

Bonett 方法

关于 BONETT 标准差比值的稳健置信区间

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

摘要

我们建议使用替代过程以改正双标准差比值的 Bonett (2006) 置信区间 (CI) 中的细微失误。Bonett 区间所基于的 Layard (1973) 检验统计量的合并峰度估计值仅在总体方差相等时是相合的。我们推导出，替代估计量在总体方差相等时是相合的，在总体方差不相等时，仍可使用这个新的估计量计算出校正的置信区间。模拟研究表明，新的置信区间的准确度和精确度通常高于基于 Levene/Browne-Forsythe 检验 W_{50} 和 Pan (1999) 检验 L_{50} 的置信区间。与 Pan 检验一样，我们观测到，基于检验 W_{50} 的置信区间在样本数量较小时精确度会降低，通常会导致无限宽的区间。对于对称和近似对称分布，基于检验 L_{50} 的置信区间效果不错，但在总体发生偏斜时，效果欠佳。

索引项：方差齐性，Levene 检验，Brown-Forsythe 检验，Layard 检验，方差比值的置信区间 (CI)

1. 简介

众所周知，经典的 F 检验和关联的置信区间 (CI) 对正态性偏离极为敏感，实际上，这种敏感已导致经典的 F 检验不适用于大多数实际应用。鉴于此原因，很多人建议使用更稳健的替代检验。其中，名为“检验 W_{50} ”的检验通常是首选检验，原因是这种检验具有非常好的 I 类错误属性，并且计算和解释起来也很简单。（对于比较分析，请参见 Conover 等人 (1981)、Balakrishnan 和 Ma (1990) 以及 Lim 和 Loh (1996)。）检验 W_{50} 基于最初由 Levene (1960) 建议的过程，后来由 Brown 和 Forsythe (1974) 进行了功能改进。检验 W_{50} 被广泛采用，可用于大多数知名的统计软件包，如 Minitab 统计软件、SAS、R 和 JMP。

检验 W_{50} 的 II 类错误属性的改进没有其 I 类错误属性改进得那么出色。Pan (1999) 指出，对于某些分布（包括正态分布），双样本问题中的检验 W_{50} 的功效具有一个远低于 1 的上限。并且，此上限不受总体方差之间差异的量值之影响。此缺陷会自然扩展到基于检验 W_{50} 的置信区间。Pan 指出，存在一个不可忽略的概率，即基于检验 W_{50} 的总体方差比值的置信区间将是无限的 $(0, +\infty)$ ，因而提供不了任何有意义的信息。Pan 观测值与我们自己模拟的结果一致，我们将在后文介绍。

Pan 建议使用名为 L_{50} 的替代过程来改进 W_{50} 过程的局限性。根据模拟结果，Pan 得出以下结论：检验 L_{50} 的功效高于检验 W_{50} ，且稳健性相当，可以共享其所希望有的渐进属性。但是，Pan 模拟的样本是从具有重尾到轻尾的对称或适度偏斜分布中抽取的。此处没有专门介绍分布偏度对小样本的 L_{50} 检验的潜在影响。

Pan 还指出， L_{50} 过程的功效与其他非常稳健的过程（如已修改的 Fligner-Killeen 秩检验和 Hall-Padmanabhan 自适应检验）相同。但实际上，已修改的 Fligner-Killeen 秩检验和 Hall-Padmanabhan 自适应检验的用途不及检验 L_{50} 和 W_{50} ，原因是它们需要大量耗时耗力的计算。

最近，Bonett (2006) 建议使用一种替代的置信区间过程，该过程基于方差齐性的 Layard (1973) 检验的双样本格式。Bonett 包含几次调整，以提高 Layard 过程的小样本性能。例如，Bonett 建议使用合并峰度估计量，该估计量渐进等效于 Layard 检验，但显示更小的小样本偏差。

不巧的是，当总体方差不相等时，不论 Layard 的原始合并峰度估计量，还是 Bonett 建议的替代者都不具有相合性。因此，Bonett (2006) 建议的区间不是正确的置信区间，但可以更好地描述为方差等同性检验的验收区间。因此，从 1 中扣减 Bonett (2006) 中报告的模拟覆盖概率将得到方差齐性检验的 I 类错误率。将这些 I 类错误率与 Layard 原始检验的错误率进行比较，结果表明，Bonett 的调整可成功增强 Layard 检验的小样本性能。但是，必须重新参考 Bonett 建议的方差比值的置信区间。

还请注意，Bonett 比较了基于 Shoemaker (2003) 近似 F 检验的置信区间与他自己建议的区间。但是，与 Shoemaker 检验关联的方差比值的置信区间（在 Shoemaker 文章第 106 页中有简要介绍）也是基于 Layard 的合并峰度估计。因此，在 Shoemaker 文章的第 7 部分中计算的置信区间也最好地描述了方差齐性检验的接收区间。虽然会出现这些误差，但仍可以从 Bonett 模拟结果中得出以下结论：调整改进了 Layard 方差齐性检验的小样本性能，并且最终的方差齐性检验的效果比 Shoemaker 检验更好。

在本白皮书中，我们通过扩展双样本形式的 Layard 检验来改正 Bonett (2006) 中的错误，以检验有关方差或标准差比值的原假设。为完成此任务，我们建议使用对任何给定的假设比值皆有相合性的合并峰度估计量。然后，我们将对此检验统计量反解，就能获得该比值的置信区间。最终，我们进行了模拟研究，以评估小样本设计中的新置信区间的稳健性。此外，我们还在新置信区间的小样本性能和与经典的 F 检验、 W_{50} 检验和 L_{50} 检验关联的置信区间的性能之间进行比较。

2. Layard 检验和一些推广

设 $Y_{i1}, \dots, Y_{in_i}, \dots, Y_{k1}, \dots, Y_{kn_k}$ 为 k 个独立样本，每个样本都是独立同分布的，其均值为 $E(Y_{ij}) = \mu_i$ ，方差为 $\text{Var}(Y_{ij}) = \sigma_i^2 > 0$ 。此外，假设样本源自共同峰度为 $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$ 的分布。我们注意到，Layard 使用峰度过高 $\gamma_e = \gamma - 3$ 。

设 \bar{Y}_i 和 S_i 分别为样本 i 的均值和标准差。也可以设 $\tau^2 = 2 + (1 - 1/\bar{n})\gamma_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$ ，其中 $\bar{n} = \sum n_i / k$ 。正如 Layard (1973) 中所述，对于大样本数量， $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$ 。

为检验方差齐性的原假设，Layard 对分量为 $Z_i = (n_i - 1)^{1/2} \ln S_i^2 / \tau$ 的向量进行正交变换，这时，在原假设条件下，此向量之渐进分布为标准正态分布。然后，他使用正交变换下距离保持不变的性质，指出在方差齐性的原假设条件下，检验统计量 S' （给定如下）渐进分布为自由度为 $k - 1$ 的卡方分布：

$$S' = \sum_{i=1}^k (n_i - 1) \left(\ln S_i^2 - \frac{\sum_{i=1}^k (n_i - 1) \ln S_i^2}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

