

单样本标准差检验

概述

单样本标准差检验用于估计过程的变化性并将变化性与目标值进行比较。通常，变化性使用方差（也即标准差）进行测量。

现在虽然研制出了许多用于评估总体方差的统计方法，但是每种方法都有各自的优缺点。用于检验方差的传统卡方检验方法可能是最常用的方法，但它对正态性假设极为敏感，并会在数据呈偏态或重尾分布时生成非常不准确的结果。同时还研制出了一些其他方法，但是这些方法也有缺点。例如，有些方法只适用于大样本或呈对称分布的数据（参见附录 A）。

在 Minitab 15 中，我们将使用另一种大样本方法，此方法是从样本方差的扩展卡方近似分布（由 Box 于 1953 年提出）派生的。这种方法被称作调整自由度 (AdjDF) 方法，对足够大的样本的正态假设不太敏感，已证明能生成比其他方法更准确的置信区间 (Hummel、Banga 和 Hettmansperger, 2005 年)。不过，Bonett (2006) 最近研制出了经过修订的统计方法，似乎能够提供更准确的近似置信区间。

在本文中，我们将评估 Bonett 方法的性能。此外，为了规划样本量，我们还研究了与 Bonett 的置信区间有关的等效检验程序的功效。根据我们的研究结果，我们使用 Bonett 方法来在 Assistant 中开展单样本标准差检验。我们还检查自动执行并在 Assistant Report Card 中显示的以下数据检查，并解释它们对结果造成的影响：

- 异常数据
- 检验的有效性
- 样本量

单样本标准差方法

Bonett 方法与 AdjDF 方法

在 Bonett 方法 (2006) 公布前, 对总体方差进行推测的最强大的程序可能莫过于 AdjDF 方法了。然而, Bonett 公布的结果证明, 当从非正态总体提取到大小适中的样本时, Bonett 方法提供接近目标水平的稳定的置信水平。因此, Bonett 方法可能是对总体的标准差或方差进行推测的优选方法。

目标

在对单个总体的方差进行推测时, 我们想要比较 Bonett 方法与 AdjDF 方法的性能。具体来说, 当从非正态总体中生成不同大小的样本时, 我们想要确定哪种方法将为方差 (或标准差) 生成更准确的置信区间。

我们比较置信区间是因为 Bonett 方法可直接应用于置信区间。可以得到与 Bonett 置信区间有关的等效假设检验程序。不过, 为了将我们的研究结果直接与 Bonett (2006) 中发布的结果进行比较, 我们检查了置信区间, 而不是假设检验。

方法

AdjDF 方法和 Bonett 方法都正式在附录 B 中定义。要比较每种方法的置信区间的准确性, 我们进行了下面的模拟。首先, 我们从特性各不相同的分布 (如偏态和重尾、对称和重尾以及对称和轻尾分布) 中生成了不同大小的随机样本。对于每个样本量, 我们从各个分布中提取了 10,000 个重复样本, 并使用各种方法为分布的真实方差计算 95% 的双侧置信区间。然后, 我们计算出了包含真实方差的 10,000 个区间的比例, 称为模拟覆盖概率。如果置信区间准确无误, 则模拟覆盖概率应接近 0.95 的目标范围概率。此外, 我们还计算了与每种方法的置信区间有关的平均宽度。如果这两种方法的置信区间具有大致相同的模拟覆盖概率, 则产生更短间隔 (平均而言) 的方法更准确。有关详细信息, 请参见附录 C。

结果

Bonett 方法通常会比 AdjDF 方法产生更好的覆盖概率和更准确的置信区间。因此, 采用 Bonett 方法, 方差统计检验产生的类型 I 和类型 II 错误率较低。因此, Assistant 中的单样本标准差检验将采用 Bonett 方法。

此外, 我们的研究结果表明, 对于中尾至重尾分布, Bonett 方法需要较大的样本量, 以达到准确的目标水平:

- 对于正态或轻尾分布, 20 个样本量就足够了。
- 对于中度重尾分布, 样本量至少应为 80 个。
- 对于重尾分布, 样本量至少应为 200 个。

因此, 要确保您数据的单样本标准差检验或置信区间结果有效, Assistant 中的数据检查功能可同时评估样本量和数据分布的尾部 (参见下文的检验数据检查的有效性)。

理论功效

Bonett 方法可直接应用于方差（或标准差）的置信区间。但是，使用假设检验和置信区间之间的统计关系，我们可以得到一个与 Bonett 的近似置信区间有关的等效检验。由于无法为此检验提供确切的功效，我们需要得到一个功效。此外，我们还要评估理论功效对正态性假设的敏感性。

目标

我们想要确定我们是否可以使用与 Bonett 置信区间有关的检验的理论功效来评估 Assistant 中的单样本标准差检验的能力和样本量要求。为此，在对正态和非正态数据进行分析时，我们需要评估这一理论功效是否准确地反映了检验的实际能力。

方法

有关使用 Bonett 方法的检验的理论功效，请参见附录 C。我们进行了模拟，以评估使用 Bonett 方法的实际能力水平（我们称之为模拟能力水平）。首先，我们从前面的研究所述的分布中生成了不同大小的随机样本：偏态和重尾、对称和重尾以及对称和轻尾分布。对于每一种分布，我们对每 10,000 个重复样本执行了检验。对于每种样本量，我们计算了检验的模拟能力，以检测给定的差值，并将其作为此检验对其有重要意义的 10,000 个样本的一部分。为了比较，我们还使用检验的理论功效计算了对应的能力水平。如果理论功效对正态性不是那么敏感，正态和非正态数据的理论和模拟能力水平应十分接近。有关详细信息，请参见附录 D。

结果

我们的模拟表明，当样本呈正态或轻尾分布时，使用 Bonett 方法的检验的理论和模拟能力几乎相同。当样本呈重尾分布时，理论功效可能会采取保守估计，但会高估达到给定能力所需的样本量。因此，检验的理论功效将确保样本量足够大，无论哪种分布都能够检测到标准差中的实际重要差值。然而，如果数据呈重尾分布，估计的样本量可能会大于实际需要的样本量，可能会在取样时导致超支。

