 区间可确认，炉子 3 的变化与炉子 2 或炉子 1 不同。对于炉子 1、2 和 3，MC 区间分别为 (0.896, 2.378)、(1.072, 2.760) 和 (4.366, 12.787)。

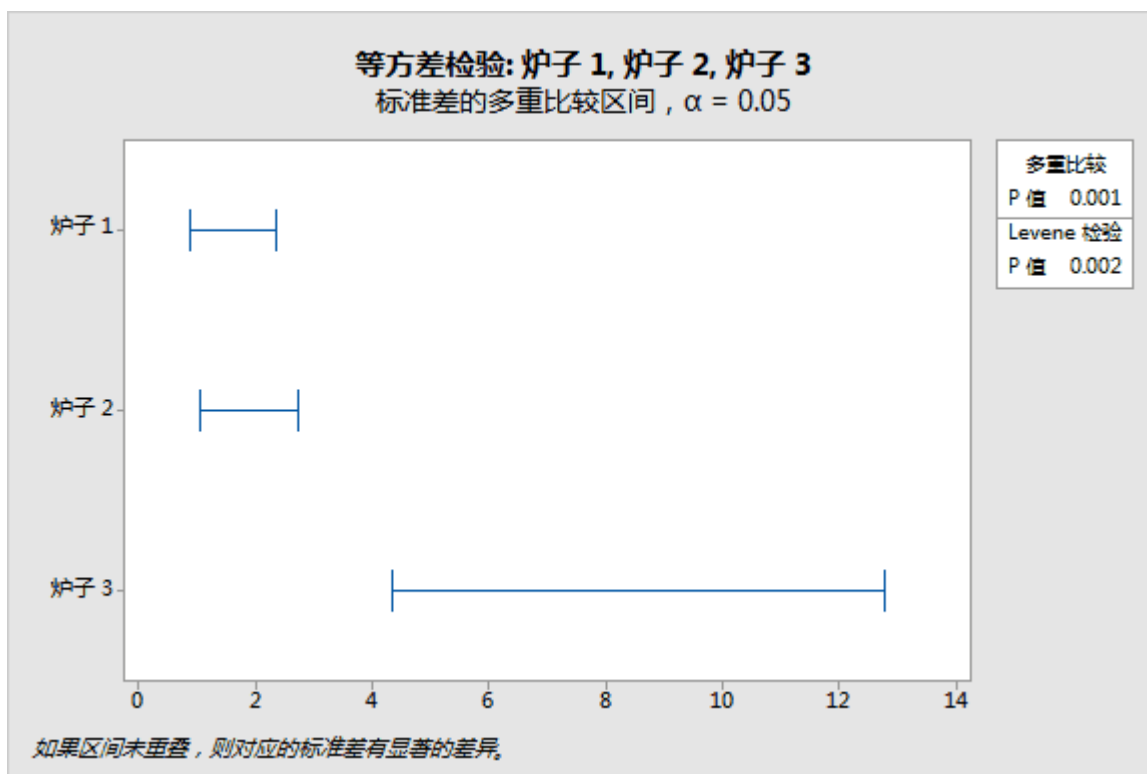


图 2 MC 检验和检验 W_{50} (Levene 检验) 的 MC 区间和 p 值

6. 结论

总之，模拟结果显示，对于采用多个小样本的设计，MC 检验的性能类似于检验 W_{50} 的性能。MC 检验更适用于轻尾到中尾对称或近对称分布，而当数据抽取自高度偏态分布和重尾分布时，检验 W_{50} 的效果可能更好。MC 过程的一个明显优势是，在标准差同质性的整体检验效果显著时，它可提供用于筛选具有不同标准差或方差的功能强大的可视化工具。Minitab 版本 17 中提供图形化 MC 过程。

7. 附录

当且仅当满足以下条件时，双样本设计中的 Bonett (2006) 的 Layard 检验调整将否定方差同质性的原假设，

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

或等效情况

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

其中

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

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

因此，如果设计是平衡设计，则 $c_{\alpha/2} = 1$ ，检验的 p 值显示如下

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

其中

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

如果设计是不平衡设计，则 $P = 2 \min(\alpha_L, \alpha_U)$ ，其中，

α_L 是此方程中 α 的最小解

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

并且， α_U 是此方程中 α 的最小解

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

对 α 的这些方程求解的方法是首先对 $z \equiv z_\alpha$ 的方程求解，然后获取 $\alpha = \Pr(Z > z)$ ，其中，随机变量 Z 呈标准正态分布。在我们描述如何对这些方程求解之前，我们注意到，方程 (1) 可以重新表示为方程 $L(z) = 0$ ，其中，

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

类似地，方程 (2) 等效于方程 $U(z) = 0$ ，其中，

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

我们发现， $L(z, n_2, n_1, S_2, S_1) = -U(z, n_1, n_2, S_1, S_2)$ 。因此，只需找到两个函数的其中一个函数的根即可。

对方程 (1) 或 (2) 求解的算法根据以下结果推导出来：

结果

设 n_1, n_2, S_1 和 S_2 为给定值和固定值。对于不平衡设计，函数 $L(z, n_1, n_2, S_1, S_2)$ 最多有两个根。

4. 如果 $n_1 < n_2$ ，则 $L(z, n_1, n_2, S_1, S_2)$ 是凸函数。它满足 $L(-\infty, n_1, n_2, S_1, S_2) = L(n_1, n_1, n_2, S_1, S_2) = +\infty$ ，并在满足以下条件时达到最小值