通常， $Z_i = (n_i - 1)^{1/2} (\ln S_i^2 - \ln \sigma_i^2) / \tau$ 渐进分布为标准正态分布。因此，可以应用 Layard 技术推导出更广义化的检验统计量 T_k' ：

$$T_k' = \sum_{i=1}^k (n_i - 1) \frac{(\ln S_i^2 - \ln \sigma_i^2)^2}{\tau^2} - \left(\sum_{i=1}^k (n_i - 1) \frac{\ln S_i^2 - \ln \sigma_i^2}{\tau \sqrt{\sum_{i=1}^k (n_i - 1)}} \right)^2$$

在原假设和备择假设两种条件下， T_k' 之渐进分布皆为自由度为 $k - 1$ 的卡方分布。

可以将 T_k' 表示为更类似于 S' 的形式。将平方项表示为双重求和，并进行一些代数运算，会得到以下结果：

$$T_k' = \sum_{i=1}^k (n_i - 1) \left(\ln S_i^2 - \ln \sigma_i^2 - \frac{\sum_{i=1}^k (n_i - 1) (\ln S_i^2 - \ln \sigma_i^2)}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

如果所有方差都相等，则 $T_k' = S'$ 。因此，在检验方差齐性的原假设时， S' 和 T_k' 的检验统计量是相同的。但是， T_k' 也被更广泛地用于检验能表示为方差函数的任何假设。例如，对于任何给定的 $\sigma_{0i} > 0, i = 1, \dots, k$ ，可以使用 T_k' 检验 $H_0: \sigma_i = \sigma_{0i}$ 形式的任何原假设。

由于 $\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$ 未知，因此，基于 S' 或 T_k' 的检验就需要总体共同峰度的估计量 γ 。例如，为检验方差齐性的原假设，Layard 建议使用以下共同峰度的合并估计量：

$$\hat{\gamma} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^4}{\left[\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right]^2} \sum_{i=1}^k n_i$$

但 Layard 指出，在方差不相等时， $\hat{\gamma}$ 不一定是共同峰度的相合估计量。

在双样本设计的特定情况下，可以通过针对给定假设比值 $\rho_0 > 0$ 检验原假设 $H_0: \sigma_1/\sigma_2 = \rho_0$ 来评估标准差之间的差别量值。但是，通过计算标准差比值的置信区间，可以更直接地评估此差值。

如果 $\rho_0 = 1$ ，则原假设等效于方差齐性假设。因此，我们可以将此检验基于 $T_2' = S'$ ，在替换 $\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$ 表达式中的 γ 的 Layard 峰度估计量的双样本格式后，就可以得到 \hat{t}^2 。

但是，如果 $\rho_0 \neq 1$ ，则必须将检验基于 T_2' 而不是 S' 之上。此外，如果 $\rho_0 \neq 1$ ，则 Layard 的合并峰度估计值不一定有相合性，因此，无法用于估计总体的共同峰度。因此，需要使用替代的合并峰度估计量（它对于任何假设比率 $\rho_0 > 0$ 都是相合估计量）。

接下来，我们推导出一个估计量。由于它是 ρ_0 的函数，我们将此估计量表示为 $\hat{\gamma}_P(\rho_0)$ 。我们也定义了检验统计量 $T_2 = \tau^2 T_2' / \hat{\tau}^2$ ，其中， $\hat{\tau}^2 = 2 + (1 - 1/\bar{n})(\hat{\gamma}_P(\rho_0) - 3)$ 。根据 Slutsky 定理， T_2 的渐进分布是自由度为 1 的卡方分布。最后，我们反解 T_2 ，就可以得到 $\rho = \sigma_1/\sigma_2$ 的置信区间。

3. 标准差比值的置信区间

前一部分详细介绍了按方差或标准差比值形式给出检验的原假设时，需要使用替代的峰度估计量的情况。以下结果提供此估计量。

结果 1

对于任何给定的 $\rho = \sigma_1/\sigma_2 > 0$ ，双样本模型中共同总体峰度的相合估计量给定为

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2}$$

可在附录 A 中找到此结果的论证。

按预期， $\hat{\gamma}_P(1)$ 与 Layard 的合并峰度估计量 $\hat{\gamma}$ 相同，原因是 $\sigma_1/\sigma_2 = 1$ 意味着标准差（方差）相等。

统计量 T_2' （即一般统计量 T_k' 的双样本格式）给定为

$$T_2' = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right) \tau^2}$$

其中， $\hat{\rho} = S_1/S_2$ ， $\rho = \sigma_1/\sigma_2$ 而且 $\tau^2 = 2 + (1 - 1/\bar{n})\hat{\gamma}_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$ 。

按 Layard (1973) 中所述，在大样本中， $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$ 。Bonett (2006) 使用替代近似值，该近似值也被 Shoemaker (2003) 在 $\text{Var}((n_i - 1)^{1/2} \ln S_i^2) \cong \gamma - (n_i - 3)/n_i$ 中采用。在大样本中，这些近似值等效。但是，Shoemaker 报告，在小样本条件下进行方差齐性检验时，后者更具有优势。通过此调整，统计量 T_2' 可修改为

$$T_2' = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\gamma - g_1}{n_1 - 1} + \frac{\gamma - g_2}{n_2 - 1}}$$

其中， $g_i = (n_i - 3)/n_i$ 。

然后，用于检验原假设 $H_0: \rho = \rho_0$ 的检验统计量 $T_2 = \tau^2 T_2' / \hat{\tau}^2$ 给定为

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{\gamma}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho_0) - g_2}{n_2 - 1}}$$

在 T_2 的这一表达式中，可以将分母的平方根当作合并峰度标准误的大样本估计值。

此外，在 $\hat{\gamma}_P(1) \equiv \hat{\gamma}$ 的表达式中，Bonett (2006) 使用了以调整的比率 $1/[2(n_i - 4)^{1/2}]$ 得到的调整的样本均值。相应地，我们对合并峰度估计量进行了相同的调整：

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

其中, m_i 是样本 i 的调整均值, 其调整比率为 $1/[2(n_i - 4)^{1/2}]$ 。由于调整均值 m_i 是总体均值 μ_i 的相合估计量, 因此, 此格式的合并峰度估计量和以前格式的估计量是渐进等效的。但是, 基于 T_2 的检验, 此替代格式可以提高小样本的检验性能。

检验统计量 T_2 现在已反解, 可推导出方差或标准差比值的近似置信区间。但是首先, 我们将简要介绍标准差比值的 Bonett (2006) 置信区间在推导中的失误。

3.1 Bonett 区间

Bonett (2006) 对以下统计量反解 (而不是对 T_2 反解) 来获得置信区间

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

结果是, 所得的区间只是方差齐性检验的接收区域。其原因是, 合并峰度估计量 $\hat{\rho}_P(1)$ 是相合估计量仅在方差相等, 或者假设比值为 1 的等价情况下才成立的。所得的区间在 Bonett (2006) 中报告为

$$\exp[\ln(c S_1^2/S_2^2) \pm z_{\alpha/2} se]$$

其中

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

包括作为小样本调整的常数 c , 以降低不平衡设计中不等尾误差概率的影响。此常数按如下给定

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

在设计平衡并且其影响随着样本数量增大而可以被忽略时, 此常数就会消失。

表 1 说明了将上述区间误当成置信区间的后果。这些结果基于小样本模拟研究, 在该研究中, 我们基于 Bonett (2006) 区间计算模拟覆盖概率。对于等方差情况 (左列), 我们从标准正态分布中抽取两个独立样本。对于不等方差情况 (右列), 我们按常数因子 4 调整第二个样本的观测值。估计覆盖概率基于 100,000 次重复。目标名义覆盖率为 0.95。