数据检查

异常数据

异常数据是非常大或非常小的数据值，也称为异常值。异常数据会对分析结果产生巨大的影响，并且会影响发现具有重要统计意义的结果的概率，特别是当样本量较小时。异常数据可以表明数据收集问题或者您正在研究的过程的异常表现引起的问题。因此，这些数据点往往值得研究，如果可能的话应予以更正。

目标

我们想研制一种方法，来检查相对于总体样本而言非常大或非常小并且可能影响分析结果的数据值。



方法

我们研制出了一种方法，根据 Hoaglin, Iglewicz, and Tukey (1986) 介绍的用于识别箱线图上的异常值的方法来检查异常数据。

结果

如果数据点在分布的下限或上限范围内超过 1.5 倍的四分位范围，Assistant 会将该数据点识别为异常数据点。上、下四分位范围分别是数据的 25% 和 75%。四分位范围是两个四分位数之间的差值。即使有多个异常值，这种方法也应用得很好，因为它可以检测每一个具体的异常值。

检查异常数据时，Assistant 会在 Report Card 中显示以下状态指标：

状态	条件
	没有异常数据点。
	至少有一个异常数据点，可能会对结果产生巨大的影响。

检验的有效性

在上面的单样本标准差方法部分，我们发现，Bonett 方法通常提供比 AdjDF 方法更好的结果。然而，当呈重度重尾分布时，Bonett 方法需要较大的样本量，才能获得准确的结果。因此，评估检验的有效性的方法必须不仅取决于样本量，还取决于母体分布的尾部重量。Gel et al. (2007) 研制出了一种检验方法，用于确定样本是否呈重尾分布。该检验被称为 SJ 检验，结果取决于样本标准差与尾部估计值 J 之比（详见附录 E）。

目标

对于给定的数据样本，我们需要制定一个规则，通过评估数据中的尾部重量来评估 Bonett 方法的有效性。



方法

我们通过模拟来研究 SJ 检验的能力，以确定重尾分布。如果 SJ 检验适用于中等大小的样本，那么它可以为我们区分出重尾和轻尾分布。有关详细信息，请参见附录 F。

结果

我们的模拟表明，当样本足够大时，SJ 检验可用于区分重尾和轻尾分布。对于中等或大量样本，较小的 p 值表示重度重尾，较大的 p 值表示轻度轻尾。然而，由于大样本的 p 值往往比小样本的小，我们在确定尾部重量时也会考虑样本量。因此，我们研制出了一套规则，Assistant 可根据样本量和 SJ 检验的 p 值对每个样本的分布尾部进行分类。要查看 p 值的具体范围，以及与轻尾、中尾和重尾分布有关的样本量，请参见附录 F。

根据这些结果，Assistant Report Card 将显示以下状态指标，以评估数据的单样本标准差检验（Bonett 方法）的有效性：

状态	条件
	没有证据表明您的样本呈重尾分布。您的样本量足够大，可以可靠地检查此条件。 或 您的样本呈中度重尾或重尾分布。不过，您的样本量足够大，可以弥补差值，因此 p 值应该是准确的。
	您的样本呈中度重尾或重尾分布。您的样本量不够大，无法弥补差值。解释结果时应小心。 或 您的样本不够大，无法可靠地检查是否有重尾。解释结果时应小心。

样本量

通常情况下，执行统计假设检验可收集证据，以拒绝“无差值”零假设。如果样本量太小，检验可能无法检测到切实存在的差值，从而导致类型 II 错误。因此，一定要确保样本量足够大，可以高概率地检测到具有重要意义的实际差值。

目标

如果数据没有提供足够的证据来拒绝零假设，我们要确定样本量是否足够大，以便检验能够高概率地检测到所需的实际差值。尽管规划样本量的目标是确保样本量足够大，可以高概率地检测到重要差值，它们也不应该如此之大，使得无意义的差值高概率地成为具有重要统计意义的差值。






方法

单样本标准差检验的能力和样本量分析基于检验的理论功效。当数据接近正态分布的尾部或呈轻尾分布时，此功效可提供有用的估计值，但是当数据呈重尾分布时，可能会生成保守的估计值（参见上述单样本标准差方法部分中的理论功效中总结的模拟结果）。

结果

当数据未提供针对零假设的足够证据时，Assistant 使用正态近似检验的功效来计算可以用 80% 和 90% 的概率检测到的给定样本量的实际差值。此外，如果用户提供所需的特定实际差值，Assistant 会采用正态近似检验的功效，计算差值检测概率为 80% 和 90% 的样本量。

为了帮助解释结果，用于单样本标准差检验的 Assistant Report Card 在检查能力和样本量时，会显示以下状态指标：

状态	条件
	检验发现标准差与目标值之间的差值，因此能力不是问题。 或 能力是足够的。检验没有发现标准差与目标值之间的差值，但是样本足够大，提供至少 90% 的概率来检测给定差值。
	能力可能足够。检验没有发现标准差与目标值之间的差值，但是样本足够大，提供 80%~90% 的概率来检测给定差值。报告了实现 90% 的能力所需的样本量。
	能力可能不够。检验没有发现标准差与目标值之间的差值，并且样本足够大，提供 60%~80% 的概率来检测给定差值。报告了实现 80% 和 90% 的能力所需的样本量。
	能力不够。检验没有发现标准差与目标值之间的差值，并且样本不够大，无法提供至少 60% 的概率来检测给定差值。报告了实现 80% 和 90% 的能力所需的样本量。
	检验未发现标准差与目标值之间的差值。您没有指定要检测的实际差值。根据您的数据，该报告可能指示，您可以用 80% 和 90% 的概率，根据样本量和 alpha 值检测到的差值。