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

因此, 如果 $L(z_m, n_1, n_2, S_1, S_2) \leq 0$, 则有两个根, 一个在区间 $(-\infty, z_m]$ 内, 另一个在区间 $[z_m, n_1)$ 内。另外, 如果 $L(z_m, n_1, n_2, S_1, S_2) > 0$, 则函数 $L(z, n_1, n_2, S_1, S_2)$ 无根。

5. 如果 $n_1 > n_2$, 则 $L(z, n_1, n_2, S_1, S_2)$ 会从 $+\infty$ 到 $-\infty$ 单调递减, 从而具有唯一的根。如果 $L(0, n_1, n_2, S_1, S_2) = \ln S_1^2/S_2^2 \geq 0$, 则根位于区间 $[0, n_2)$ 内, 否则位于区间 $(-\infty, 0)$ 内。

论证

在以下情况下, 我们设 $L(z) \equiv L(z, n_1, n_2, S_1, S_2)$ 。

首先, 我们想要证明, 如果 $n_1 < n_2$, 则 $L(z)$ 是凸函数, 并在满足以下条件时达到最小值

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

正如前面所定义的那样

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

因此, 我们得到 $\lim_{z \rightarrow -\infty} L(z) = +\infty$, 并且

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

此外, 我们注意到 $L(z)$ 的导数满足

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

设

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

如果 $n_1 < n_2$, 则二次 $Q(z)$ 具有两个根, 按如下给定

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

与

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

由于 $Q(n_1) = \frac{n_1 - n_2}{se} < 0$, 我们得到 $z_1 < n_1 = \min(n_1, n_2) < z_2$, 因此, 对于 $(-\infty, z_1)$ 中的 z , 根为 $Q(z) > 0$, 同时, 对于 (z_1, n_1) 中的 z , 根为 $Q(z) < 0$ 。它遵循以下原则, 对于 $(-\infty, z_1)$ 中的 z , 根为 $L'(z) < 0$, 对于 (z_1, n_1) 中的 z , 根为 $L'(z) > 0$ 。因此, 在域 $(-\infty, \min(n_1, n_2))$ 上, $L(z)$ 是凸函数, 并在 $z_1 \equiv z_m$ 时达到最小值。

如果 $n_1 > n_2$, 则存在两种情况: 一种情况是 $n_1 - n_2 > 4/se$, 另一种情况是

$0 < n_1 - n_2 < 4/se$ 。在第一种情况中， z_1 和 z_2 是 $Q(z)$ 的根，因此 $n_2 = \min(n_1, n_2) < z_1 < z_2$ 。（这是因为 $n_2 - \frac{z_1 + z_2}{2} = \frac{n_2 - n_1}{2} < 0$ ）。因此对于域 $(-\infty, \min(n_1, n_2))$ 中的 z ，根为 $Q(z) > 0$ 。在第二种情况中， $Q(z)$ 无根，因此 $Q(z) > 0$ 在此域中。

它遵循以下原则，如果 $n_1 > n_2$ ，则 $L'(z) < 0$ ，以使 $L(z)$ 从 $+\infty$ 到 $-\infty$ 单调递减。

8. 参考书

Banga, S. J. and Fox, G. D. (2013)。关于 Bonett 的标准差比率稳健性置信区间。印制中。

Barnard, J. (1978)。正常范围的概率积分。 *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 27, 197 - 198.

Bonett, D. G. (2006)。标准差比率的稳健性置信区间。 *Applied Psychological Measurements*, 30, 432 - 439.

Brown, M. B., and Forsythe A. B. (1974)。方差等同性的稳健性检验。 *Journal of the American Statistical Association*, 69, 364 - 367.

Conover, W. J., Johnson, M. E. and Johnson, M. M. (1981)。方差同质性检验的比较研究以及外大陆架招标数据的应用。 *Technometrics*, 23, 351 - 361.

Hochberg, Y., Weiss, G., and Hart S. (1982)。关于多重比较的图形化过程。 *Journal of the American Statistical Association*, 77, 767 - 772.

Kramer, C. Y. (1956)。具有不等仿行数的组均值多重范围检验扩展。 *Biometrics*, 12, 307 - 310.

Layard, M. W. J. (1973)。方差同质性的稳健性大样本检验。 *Journal of the American Statistical Association*, 68, 195 - 198.

Levene, H. (1960)。“方差等同性的稳健性检验”，I. Olkin, ed., *Contributions to Probability and Statistics*, Palo Alto, CA: Stanford University Press, 278 - 292.

Miller, R. G. (1968)。折裂方差。 *Annals of Mathematical Statistics*, 39, 567 - 582.

Nakayama, M. K. (2009)。渐进有效的单阶段多重比较过程。 *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)。关于双方差等同性的 Levene 类型检验。 *Journal of Statistical Computation and Simulation*, 63, 59 - 71.

Stoline, M. R. (1981)。多重比较的状态：单因子方差分析设计中所有配对比较的同步估计。 *The American Statistician*, 35, 134 - 141.

Tukey, J. W. (1953)。多重比较问题。 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. All other marks referenced remain the property of their respective owners.