表 1 总体方差不等对 Bonett (2006) 置信区间 ($\alpha = 0.05$) 的影响

n_1, n_2	模拟覆盖概率	
	等方差	不等方差
10, 10	0.963	0.972
50, 50	0.952	0.991
100, 100	0.952	0.994

如果这些区间基于有相合性的合并峰度估计量, 则预计这两种情况下的覆盖概率相同。但请注意, 在方差不相等时, 这些区间始终比较保守。此外, 随着样本数量增大, 覆盖概率趋向于 1。请注意, 通过 Shoemaker (2003) 近似置信区间可获得类似结果。

3.2 置信区间的计算

分析根据备择假设 $H_A: \rho \neq \rho_0$ 检验原假设 $H_0: \rho = \rho_0$ 的问题, 其中, $\rho = \sigma_1/\sigma_2$ 和 $\rho_0 > 0$ (基于前面给定的检验统计量 T_2)。在原假设条件下, 检验统计量

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1}}$$

渐进分布为自由度为 1 的卡方分布。因此, 当且仅当满足以下条件, 此检验将在显著性水平为 α 时否定原假设

$$(\ln \hat{\rho}^2 - \ln \rho_0^2)^2 > z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

其中, z_α 表示标准正态分布的第 $\alpha \times 100$ 个上百分位点。请注意, 自由度为 1 的卡方分布的第 $\alpha \times 100$ 个上百分位点 $\chi_{1,\alpha}^2$ 符合以下条件: $\chi_{1,\alpha}^2 = z_{\alpha/2}^2$ 。

Bonett (2006) 模拟结果显示, 在小样本检验时, 此调整用于降低不平衡设计中不等尾误差概率的影响其效果良好。因此, 我们可以对基于 T_2 的检验进行类似调整。在进行此调整后, 当且仅当满足以下条件时, 检验才拒绝原假设

$$(\ln \rho_0^2 - \ln(c\hat{\rho}^2))^2 > z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

其中, c 是 Bonett 的调整常数, 给定为

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

等价地, 基于 T_2 的 $\rho = \sigma_1/\sigma_2$ 的置信度近似百分之 $(1 - \alpha)100$ 的集合按如下给定

$$\left\{ \rho \in (0, \infty): (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho) - g_2}{n_2 - 1} \right) \leq 0 \right\}$$

请注意, c 对平衡设计不产生任何影响, 并且, 在大样本不平衡设计中只有一些可忽略的影响。

下一个结果提供了置信集合的另一种表达式, 按此更便于描述其自然的形式。在此表达式中, 合并峰度估计量将根据各个样本峰度重写, 按如下给定

$$\hat{\gamma}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1, 2$$

结果 2

基于 T_2 的 $\rho = \sigma_1/\sigma_2$ 的置信度近似百分之 $(1 - \alpha)100$ 的集合按如下给定

$$\hat{\rho}\sqrt{c} \{r \in (0, \infty): H(r^2) \leq 0\}$$

或等价地, $\rho^2 = \sigma_1^2/\sigma_2^2$ 的置信集合可以表示为

$$c\hat{\rho}^2 \{r \in (0, \infty): H(r) \leq 0\}$$

其中

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 se^2(cx), x > 0$$

$$se^2(x) = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 x^2/n_2}{(K+x)^2} - B$$

$$A = \frac{(n_1+n_2)(n_1+n_2-2)}{(n_1-1)(n_2-1)}, B = \frac{g_1}{n_1-1} + \frac{g_2}{n_2-1}, K = \frac{n_1-1}{n_2-1}$$

有关此结果的论证，请参见附录 B。

很容易验证函数 $H(x)$ 在正实线上的连续性，且有 $H(0) = H(+\infty) = +\infty$ 及 $H(1) < 0$ 。因此，根据微分中值定理，函数 $H(x)$ 在区间 $(0, 1)$ 中至少有一个根，在区间 $(0, +\infty)$ 中至少有一个根。

下一个结果介绍了作为一个区间或一组分离区间的置信集合。

结果 3

如果函数 $H(x)$ 只有两个根， x_L 和 x_U ，则 $0 < x_L < 1 < x_U$ 和 $\rho^2 = \sigma_1^2/\sigma_2^2$ 的置信集合是按如下给定的区间

$$[c\hat{\rho}^2 x_L, c\hat{\rho}^2 x_U]$$

然后，它遵循以下原则： $\rho = \sigma_1/\sigma_2$ 的置信区间是区间

$$[\hat{\rho}\sqrt{cx_L}, \hat{\rho}\sqrt{cx_U}]$$

另一方面，如果函数 $H(x)$ 超过两个根，则 $\rho^2 = \sigma_1^2/\sigma_2^2$ 的置信集合为一组非重叠区间。如果函数开口向上，则每个区间的端点是相邻的两个根。

有关此结果的论证，请参见附录 C。

备注

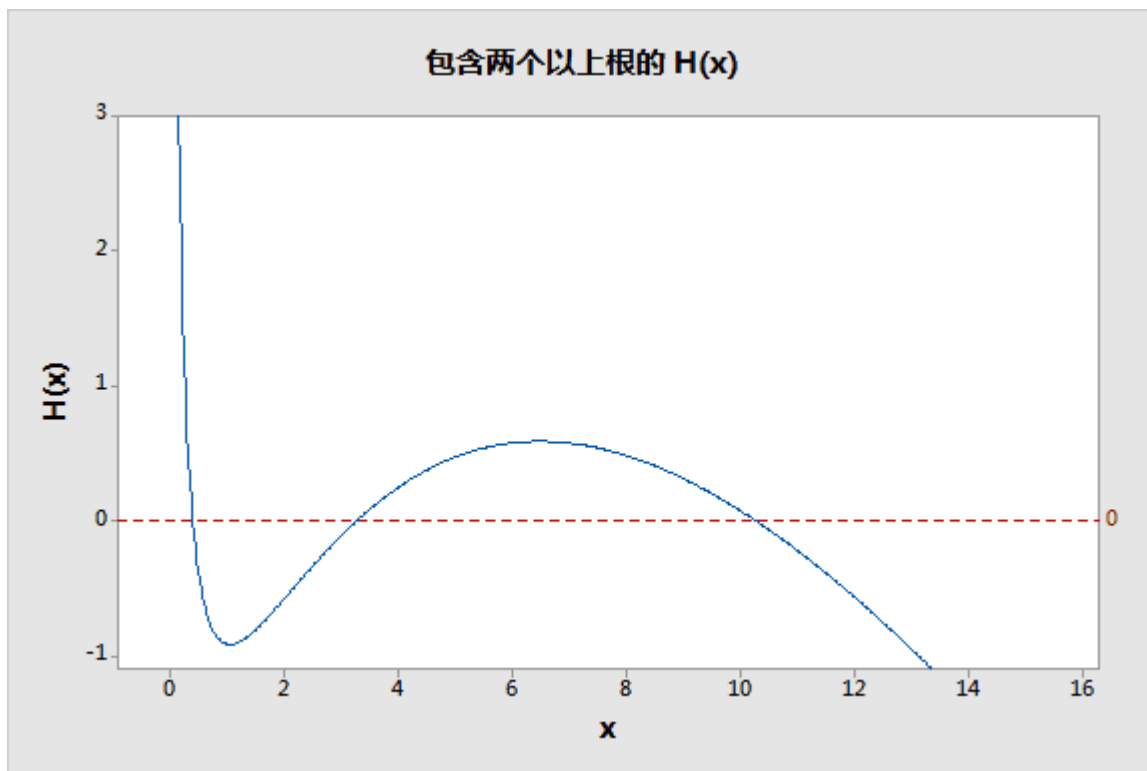
虽然在数学上，函数 $H(x)$ 允许两个以上的根，但我们观测到，只有在极其异常且实际上没有任何意义的设计中才会出现这种情况，在这类设计中，一个或两个样本太小，或严重不平衡。我们推测， $H(x)$ 只能具有两个根或四个根。