参考书

- Bonett, D.G. (2006)。非正态分布标准差的近似置信区间 (Approximate confidence interval for standard deviation of nonnormal distributions)。《统计计算与数据分析》(Computational Statistics & Data Analysis), 50, 775-782。
- Box, G.E.P. (1953)。非正态性和方差检验 (Non-normality and tests on variances)。《Biometrika》, 40, 318。
- Efron, B., & Tibshirani, R. J. (1993)。引导程序简介 (An introduction to the bootstrap)。Boca Raton, FL: Chapman and Hall/CRC。
- Gel, Y. R., Miao, W., & Gastwirth, J. L. (2007)。针对重尾的可靠的定向正态性检验 (Robust directed tests of normality against heavy-tailed alternatives)。《统计计算与数据分析》(Computational Statistics & Data Analysis), 51, 2734-2746。
- Hummel, R., Banga, S., & Hettmansperger, T.P. (2005)。随机样本中的方差的置信区间更好 (Better confidence intervals for the variance in a random sample)。《Minitab Technical Report》。
- Lee, S.J., & Ping, S. (1996)。检验对称重尾分布的方差。《统计计算与模拟杂志》(Journal of Statistical Computation and Simulation), 56, 39-52。

附录 A：检验方差（或标准差）的方法

下表总结了与检验方差有关的各种方法的优缺点。

方法	注释
传统的卡方程序	对正态性假设极为敏感。不论样本多大，即使正态性分布的小偏离也会生成不准确的结果。事实上，当数据偏离正态时，增加样本量会降低程序的准确性。
大样本方法基于样本方差对数变换的渐近正态分布	一般比传统的卡方方法好，但需要对正态假设不敏感的较大样本量。
大样本方法基于单侧（上尾）检验的 Edgeworth 扩展 参见 Lee and Ping (1996)。	产生可接受的类型 I 错误率，但要求数据呈对称分布。
大样本方法基于近似于样本方差分布的扩展的卡方分布。这种方法被称作调整自由度 (AdjDF)。 参见 Hummel, Banga, and Hettmansperger (2005)。	提供比基于样本方差对数变换的渐近正态分布方法以及置信区间的非参数 ABC 引导估算法 (Efron and Tibshirani, 1993) 更好的覆盖概率。 用于 Minitab 15 中的单方差检验。
Bonett 经调整的样本方差对数变换的渐近分布 参见 Bonett (2006)。	即使是中等大小的样本，也要为其置信区间提供有用的覆盖概率。不过，当数据呈重尾分布时，需要提供更大的样本。 用于 Minitab 16 中的单方差检验和 Assistant 单样本标准差检验。

附录 B: Bonett 方法和 AdjDF 方法的定义

使 x_1, \dots, x_n 成为有限四阶矩总体中大小为 n 的随机观测样本。使 \bar{x} 和 s 分别成为观测样本平均值和标准差。同时使 γ 和 γ_e 分别成为总体峰度和峰度超值，然后成为 $\gamma_e = \gamma - 3$ 。因此，对于正态总体，分别为 $\gamma = 3$ 和 $\gamma_e = 0$ 。同时使 σ^2 成为未知的总体方差。在下面的章节中，我们展示了两种有关 σ^2 的推测方法、调整自由度 (AdjDF) 方法，以及 Bonett 方法。

公式 B1: AdjDF 方法

AdjDF 方法基于近似于样本方差分布的扩展的卡方分布 (参见 Box, 1953)。更准确地说，样本方差的前两个阶矩都与扩展的卡方分布的阶矩匹配，用于确定未知比例和自由度。这种方法将为方差产生如下 $(1 - \alpha)100\%$ 的近似双侧置信区间：

$$\left[\frac{rs^2}{\chi_{r,\alpha/2}^2}, \frac{rs^2}{\chi_{r,1-\alpha/2}^2} \right]$$

其中

$$r = \frac{2n}{\hat{\gamma}_e + 2n/(n-1)}$$

$$\hat{\gamma}_e = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

峰度超值的这个估计值与 Minitab 中的 Basic Statistics 命令使用的值相同。

公式 B2: Bonett 方法

Bonett 方法依赖于众所周知的传统方法，即用中心极限定理和 Cramer δ 方法获得样本方差对数变换的渐近分布。对数变换用于加快正态性的收敛速度。使用这种方法，方差 $(1 - \alpha)100\%$ 的近似双侧置信区间被定义为：

$$\left[s^2 \exp(-z_{\alpha/2} se), s^2 \exp(z_{\alpha/2} se) \right]$$

其中 z_α 是标准正态分布的上百分位，并且 se 是对数变换样本方差的标准误差的渐近估计值，可表示为：

$$se = \sqrt{\frac{\hat{\gamma} - (n-3)/n}{n-1}} = \sqrt{\frac{\hat{\gamma}_e + 2 + 3/n}{n-1}}$$

以前，Hummel et al. (2005) 曾进行过模拟研究，证明了 AdjDF 方法优于这个传统方法。不过，Bonett 为克服其缺点对此传统方法做出了两项调整。第一个调整涉及峰度估计。Bonett 使用以下公式来估计峰度：

$$\hat{\gamma}_e = \frac{n}{(n-1)^2} \sum_{i=1}^n \left(\frac{x_i - m}{s} \right)^4 - 3$$

其中 m 是调整后的平均值，调整比例等于 $1/2\sqrt{n-4}$ 。此峰度估计容易提高重尾（对称或偏态）分布置信水平的准确性。

对于第二个调整，Bonett 凭经验确定样本方差和标准误差的固定乘数。当样本量较小时，此固定乘数约等于尾部概率，并且可表示为：

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

这些调整为方差产生 Bonett 的 $(1 - \alpha)100\%$ 的近似双侧置信区间：

$$\left[cs^2 \exp(-c z_{\alpha/2} se), cs^2 \exp(c z_{\alpha/2} se) \right]$$

附录 C: Bonett 方法与 AdjDF 方法的准确性

模拟 C1: 置信区间比较

我们想要比较使用 AdjDF 方法和 Bonett 方法计算出的方差的置信区间的准确性。我们从几个分布生成不同大小的随机样本 ($n = 20, 30, 40, 50, 60, 80, 100, 150, 200, 250, 300$) 并使用各种方法计算置信区间。此分布包括:

- 标准正态分布 ($N(0, 1)$)
- 对称和轻尾分布, 包括均匀分布 ($U(0, 1)$) 和两个参数均设置为 3 ($B(3, 3)$) 的 Beta 分布
- 对称和重尾分布, 包括自由度为 5 和 10 的 t 分布 ($t(5), t(10)$) 以及位置为 0、标度为 1 ($Lp1$) 的 Laplace 分布
- 偏态和重尾分布, 包括标度为 1 (Exp) 的指数分布以及自由度为 3、5 和 10 的卡方分布 ($Chi(3), Chi(5), Chi(10)$)
- 左偏和重尾分布; 具体来说, 即参数分别设置为 8 和 1 的 Beta 分布 ($B(8, 1)$)

此外, 为了评估异常值的直接影响, 我们从受污染的正态分布中生成了样本, 定义如下:

$$CN(p, \sigma) = pN(0, 1) + (1 - p)N(0, \sigma)$$

其中 p 是混合参数, $1 - p$ 是污染比例 (等于异常值的比例)。我们选择了两个受污染的正态总体作为研究对象: 一个是 $CN(0.9, 3)$, 其中 10% 的总体是异常值; 另一个是 $CN(0.8, 3)$, 其中 20% 的总体是异常值。这两个分布是对称的, 并且由于异常值呈长尾分布。

对于各个样本量, 将从每个分布中提取 10,000 个重复样本, 并使用各种方法计算 95% 的双侧置信区间。随机样本生成器被设定为种子, 以便将这两种方法应用于同一样本。根据这些置信区间, 我们再计算出每种方法的模拟覆盖概率 (CovP) 和平均间隔宽度 (AveW)。如果这两种方法的置信区间具有大致相同的模拟覆盖概率, 则产生更短间隔 (平均而言) 的方法更准确。因为我们使用 95% 的目标置信水平, 模拟误差为 $\sqrt{0.95(0.05)/10,000} = 0.2\%$ 。

模拟结果被记录在表 1 和表 2 中, 如下所示。

表 1 使用 AdjDF 和 Bonett 方法计算出的方差的 95% 的双侧置信区间的模拟覆盖概率。这些样本从轻尾、正态、接近正态或重尾的对称分布中生成。

分布		轻尾、正态或接近正态的对称分布				重尾的对称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(0.8, 3)	CN(0.9, 3)	T(5)
偏度		0	0	0	0	0	0	0	0
峰度 (γ_e)		-1.20	-0.667	0	1.00	3.00	4.544	5.333	6.00
n = 10									
AdjDF	CovP	0.91	0.909	0.903	0.883	0.853	0.793	0.815	0.858
	AveW	0.154	0.087	3.276	5.16	13.924	21.658	14.913	11.742
Bonett	CovP	0.972	0.967	0.962	0.952	0.919	0.891	0.92	0.935
	AveW	0.242	0.115	3.71	5.134	10.566	15.335	10.367	8.578
n = 20									
AdjDF	CovP	0.937	0.937	0.923	0.909	0.881	0.819	0.817	0.868
	AveW	0.080	0.045	1.572	2.463	5.781	9.265	6.539	5.151
Bonett	CovP	0.953	0.954	0.946	0.934	0.909	0.856	0.864	0.904
	AveW	0.10	0.051	1.683	2.422	4.932	7.282	4.945	4.026
n = 30									
AdjDF	CovP	0.946	0.942	0.933	0.917	0.894	0.851	0.823	0.882
	AveW	0.061	0.034	1.17	1.764	4.117	6.33	4.557	3.667
Bonett	CovP	0.951	0.95	0.947	0.933	0.909	0.869	0.852	0.907
	AveW	0.070	0.037	1.221	1.75	3.654	5.383	3.736	2.997
n = 40									
AdjDF	CovP	0.953	0.947	0.932	0.922	0.904	0.867	0.833	0.89
	AveW	0.051	0.028	0.971	1.489	3.246	5.131	3.654	3.024
Bonett	CovP	0.954	0.951	0.941	0.936	0.914	0.879	0.856	0.907
	AveW	0.057	0.030	1.002	1.469	2.994	4.519	3.128	2.542
n = 50									
AdjDF	CovP	0.951	0.945	0.937	0.925	0.911	0.878	0.838	0.893
	AveW	0.045	0.025	0.849	1.291	2.789	4.357	3.091	2.603
Bonett	CovP	0.951	0.947	0.944	0.938	0.918	0.888	0.855	0.908
	AveW	0.049	0.026	0.87	1.28	2.613	3.939	2.729	2.24

分布	轻尾、正态或接近正态的对称分布				重尾的对称分布			
	U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(0.8, 3)	CN(0.9, 3)	T(5)
偏度	0	0	0	0	0	0	0	0
峰度 (γ_e)	-1.20	-0.667	0	1.00	3.00	4.544	5.333	6.00

n = 60

AdjDF	CovP	0.949	0.943	0.938	0.926	0.913	0.89	0.853	0.899
	AveW	0.040	0.022	0.766	1.155	2.49	3.857	2.768	2.283
Bonett	CovP	0.949	0.947	0.943	0.935	0.918	0.896	0.868	0.91
	AveW	0.043	0.023	0.781	1.147	2.354	3.552	2.498	2.023

n = 70

AdjDF	CovP	0.948	0.945	0.94	0.93	0.913	0.89	0.858	0.896
	AveW	0.037	0.020	0.701	1.056	2.283	3.458	2.475	2.049
Bonett	CovP	0.947	0.946	0.944	0.938	0.918	0.894	0.868	0.905
	AveW	0.039	0.021	0.713	1.049	2.174	3.227	2.272	1.828

n = 80

AdjDF	CovP	0.947	0.949	0.938	0.929	0.918	0.905	0.869	0.902
	AveW	0.034	0.019	0.652	0.988	2.089	3.205	2.30	1.906
Bonett	CovP	0.946	0.95	0.942	0.935	0.923	0.907	0.877	0.911
	AveW	0.036	0.019	0.662	0.982	2.005	3.014	2.133	1.716

n = 90

AdjDF	CovP	0.946	0.947	0.948	0.929	0.918	0.908	0.869	0.901
	AveW	0.032	0.018	0.611	0.921	1.951	2.982	2.124	1.874
Bonett	CovP	0.945	0.948	0.952	0.936	0.92	0.91	0.874	0.909
	AveW	0.034	0.018	0.618	0.916	1.882	2.822	1.984	1.646

n = 100

AdjDF	CovP	0.947	0.951	0.945	0.933	0.92	0.91	0.885	0.912
	AveW	0.030	0.017	0.576	0.873	1.83	2.801	2.017	1.658
Bonett	CovP	0.946	0.953	0.948	0.937	0.923	0.912	0.891	0.916
	AveW	0.032	0.017	0.583	0.869	1.772	2.666	1.899	1.522