以下示例基于虚构的数据，以强制函数 $H(x)$ 具有两个以上的根。这些数据汇总如下： $n_1 = 169$ 、 $n_2 = 7$ 、 $S_1 = 301.855$ 、 $S_2 = 4606.170$ 、 $\hat{\gamma}_1 = 1.877$ 、 $\hat{\gamma}_2 = 6.761$ 、 $c = 0.728$ 、 $A = 30.381$ 、 $B = 0.101$ 和 $K = 28.000$ 。

对于 $\alpha = 0.05$ ，函数 $H(x)$ 给定为

$$H(x) = (\ln x)^2 - 1.96^2 \left(30.381 \frac{1.877 \times 28^2/169 + 6.761 \times (.728x)^2/7}{(28.000 + 0.728x)^2} - 0.101 \right)$$

在这种情况下，函数 $H(x)$ 具有四个根。函数图显示如下。请注意，由于第四个根太大，因此图中显示不出来。但是，我们知道，由于 $H(+\infty) = +\infty$ ，存在第四个根。



这四个根在数值上计算为 $x_1 = 0.389$ 、 $x_2 = 3.282$ 、 $x_3 = 10.194$ 和 $x_4 = 39685.0$ 。标准差比值的估计为 $\hat{\rho} = S_1/S_2 = 0.066$ 。 $\rho^2 = \sigma_1^2/\sigma_2^2$ 的置信集合表示为

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] \cup [c \hat{\rho}^2 x_3, c \hat{\rho}^2 x_4] = [0.001, 0.010] \cup [0.032, 124.072]$$

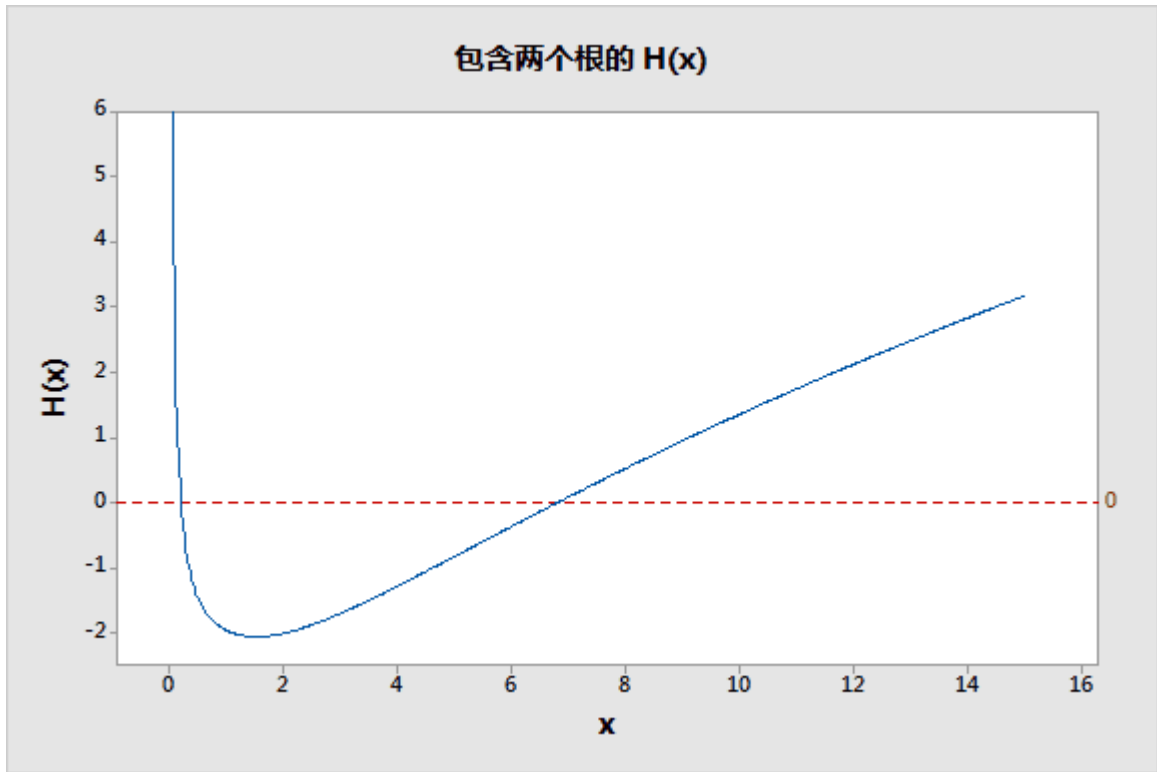
标准差比值的置信集合 ρ 可通过求区间端点的平方根来获得。

在样本不太小 ($n_i \geq 10$) 并且其数量之间的不一致性不太大时，函数 $H(x)$ 通常会有两个根。如结果 2 中所述，一个根小于 1，另一个根大于 1。此处的示例基于随机生成的数据。这些数据可以汇总如下： $n_1 = 10$ 、 $n_2 = 12$ 、 $S_1 = 1.150$ 、 $S_2 = 1.043$ 、 $\hat{\gamma}_1 = 2.704$ 、 $\hat{\gamma}_2 = 3.671$ 、 $c = 1.041$ 、 $A = 4.444$ 、 $B = 0.146$ 和 $K = 0.818$ 。

对于 $\alpha = 0.05$ ，在这种情况下，函数 $H(x)$ 给定为

$$H(x) = (\ln x)^2 - 1.960^2 \left(4.444 \frac{2.704 \times 0.818^2/10 + 3.671 \times (1.041x)^2/12}{(0.818 + 1.041x)^2} - 0.146 \right)$$

函数 $H(x)$ 具有如下所示的两个根：



这两个根在数值上计算为 $x_1 = 0.200$ 和 $x_2 = 6.824$ 。标准差比值为 $\hat{\rho} = S_1/S_2 = 1.102$ 。 $\rho^2 = \sigma_1^2/\sigma_2^2$ 的置信集合是按如下给定的区间

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] = [0.253, 8.634]$$

标准差比值的置信区间 ρ 可通过求上述区间端点的平方根来获得。

现在，我们介绍两种用于查找置信限的算法。

第一种算法包括使用数值根查找器过程来查找函数 $H(x)$ 的根。与方差比值的置信下限对应的根限制在区间 $(0, 1)$ 内。如果我们通过 x_L 表示此根，则按结果 3，方差比值的置信下限计算为 $c \hat{\rho}^2 x_L$ ，标准差比值的置信下限为 $\hat{\rho} \sqrt{c x_L}$ 。类似地，方差比值的置信上限为 $c \hat{\rho}^2 x_U$ ，标准差比值的置信上限为 $\hat{\rho} \sqrt{c x_U}$ ，其中 $x_U > 1$ 是 $H(x)$ 的另一个根。查找置信上限的简单方法是利用 $1/\rho^2$ 的下限就是 ρ^2 的上限的事实。首先，第一个样本和第二个样本的角色在函数 $H(x)$ 的表达式中互换，就像为比值 $1/\rho^2 = \sigma_2^2/\sigma_1^2$ 计算置信限一样。其次，对新函数 $H(x)$ 应用查找下限的算法。最后，对所得的限值求倒数，以获得所需的置信上限。

另一种方法是使用按如下给定的递归关系，按递归公式计算方差比值的置信下限

$$\rho_0^2 = 1$$

$$\rho_{i+1}^2 = \exp \left[\ln(c \hat{\rho}^2) - z_{\alpha/2} \sqrt{\frac{\hat{Y}_P(\rho_i) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_i) - g_2}{n_2 - 1}} \right], i = 0, 1, 2, \dots$$

方差比值的置信下限是 ρ_{j+1}^2 ，使得 $|\rho_{j+1}^2 - \rho_j^2| < \varepsilon$ ，其中 $j > 0$ 和 ε 被选为很小的值（例如 $\varepsilon = 10^{-6}$ ）。为找到置信上限，我们只要使用上面的 $+z_{\alpha/2}$ 替换 $-z_{\alpha/2}$ 即可。

显然，计算置信限的这两种算法等效，因为递归过程本质上就是对 ρ^2 的方程 $H(\rho^2 / (c \hat{\rho}^2)) = 0$ 求解的迭代过程。递归算法易于实施，因此，在根查找器过程不可用时，递归算法是一种有用的替代方法。

4. 模拟研究和结果

在本白皮书中，我们推导出一个推广的 Layard 的双方差齐性检验的过程，以检验方差比值。我们将此过程称作推广的 Layard 比值检验，简称 ELTR。在这部分中，我们根据 ELTR 过程对置信区间的小样本属性进行了调查。我们遵循 Bonett (2006) 采用的一般方法。

我们将基于 ELTR 过程的置信区间与基于检验 L_{50} (Pan, 1999) 和检验 W_{50} (Levene/Brown-Forsythe 检验) 的置信区间进行比较。对于研究 1，我们还包括基于经典的 F 检验的置信区间，以进行比较。我们知道，在数据呈正态分布时，经典的 F 检验是最优的。请注意，基于检验 W_{50} 和 L_{50} 的置信区间的计算已在 Pan (1999) 中给定。可以在许多介绍性的统计学教科书中找到基于 F 检验的置信区间的计算方法。它们也在 Bonett (2006) 中给定。

我们进行了三项模拟研究，每项研究都有 100,000 次重复。每次重复包含数量为小到中等的两个独立样本。每个样本均抽取自具有已知属性（包括对称、不对称、重尾和轻尾）的父分布。对于 90%、95% 和 99% 的名义置信水平，与每个模拟关联的标准误分别大约为 0.0009、0.0007 和 0.0003。

为评估每个过程的性能，我们报告了所取得的覆盖概率和方差比值的模拟区间的平均宽度。与检验 W_{50} 关联的一些区间具有无限宽度（由 Pan (1999) 揭示出其可能性）。在这些情况下，我们报告了有限区间的平均宽度和有无限宽度区间所占的百分比。所有的模拟都采用 Mathematica 软件包版本 8 来进行。

研究 1：正态数据的覆盖概率比较

在第一项研究中，我们生成了服从正态分布的各种数量的随机样本。结果如表 2 中所示。

表 2 覆盖概率和平均区间宽度的比较

$1 - \alpha$	n_1, n_2	测量值	过程			
			F	ELTR	L_{50}	W_{50}
0.90	10, 10	覆盖率	0.898	0.918	0.913	0.921
		宽度	3.72	5.06	4.72	8.03 (0.01%)
	30, 10	覆盖率	0.90	0.909	0.897	0.911
		宽度	2.42	3.01	3.58	3.17
	25, 25	覆盖率	0.902	0.907	0.914	0.916
		宽度	1.61	1.73	1.85	1.938
	50, 50	覆盖率	0.90	0.901	0.906	0.907
		宽度	1.03	1.06	1.13	1.15

$1 - \alpha$	n_1, n_2	测量值	过程			
			F	ELTR	L_{50}	W_{50}
0.95	10, 10	覆盖率	0.949	0.963	0.958	0.964
		宽度	4.90	7.72	6.52	497.24 (0.20%)
	30, 10	覆盖率	0.95	0.957	0.945	0.959
		宽度	2.98	4.91	4.67	4.07
	25, 25	覆盖率	0.951	0.955	0.958	0.961
		宽度	1.99	2.24	2.31	2.49
	50, 50	覆盖率	0.951	0.952	0.953	0.954
		宽度	1.25	1.31	1.38	1.41
0.99	10, 10	覆盖率	0.989	0.993	0.992	0.994
		宽度	8.29	17.76	12.52	$> 10^4$ (8.8%)
	30, 10	覆盖率	0.99	0.992	0.986	0.994
		宽度	4.26	15.76	8.26	6.77
	25, 25	覆盖率	0.99	0.992	0.992	0.993
		宽度	2.86	3.66	3.43	4.03
	50, 50	覆盖率	0.99	0.991	0.991	0.991
		宽度	1.71	1.89	1.92	2.02

阴影行显示每个过程在每个置信水平 ($1 - \alpha$) 和每种样本数量组合 (n_1, n_2) 得到的覆盖概率 (覆盖率)。区间宽度 (宽度) 的均值显示在每个覆盖概率下方。如果某个条件下的任何区间是无限的, 则我们会报告有限区间的均值和无限区间所占的百分比。

不出所料, 结果显示, 与 F 过程关联的置信区间不仅最准确, 而且也最精确。F 过程取得的覆盖概率与目标覆盖率的接近程度比其他过程高。与 F 过程关联的区间的平均宽度比其他过程关联的区间小。但是, 此表也显示了使用 ELTR 和 L_{50} 过程构建的置信区间几乎与基于 F 过程的置信区间一样精确和准确。

基于检验 W_{50} 的区间也相当准确。但是, W_{50} 区间非常宽, 甚至具有无限宽度, 具体取决于样本大小。请注意, 在这两个样本仅具有 10 个观测值时, W_{50} 过程生成的区间中, 至少有 0.01% 的区间宽度无限。在目标覆盖率增大时, 无限区间的百分比也会增大。在大多数情况下, ELTR 和 L_{50} 区间的平均宽度小于 W_{50} 区间的平均宽度。