分布	轻尾、正态或接近正态的对称分布				重尾的对称分布			
	U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(0.8, 3)	CN(0.9, 3)	T(5)
偏度	0	0	0	0	0	0	0	0
峰度 (γ_e)	-1.20	-0.667	0	1.00	3.00	4.544	5.333	6.00

n = 150

AdjDF	CovP	0.949	0.951	0.947	0.936	0.932	0.925	0.896	0.912
	AveW	0.024	0.014	0.464	0.70	1.47	2.228	1.602	1.325
Bonett	CovP	0.948	0.952	0.949	0.939	0.933	0.924	0.898	0.915
	AveW	0.025	0.014	0.467	0.698	1.438	2.156	1.539	1.251

n = 200

AdjDF	CovP	0.943	0.949	0.948	0.938	0.927	0.93	0.914	0.918
	AveW	0.021	0.012	0.40	0.605	1.265	1.906	1.373	1.178
Bonett	CovP	0.942	0.951	0.949	0.94	0.928	0.93	0.915	0.92
	AveW	0.021	0.012	0.402	0.603	1.245	1.86	1.333	1.106

n = 250

AdjDF	CovP	0.952	0.952	0.949	0.942	0.938	0.929	0.909	0.915
	AveW	0.019	0.010	0.355	0.538	1.12	1.69	1.219	1.037
Bonett	CovP	0.951	0.952	0.949	0.944	0.941	0.929	0.909	0.916
	AveW	0.019	0.010	0.357	0.537	1.106	1.657	1.19	0.986

n = 300

AdjDF	CovP	0.95	0.948	0.951	0.94	0.938	0.936	0.92	0.914
	AveW	0.017	0.009	0.324	0.49	1.019	1.544	1.115	0.933
Bonett	CovP	0.95	0.947	0.951	0.942	0.937	0.929	0.92	0.916
	AveW	0.017	0.010	0.325	0.489	1.009	1.657	1.093	0.897

表 2 使用 AdjDF 和 Bonett 方法计算出的方差的 95% 的双侧置信区间的模拟覆盖概率。这些样本从接近正态、中度重尾或重尾的偏态分布中生成。

分布		接近正态或中度重尾的偏态分布			重尾偏态分布	
		Chi (10)	B (8, 1)	Chi (5)	Chi (3)	期望值
偏度		0.894	-1.423	1.265	1.633	2
峰度 (γ_e)		1.20	2.284	2.40	4.00	6
n = 10						
AdjDF	CovP	0.869	0.815	0.836	0.797	0.758
	AveW	93.383	0.065	61.994	47.821	10.711
Bonett	CovP	0.95	0.917	0.938	0.911	0.882
	AveW	91.006	0.058	53.83	38.137	7.498
n = 20						
AdjDF	CovP	0.889	0.862	0.862	0.833	0.811
	AveW	41.497	0.026	25.479	20.099	4.293
Bonett	CovP	0.932	0.912	0.913	0.893	0.877
	AveW	41.60	0.026	24.094	17.232	3.37
n = 30						
AdjDF	CovP	0.901	0.881	0.88	0.864	0.838
	AveW	30.021	0.018	18.182	13.63	2.844
Bonett	CovP	0.931	0.92	0.914	0.906	0.885
	AveW	30.462	0.019	17.858	12.634	2.441
n = 40						
AdjDF	CovP	0.909	0.882	0.885	0.867	0.862
	AveW	24.459	0.015	14.577	10.649	2.193
Bonett	CovP	0.93	0.915	0.913	0.904	0.898
	AveW	24.952	0.015	14.504	1.991	1.991
n = 50						
AdjDF	CovP	0.912	0.90	0.892	0.871	0.868
	AveW	21.373	0.013	12.694	9.115	1.861
Bonett	CovP	0.93	0.927	0.916	0.903	0.901
	AveW	21.814	0.013	12.741	8.897	1.735

分布	接近正态或中度重尾的偏态分布			重尾偏态分布	
	Chi (10)	B(8, 1)	Chi (5)	Chi (3)	期望值
偏度	0.894	-1.423	1.265	1.633	2
峰度 (γ_e)	1.20	2.284	2.40	4.00	6

$n = 60$

AdjDF	CovP	0.915	0.908	0.901	0.89	0.875
	AveW	18.928	0.011	11.338	8.211	1.645
Bonett	CovP	0.93	0.933	0.923	0.917	0.90
	AveW	19.369	0.012	11.456	8.093	1.554

$n = 70$

AdjDF	CovP	0.915	0.91	0.904	0.898	0.881
	AveW	17.513	0.010	10.307	7.461	1.488
Bonett	CovP	0.932	0.932	0.922	0.919	0.906
	AveW	17.906	0.011	10.464	7.408	1.429

$n = 80$

AdjDF	CovP	0.92	0.916	0.911	0.904	0.89
	AveW	16.157	0.009	9.604	6.892	1.349
Bonett	CovP	0.935	0.936	0.929	0.924	0.915
	AveW	16.537	0.010	9.765	6.882	1.314

$n = 90$

AdjDF	CovP	0.924	0.918	0.911	0.897	0.894
	AveW	15.25	0.009	9.007	6.323	1.255
Bonett	CovP	0.938	0.936	0.929	0.918	0.913
	AveW	15.609	0.009	9.175	6.366	1.23