研究 2：非正态数据的覆盖概率比较

第二项研究旨在当父分布不是正态分布时评估和比较 ELTR、 L_{50} 和 W_{50} 过程的性能。为了评估异常值对过程性能的影响，我们还会包含污染正态分布。我们将此污染分布指定为 $CN(0.1, 3)$ ，以表明在 90% 的观测值抽取自标准正态分布时，剩余的 10% 抽取自均值为 0 且标准差为 3 的正态分布中。结果如表 3 中所示。

表 3 名义置信水平为 $1 - \alpha = 0.95$ 的某些非正态模型中的覆盖概率和平均区间宽度的比较

分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}	分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}
均匀				$\chi^2(5)$			
[1.8]				[5.4]			
10, 10	0.971	0.971	0.966	10, 10	0.956	0.938	0.956
	5.27	4.87	42.08 (0.1%)		11.61	8.78	$> 10^4$ (2.6%)
10, 30	0.964	0.961	0.957	10, 30	0.959	0.923	0.956
	2.51	2.4	2.89		6.25	4.14	190.645 (0.3%)
25, 25	0.967	0.972	0.968	25, 25	0.956	0.944	0.954
	1.43	1.79	1.88		3.66	2.92	3.26
50, 50	0.959	0.962	0.959	50, 50	0.959	0.946	0.952
	0.83	1.06	1.08		2.07	1.7	1.77
Beta (3, 3)				指数			
[2.5]				[9]			
10, 10	0.968	0.966	0.966	10, 10	0.947	0.916	0.95
	6.26	5.59	254.62 (0.1%)		20.99	14.47	$> 10^4$ (9.1%)
10, 30	0.96	0.954	0.96	10, 30	0.954	0.896	0.953
	3.14	2.76	3.71		10.46	6.19	$> 10^4$ (4.1%)
25, 25	0.959	0.966	0.965	25, 25	0.956	0.931	0.951
	1.81	2.06	2.18		6.09	4.13	5.48 (0.008%)

分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}	分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}
50, 50	0.957	0.959	0.958	50, 50	0.962	0.942	0.952
	1.06	1.23	1.26		3.18	2.24	2.38
Laplace				$\chi^2(1)$			
[6]				[15]			
10, 10	0.946	0.935	0.961	10, 10	0.928	0.889	0.947
	13.47	10.45	$> 10^4$ (3.0%)		55.09	37.4	$> 10^5$ (25.1%)
10, 30	0.947	0.919	0.957	10, 30	0.943	0.882	0.956
	6.78	4.82	$> 10^4$ (0.4%)		18.71	11.14	$> 10^6$ (25.7%)
25, 25	0.945	0.94	0.952	25, 25	0.952	0.925	0.954
	4.00	3.372	3.86		10.97	6.84	$> 10^4$ (0.4%)
50, 50	0.952	0.949	0.955	50, 50	0.958	0.936	0.951
	2.19	1.91	1.99		5.08	3.31	3.75 (0.001%)
t(5)				对数正态			
[9]				[113.9]			
10, 10	0.957	0.946	0.965	10, 10	0.923	0.876	0.955
	11.07	8.81	$> 10^3$ (2.0%)		59.22	46.15	$> 10^5$ (23.0%)
10, 30	0.957	0.93	0.959	10, 30	0.949	0.866	0.958
	6.06	4.24	$> 10^3$ (0.7%)		29.13	17.67	$> 10^6$ (31.6%)
25, 25	0.954	0.948	0.96	25, 25	0.947	0.917	0.965
	3.54	2.93	4.86 (0.01%)		16.21	8.73	$> 10^4$ (2.4%)
50, 50	0.954	0.947	0.954	50, 50	0.955	0.928	0.96
	2.10	1.71	1.77 (0.003%)		8.62	4.11	164.38 (0.2%)

分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}	分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}
半正态(M)				CN(0.1, 3)			
[3.9]				[8.3]			
10, 10	0.956	0.942	0.954	10, 10	0.977	0.965	0.979
	10.41	7.89	$> 10^4$ (1.5%)		12.64	9.52	$> 10^4$ (4.9%)
10, 30	0.959	0.93	0.954	10, 30	0.981	0.952	0.979
	5.18	3.64	13.00 (0.02%)		7.82	4.71	944.68 (1.1%)
25, 25	0.959	0.952	0.959	25, 25	0.982	0.972	0.981
	3.01	2.62	2.88		4.63	3.22	3.71
50, 50	0.96	0.951	0.954	50, 50	0.983	0.972	0.978
	1.69	1.54	1.59		2.64	1.83	1.91

阴影行显示每个过程、父分布和样本数量组合得到的覆盖概率。区间宽度的均值显示在每个覆盖概率下方。如果某个条件下的任何区间是无限的，则我们会报告有限区间的均值和无限区间所占的百分比。每个父分布的峰度 (γ) 显示在方括号中。

对于对称轻尾分布，结果指示，所有三种方法都会产生类似的保守覆盖概率。但是，小样本的 ELTR 和 L_{50} 区间比 W_{50} 区间更精确。例如，在样本抽取自带有参数 (3, 3) 的 Beta 分布时，ELTR 和 L_{50} 区间取得的覆盖概率至少与 W_{50} 区间的覆盖概率一样准确，但 W_{50} 区间始终更宽。

对于对称重尾分布而言，ELTR 和 W_{50} 区间有点保守，而 L_{50} 区间比较宽松。在设计为不平衡设计时， L_{50} 区间更宽松。例如，在数量为 10 和 30 的样本抽取自 Laplace 分布时， L_{50} 区间取得的覆盖概率为 0.919。并且，在相同数量的样本抽取自自由度为 5 的 t 分布时， L_{50} 区间取得的覆盖概率为 0.930。

在小样本抽取自高度偏斜的重尾分布时， L_{50} 区间也十分宽松。例如，在样本抽取自对数正态分布时，取得的覆盖率可低至 0.866。对于这些分布， W_{50} 方法是这三种方法中变化幅度最小的方法。但是，太多 W_{50} 区间具有无限宽度。例如，当样本抽取自自由度为 1 的卡方分布 ($\chi^2(1)$) 时，超过 25% 的 W_{50} 区间具有无限宽度。ELTR 区间准确度比较低，但更窄，因此，与 W_{50} 区间相比，可提供更多信息。

最终，我们注意到，这三个过程均受异常值的负面影响。 L_{50} 方法受到的影响最小，这在意料之中，因为推导出 L_{50} 方法可降低异常值对检验 W_{50} (Pan, 1999) 的影响。在样本抽取自污染正态分布 CN(0.1, 3) 时，ELTR 和 W_{50} 过程得到的覆盖概率最小值为 0.977。其他模拟结果 (未显示) 指示这些区间随着样本数量的增大而缓慢地改进。

研究 3：等峰度假设的敏感度

我们的最终研究是探讨 ELTR 过程对于相等峰度的敏感度，这是因为由等峰度才推导出了 ELTR 过程。在父总体的峰度不相等，即 $\gamma_1 \neq \gamma_2$ 时，我们检查了 ELTR 过程的性能。我们还包含了 L_{50} 和 W_{50} 过程，因为它们是在总体类似的假设情况下推导出的。在父总体的峰度不相等时，此相似假设已被破坏。结果如表 4 中所示。

表 4 名义置信水平为 $1 - \alpha = 0.95$ 时，ELTR 过程对等峰度假设的敏感度

分布 1, 分布 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}	分布 1, 分布 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}
Beta (3, 3), 正态 [2.5, 3]				正态, CN (0.9, 3) [3, 8.3]			
10, 10	0.964	0.961	0.964	10, 10	0.955	0.948	0.951
	0.27	0.23	204.50 (0.20%)		6.88	5.16	$> 10^4$ (4.89%)
30, 10	0.946	0.939	0.946	30, 10	0.941	0.91	0.942
	0.16	0.17	0.15		5.26	3.77	3.20
10, 30	0.966	0.956	0.967	10, 30	0.961	0.95	0.958
	0.14	0.11	0.17		4.26	2.40	630.42 (1.10%)
50, 50	0.951	0.95	0.949	50, 50	0.936	0.91	0.907
	0.04	0.05	0.05		1.27	1.11	1.19
正态, Laplace [3, 6]				半正态, $\chi^2(5)$ [3.9, 5.4]			
10, 10	0.941	0.935	0.947	10, 10	0.956	0.94	0.954
	6.67	5.17	$> 10^6$ (2.90%)		0.42	0.32	304.41 (2.60%)
30, 10	0.912	0.888	0.914	30, 10	0.954	0.918	0.949
	5.06	3.85	3.21		0.33	0.22	0.20
10, 30	0.963	0.943	0.955	10, 30	0.962	0.934	0.958
	3.33	2.25	$> 10^3$ (0.40%)		0.23	0.15	3.28 (0.30%)

分布 1, 分布 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}	分布 1, 分布 2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}
50, 50	0.935	0.894	0.889	50, 50	0.955	0.941	0.945
	<i>0.98</i>	<i>1.04</i>	<i>1.12</i>		<i>0.07</i>	<i>0.06</i>	<i>0.07</i>
正态, 半正态(M) [3, 3.9]				$\chi^2(5)$, 指数 [5.4, 9]			
10, 10	0.956	0.948	0.957	10, 10	0.938	0.914	0.94
	<i>28.16</i>	<i>20.65</i>	$> 10^4$ (1.50%)		<i>211.17</i>	<i>137.88</i>	$> 10^6$ (9.10%)
30, 10	0.946	0.924	0.947	30, 10	0.928	0.875	0.929
	<i>20.59</i>	<i>14.83</i>	<i>12.78</i>		<i>194.70</i>	<i>93.02</i>	<i>83.02</i>
10, 30	0.961	0.946	0.962	10, 30	0.968	0.93	0.954
	<i>14.06</i>	<i>9.37</i>	<i>49.11</i> (0.02%)		<i>102.35</i>	<i>55.29</i>	$> 10^5$ (3.90%)
50, 50	0.953	0.95	0.952	50, 50	0.95	0.92	0.923
	<i>4.32</i>	<i>4.16</i>	<i>4.33</i>		<i>29.64</i>	<i>23.37</i>	<i>25.54</i>

阴影行显示每个过程、父分布组合（分布 1, 分布 2）和样本数量组合得到的覆盖概率。区间宽度的均值显示在每个覆盖概率下方。如果某个条件下的任何区间是无限的，则我们会报告有限区间的均值和无限区间所占的百分比。每个父分布的峰度 (γ_1, γ_2) 显示在方括号中。

通常，在样本足够大时，ELTR 过程的性能似乎不会受到不等峰度的负面影响。但是，在设计不平衡，样本数量较小并且从重尾程度更高的分布获取样本时，取得的覆盖概率变化很大。在从重尾程度更高的分布中抽取大样本时，得到的覆盖概率更好些。

在样本数量足够大时，对于因不等峰度导致的分布不相似性， L_{50} 和 W_{50} 区间通常也似乎比较稳健。但是，请注意，在样本抽取自正态分布和 Laplace 分布，或抽取自正态分布和污染正态分布时， L_{50} 和 W_{50} 区间的覆盖概率不稳定，即使样本数量大到 50 也是如此。

L_{50} 区间通常比 ELTR 和 W_{50} 区间更宽。在这三种情况下， L_{50} 区间得到的覆盖概率小于 0.90。相比之下，只有 W_{50} 区间得到的覆盖概率之一小于 0.90。ELTR 区间得到的最低覆盖概率为 0.912。

上一研究（表 3）显示，在这两个样本均抽取自污染正态分布 CN(0.1, 3) 时，这三个过程都产生了明显地更加保守的区间。目前的研究显示，当只有一个样本抽取自 CN(0.1, 3) 时，这三个过程的执行效果显著提高。但请注意，在样本数量增大到 50 时， L_{50} 和 W_{50} 区间将显著变短。

5. 示例

在此部分中，我们将所有四个过程（F、ELTR、 L_{50} 和 W_{50} ）应用到从 Pan (1999) 获取的数据集中。Ott (1993, 第 352 页) 按如下所述介绍这些数据：

铁矿方面的化学家怀疑，每磅铁矿中所含氧化铁的量（重量，盎司）的方差往往随着每磅铁矿石含氧化铁量的均值的增大而增大。要检验此理论，在两个位置中的每个位置选择了十个 1 磅的铁矿标本：一个位置（即位置 1）包含的氧化铁的均值含量高于另一个位置（即位置 2）。铁矿标本中氧化铁的含量如下所示：

位置 1	8.1	7.4	9.3	7.5	7.1	8.7	9.1	7.9	8.4	8.8
位置 2	3.9	4.4	4.7	3.6	4.1	3.9	4.6	3.5	4.0	4.2

使用四种不同方法计算的 $\sigma_2/\sigma_1 = 1/\rho$ 的 95% 置信区间在下表中给定：

过程	95% 置信区间
F	(0.262, 1.055)
ELTR	(0.277, 0.924)
L_{50} (Pan)	(0.295, 0.938)
W_{50} (Levene/Brown-Forsythe)	(0.237, 0.908)

6. 结论

我们的模拟显示，通常，基于 ELTR 过程的置信区间与根据检验 L_{50} 和 W_{50} 推导出的置信区间有同样的准确度。但是，对大多数分布来说，ELTR 区间和 L_{50} 区间的精确度要比 W_{50} 区间高。在小样本抽取自严重偏斜和重尾分布时， W_{50} 区间的准确度往往高于 ELTR 区间和 L_{50} 区间。但是，此优势通常会被明显降低的精确度所抵消。所得的 W_{50} 区间通常太宽，甚至可能无限宽。

根据设计， L_{50} 区间可提高 W_{50} 区间的精确度。但对于偏斜总体， L_{50} 区间很短，以至于产生过度宽松的覆盖概率。相比之下，ELTR 区间通常更稳定。ELTR 区间通常不太长或太短，以便覆盖概率通常不太保守或宽松。因此，ELTR 过程对于大多数有实用目的的应用来说是最佳过程。

ELTR 区间的计算比基于检验 L_{50} 或检验 W_{50} 的区间有些耗时耗力。但通常，提高精确度（与 W_{50} 区间相比）和准确度（与 L_{50} 区间相比）比额外的计算工作更加重要。ELTR 过程已作为 Minitab 统计软件版本 17 中的双样本方差分析的一部分实施，在该软件版本中，它被称作 Bonett 过程。