$n = 100$

AdjDF	CovP	0.926	0.919	0.915	0.908	0.895
	AveW	14.332	0.008	8.451	6.016	1.171
Bonett	CovP	0.935	0.936	0.931	0.924	0.916
	AveW	14.664	0.009	8.625	6.063	1.158

分布	接近正态或中度重尾的偏态分布			重尾偏态分布	
	Chi (10)	B (8, 1)	Chi (5)	Chi (3)	期望值
偏度	0.894	-1.423	1.265	1.633	2
峰度 (γ_e)	1.20	2.284	2.40	4.00	6

n = 150

AdjDF	CovP	0.933	0.925	0.923	0.913	0.911
	AveW	11.606	0.007	6.781	4.792	0.933
Bonett	CovP	0.943	0.941	0.936	0.929	0.928
	AveW	11.846	0.007	6.942	4.875	0.937

n = 200

AdjDF	CovP	0.935	0.934	0.926	0.916	0.915
	AveW	9.973	0.006	5.849	4.127	0.799
Bonett	CovP	0.942	0.948	0.936	0.93	0.931
	AveW	10.185	0.006	5.991	4.212	0.808

n = 250

AdjDF	CovP	0.938	0.939	0.934	0.926	0.922
	AveW	8.899	0.005	5.231	3.652	0.705
Bonett	CovP	0.946	0.951	0.944	0.936	0.931
	AveW	9.078	0.005	5.355	3.735	0.716

n = 300

AdjDF	CovP	0.942	0.938	0.934	0.931	0.922
	AveW	8.156	0.005	4.749	3.344	0.64
Bonett	CovP	0.947	0.948	0.943	0.941	0.933
	AveW	8.314	0.005	4.862	3.419	0.651

我们的结果与 Bonett (2006) 发布的那些结果非常一致。正如表 1 和 2 中所示, 使用 Bonett 方法计算出的置信区间均优于使用 AdjDF 方法计算出的置信区间, 因为平均而言, 它们产生的覆盖概率接近 0.95 的目标水平和较窄的置信区间。如果这两种方法的置信区间具有大致相同的模拟覆盖概率, 则产生更短间隔 (平均而言) 的方法更准确。这意味着, 基于 Bonett 方法的方差的统计检验表现更好, 并会产生较低的类型 I 和 II 错误率。当样本量较大时, 这两种方法将得到几乎相同的结果, 但是对于小到中等大小的样本量, Bonett 方法更胜一筹。

虽然 Bonett 方法通常比 AdjDF 方法表现得更好, 但即使对于非常大的样本 ($n > 100$), 它也总是会为重尾分布 (对称或偏态) 产生比 0.95 的目标覆盖率低的覆盖概率。这一点在下面

的图 1 中做出了详细说明，其中描绘了 Bonett 方法根据小、中、大样本量真正的总体峰度超值得出的模拟覆盖概率。

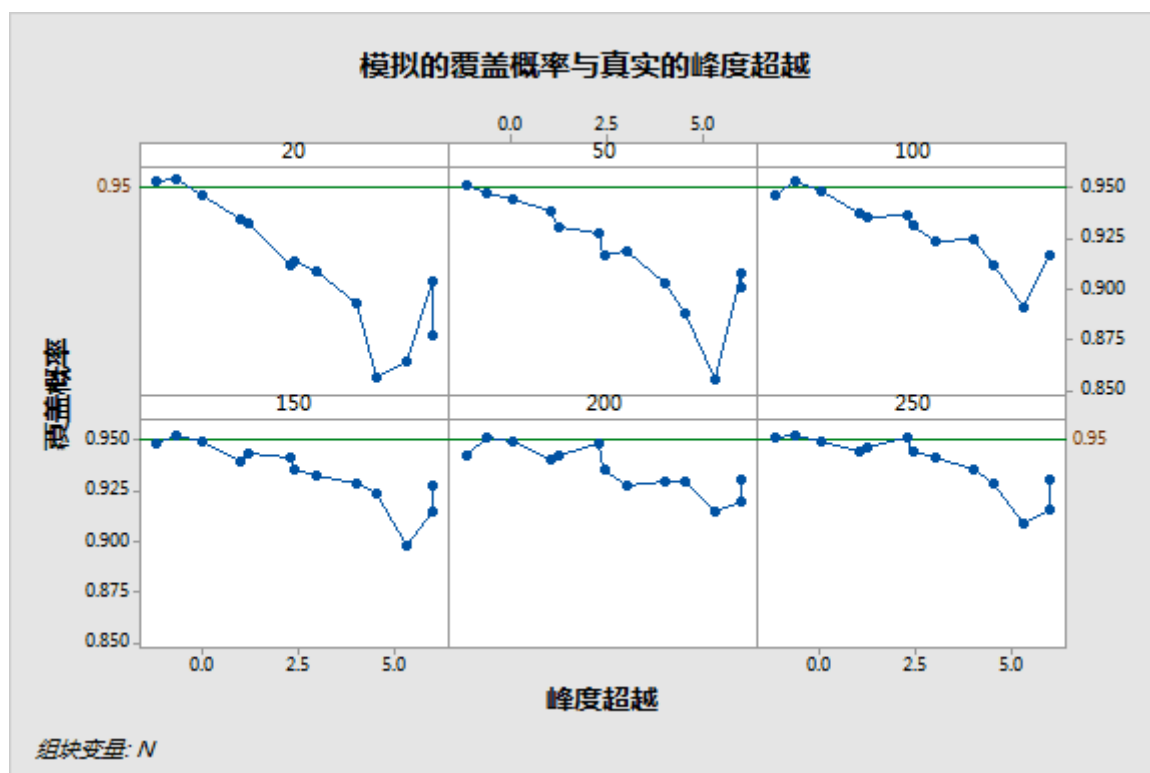


图 1 根据不同样本量的每个分布的峰度超值绘制的 Bonett 的 95% 的置信区间的模拟覆盖概率。

如图 1 中所示，峰度越大，模拟覆盖概率接近目标水平所需的样本量就越大。如前所述，对于重尾分布，用 Bonett 方法估计的模拟覆盖概率较低。不过，对于轻度轻尾分布，例如均匀分布和 Beta (3, 3) 分布，小至 20 的样本量的模拟覆盖概率既稳定又正中目标。因此，我们的基本标准是要同时根据样本量及从中提取样本的分布的尾部重量来确定 Bonett 方法的有效性。