为进行进一步研究，可以考虑在 Layard 的合并峰度估计量被 Bonett 的更稳健版本（按如下给定）替换后，考查多样本设计中 Layard 检验的小样本属性

$$\hat{\gamma}_A = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{\left[\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^2 \right]^2} \sum_{i=1}^k n_i$$

其中, m_i 是样本 i 的调整均值, 调整比率为 $1/[2(n_i - 4)^{1/2}]$ 和 $i = 1, \dots, k$ 。

此外, 最好使用 Shoemaker 的对数变换样本方差的渐进方差的近似方法。

最后, 我们注意到, Bonett (2006) 建议的区间虽然不适合作为置信区间, 但在作为双方差齐性检验的接收区域时, 对于大多数分布仍非常准确和精确。这些接收区域将非常适合作为图形过程的基础, 以用于比较多个方差。Hochberg、Weiss 和 Hart (1982) 提供了一个用于检验均值相等性的类似过程。此过程已作为 Minitab 统计软件版本 17 中的等方差检验分析的一部分实施, 在该软件版本中, 它被称作多重比较过程。

7. 附录

附录 A: 结果 1 的论证

使用第 2 部分的表示法和假设, 针对给定的 $\rho = \sigma_1/\sigma_2$, 设 $X_j = \rho Y_{2j}$ 。得到

$$\text{Var}(X_j) = \rho^2 \text{Var}(Y_{2j}) = \rho^2 \sigma_2^2 = \sigma_1^2 = \text{Var}(Y_{1j})$$

并且

$$E\left(\frac{X_j - \mu_{X_j}}{\sigma_{X_j}^2}\right)^4 = \rho^4 E(Y_{2j} - \mu_2)/(\rho^4 \sigma_2^4) = E(Y_{2j} - \mu_2)/\sigma_2^4 = \gamma$$

由于按假设, $E(Y_{1j} - \mu_1)^4/\sigma_1^4 = \gamma$, 它遵循以下原则, 两个样本 Y_{1j} 和 $X_j = \rho Y_{2j}$ 的父总体具有相同的方差 σ_1^2 和相同的峰度 γ 。按 Layard (1973), 基于两个样本 Y_{1j} 和 X_j 的有相合性的合并峰度估计量 $\hat{\gamma}$ 给定为

$$\begin{aligned} \hat{\gamma}' &= (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \sum_{j=1}^{n_2} (X_j - \bar{X})^4}{[(n_1 - 1)S_1^2 + (n_2 - 1)S_X^2]^2} \\ &= (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = \hat{\gamma}_P(\rho) \end{aligned}$$

视需求而定。

附录 B: 结果 2 的论证

我们已经确定, 基于 T_2 的 $\rho = \sigma_1/\sigma_2$ 的近似 $(1 - \alpha)100$ 百分比置信集合按如下给定

$$\left\{ \rho \in (0, \infty): (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \leq 0 \right\}$$

合并峰度估计量将表示为各个样本的峰度估计量, 按如下给定

$$\hat{\gamma}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1, 2$$

具体地说, 如果我们设 $t = \rho/\hat{\rho}$, 则

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - m_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = (n_1 + n_2) \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2}$$

其中, $K = (n_1 - 1)/(n_2 - 1)$ 。

结果是，可以将标准误的平方项表示为

$$\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2} - B$$

其中

$$A = \frac{(n_1 + n_2)(n_1 + n_2 - 2)}{(n_1 - 1)(n_2 - 1)}, B = \frac{k_1}{n_1 - 1} + \frac{k_2}{n_2 - 1}$$

因此，如果我们设 $r^2 = \rho^2/(c\hat{\rho}^2)$ ，则很容易得到

$$\begin{aligned} (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \\ = (\ln r^2)^2 - z_{\alpha/2}^2 \left(A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 c^2 r^4/n_2}{(K + c r^2)^2} - B \right) \end{aligned}$$

它遵循以下原则，基于 T_2 的对于 $\rho = \sigma_1/\sigma_2$ 的近似百分之 $(1 - \alpha)100$ 的置信集合可以给定

$$\hat{\rho}\sqrt{c} \{r \in (0, \infty): H(r^2) \leq 0\}$$

或等价地， $\rho^2 = \sigma_1^2/\sigma_2^2$ 的置信集合可以表示为

$$c\hat{\rho}^2 \{r \in (0, \infty): H(r) \leq 0\}$$

其中

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 se^2(cx), x > 0$$

并且

$$se^2(x) = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 x^2/n_2}{(K + x)^2} - B$$

附录 C: 结果 3 的论证

在 $H(0) = H(+\infty) = +\infty$ 和 $H(1) < 0$ 情况下，很容易验证函数 $H(x)$ 在正实线上的连续性。根据微分中值定理，函数 $H(x)$ 在区间 $(0, 1)$ 中至少有一个根，在区间 $(0, +\infty)$ 中至少有一个根。因此，如果函数 $H(x)$ 具有两个根，则一个根小于 1，另一个根大于 1。由于此函数开口向上，则在 r 介于这两个根之间时，满足不等式 $H(r) \leq 0$ 。这些根定义了 $\rho^2/(c\hat{\rho}^2)$ 的置信区间的端点。因此，如果我们设 $x_L < 1 < x_U$ 为两个根，则按结果 2，方差比值的置信下限 ρ^2 计算为 $c\hat{\rho}^2 x_L$ ，标准差比值的置信下限为 $\hat{\rho}\sqrt{c x_L}$ 。类似地，方差比值的置信上限为 $c\hat{\rho}^2 x_U$ ，标准差比值的置信上限为 $\hat{\rho}\sqrt{c x_U}$ 。

另一方面，如果函数 $H(x)$ 具有两个以上的根，则当 r 介于两个连续的根之间且函数开口向上时，满足不等式 $H(r) \leq 0$ 。因此，置信集合是一组非重叠区间。

8. 参考书

Balakrishnan, N. and Ma, C. W. (1990). A Comparative Study of Various Tests for the Equality of Two Population Variances. *Journal of Statistical Computation and Simulation*, 35, 41 - 89.

- Bonett D. G. (2006). Robust Confidence Interval for a Ratio of Standard Deviations. *Applied Psychological Measurements*, 30, 432 - 439.
- Boos, D. D. and Brownie, C. (1989). Bootstrap Methods for Testing Homogeneity of Variances. *Technometrics*, 31, 69 - 82.
- 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.
- Fligner, M. A. and Killeen, T. J. (1976). Distribution-Free Two-Sample Tests for Scale. *Journal of the American Statistical Association*, 71, 210 - 213.
- Hall, P. and Padmanabhan, A. R. (1997). Adaptive Inference for the Two-Sample Scale Problem. *Technometrics*, 39, 412 - 422.
- 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., *Contributions to Probability and Statistics*, Palo Alto, CA: Stanford University Press, 278 - 292.
- Lim, T.-S. and Loh, W.-Y. (1996). A Comparison of Tests of Equality of Variances. *Computational Statistics and Data Analysis*, 22, 287 - 301.
- Ott, L. (1993). *An Introduction to Statistical Methods and Data Analysis*, Belmont, CA: Duxbury Press.
- 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, 105 - 114.
- Wolfram, S. (1999). *The Mathematica Book*, 4th ed. Wolfram Media/Cambridge University Press.

© 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.