作为制定这一标准的第一步，我们根据尾部重量将分布分成以下三类：

- **轻尾或正态分布 (L 型)**：这些是 Bonett 的置信区间产生接近目标覆盖水平的稳定的覆盖概率的分布。对于这些分布，低至 20 的样本量可产生准确的结果。示例包括均匀分布、Beta(3, 3) 分布、正态分布、自由度为 10 的 t 分布，以及自由度为 10 的卡方分布。
- **中度重尾分布 (M 类型)**：对于这些分布，Bonett 方法需要至少为接近目标范围的模拟覆盖概率提供 80 个样本量。示例包括自由度为 5 的卡方分布和 Beta(8, 1) 分布。
- **重尾分布 (H 类型)**：这些是 Bonett 的置信区间产生的覆盖概率远远低于目标范围的分布，除非样本量非常大 ($n \geq 200$)。示例包括自由度为 5 的 t 分布、Laplace 分布、自由度为 3 的卡方分布、指数分布以及两个受污染的正态分布、CN(0.9, 3) 和 CN(0.8, 3)。

因此，评估 Bonett 方法的有效性的一般规则要求我们开发一个程序，以识别样本数据呈哪 3 种分布类型。为了评估检验数据检查的有效性，我们开发了 this 程序。有关详细信息，请参见附录 E。

附录 D：理论功效

我们得出了与 Bonett 方法关联的检验的理论功效，并进行了模拟，以比较检验的理论功效和模拟功效。如果理论功效和模拟功效曲线彼此接近，那么基于理论功效的能力和样本量分析应会产生准确的结果。

公式 D1：Bonett 方法的理论功效

如前所述，Bonett 方法基于众所周知的传统方法，即用中心极限定理和 Cramer δ 方法获得样本方差对数变换的渐近分布。更准确地说，在大样本中，公认 $\frac{\ln S^2 - \ln \sigma^2}{se}$ 大致呈标准正态分布。分母 se 是对数变换样本方差的大样本标准误差，并且可表示为：

$$se = \sqrt{\frac{\gamma - (n-3)/n}{n-1}}$$

其中 γ 是未知母体的峰度。

由此可见，使用 Bonett 方法的双侧检验的近似 α 水平的近似功效可表示为样本量的功效，比率 $\rho = \sigma/\sigma_0$ 和母体峰度 γ 可表示为：

$$\pi(n, \rho, \gamma) = 1 - \Phi\left(z_{\alpha/2} - \frac{\ln \rho^2}{\sqrt{\frac{\gamma - 1 + 3/n}{n-1}}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\ln \rho^2}{\sqrt{\frac{\gamma - 1 + 3/n}{n-1}}}\right)$$

其中 σ_0 是未知标准差的推测值， Φ 是标准正态分布的 CDF，并且 z_α 是标准正态分布的上 α 百分位点。也可以从这些计算中取得单侧功效。

注意：在规划样本量以供研究时，可以用峰度的估计值代替真正的峰度。此估计值通常取决于专家意见或先前的实验结果。如果这些信息不可用，进行小规模试点研究，以制定主要研究计划，往往是一种不错的做法。使用从试点研究中得到的样本，峰度可估计为

$$\hat{\gamma} = \frac{n}{(n-1)^2} \sum_{i=1}^n \left(\frac{x_i - m}{s}\right)^4$$

其中 m 是调整后的平均值，调整比例等于 $1/2\sqrt{n-4}$ 。

模拟 D1：实际能力与理论能力的比较

我们设计了一种模拟方法，以在使用 Bonett 方法检验方差时，将估计的实际能力水平（称为模拟水平）与理论能力水平（称为近似能力水平）进行比较。

在每个实验中，我们从每个分布中生成了 10,000 个重复样本，每个大小为 n ，其中 $n = 20, 30, 40, 50, \dots, 120$ 在模拟 C1（参见附录 C）中有介绍。对于每种分布和样本量 n ，我们计算了模拟能力水平，并将其作为 α 水平为 $\alpha = 0.05$ 的双侧检验对其具有重要意义 10,000 个随机重复样本的一部分。在计算模拟能力时，我们使用 $\rho = \sigma/\sigma_0 = 1.25$ 来获取相对较小的能力水平。然后，我们使用理论功效计算出相应的能力水平，以供比较。

结果显示在表 3 和表 4 中并以图形方式显示在下面的图 2 中。

表 3 根据 Bonett 方法与理论（正态近似）能力水平进行比较得到的方差双侧检验的模拟能力水平（评估时间为 $\rho = \sigma/\sigma_0 = 1.25$ ）。从轻尾、正态、接近正态或重尾的对称分布生成样本。

n	能力	轻尾、正态或接近正态的对称分布				重尾的对称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(.8, 3)	CN(.9, 3)	t(5)
20	相同值	0.521	0.39	0.31	0.237	0.178	0.152	0.139	0.172
	近似值	0.514	0.359	0.264	0.195	0.137	0.117	0.109	0.104
30	相同值	0.707	0.551	0.441	0.337	0.225	0.186	0.169	0.228
	近似值	0.717	0.519	0.382	0.276	0.186	0.154	0.143	0.135
40	相同值	0.831	0.679	0.526	0.427	0.285	0.266	0.203	0.285
	近似值	0.846	0.651	0.49	0.356	0.236	0.192	0.176	0.165
50	相同值	0.899	0.753	0.621	0.505	0.332	0.255	0.238	0.34
	近似值	0.921	0.754	0.586	0.431	0.284	0.229	0.21	0.196
60	相同值	0.942	0.822	0.701	0.57	0.38	0.285	0.274	0.384
	近似值	0.961	0.83	0.668	0.501	0.332	0.266	0.243	0.227
70	相同值	0.964	0.866	0.757	0.632	0.424	0.327	0.314	0.439
	近似值	0.981	0.885	0.737	0.566	0.379	0.303	0.276	0.257
80	相同值	0.981	0.909	0.815	0.689	0.481	0.372	0.347	0.483
	近似值	0.991	0.923	0.794	0.624	0.423	0.34	0.309	0.288
90	相同值	0.988	0.937	0.851	0.724	0.514	0.40	0.377	0.523
	近似值	0.996	0.95	0.84	0.676	0.467	0.375	0.342	0.318
100	相同值	0.994	0.961	0.88	0.779	0.558	0.43	0.411	0.566
	近似值	0.998	0.967	0.876	0.722	0.508	0.41	0.373	0.347
110	相同值	0.997	0.967	0.909	0.803	0.591	0.471	0.449	0.592
	近似值	0.999	0.979	0.905	0.763	0.547	0.443	0.404	0.376
120	相同值	0.999	0.982	0.929	0.844	0.629	0.502	0.476	0.63
	近似值	1.00	0.987	0.928	0.799	0.584	0.476	0.434	0.405

表 4 根据 Bonett 方法与理论（正态近似）能力水平比较得到的方差双侧检验的模拟能力水平（评估时间为 $\rho = \sigma/\sigma_0 = 1.25$ ）。从接近正态、中度重尾或重尾的偏态分布生成样本。

n	能力	接近正态或中度重尾的偏态分布			重尾的偏态分布	
		Chi (10)	B (8, 1)	Chi (5)	Chi (3)	期望值
20	相同值	0.222	0.166	0.172	0.139	0.128
	近似值	0.186	0.152	0.149	0.123	0.104
30	相同值	0.314	0.216	0.234	0.19	0.151
	近似值	0.263	0.263	0.205	0.164	0.135
40	相同值	0.387	0.266	0.292	0.223	0.186
	近似值	0.338	0.266	0.261	0.204	0.165
50	相同值	0.455	0.324	0.349	0.263	0.208
	近似值	0.409	0.323	0.316	0.245	0.196
60	相同值	0.521	0.376	0.399	0.302	0.239
	近似值	0.477	0.377	0.369	0.286	0.227
70	相同值	0.583	0.419	0.463	0.361	0.269
	近似值	0.539	0.43	0.42	0.325	0.257
80	相同值	0.646	0.473	0.499	0.394	0.299
	近似值	0.597	0.479	0.469	0.365	0.288
90	相同值	0.688	0.517	0.561	0.428	0.327
	近似值	0.649	0.526	0.516	0.403	0.318
100	相同值	0.738	0.561	0.591	0.469	0.368
	近似值	0.695	0.571	0.56	0.44	0.347
110	相同值	0.779	0.608	0.637	0.495	0.394
	近似值	0.737	0.611	0.60	0.475	0.376
120	相同值	0.81	0.635	0.679	0.538	0.416
	近似值	0.774	0.65	0.638	0.509	0.405

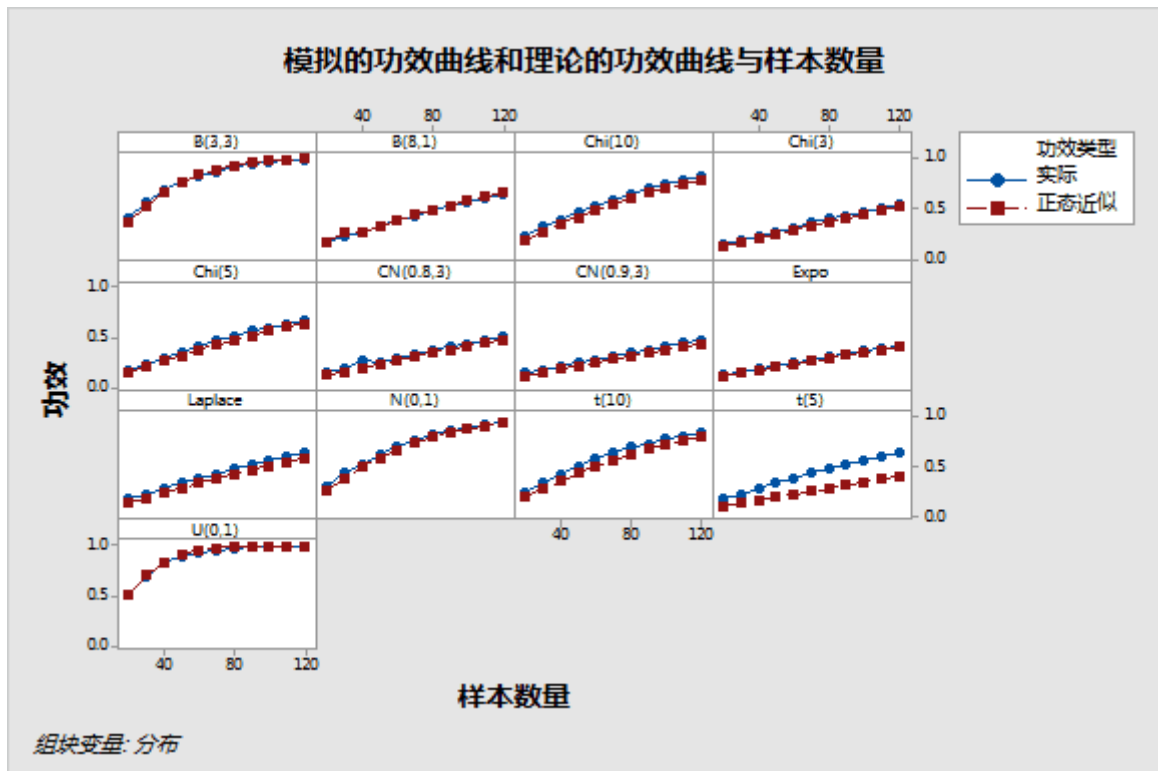


图 2 与各种分布的理论能力曲线比较的模拟能力曲线

表 3、表 4 和图 2 中的结果表明，当从诸如均匀分布、Beta (3, 3) 分布、正态分布、自由度为 10 的 t 分布、自由度为 10 的卡方分布等轻度轻尾分布（L 型分布，参见附录 C 中的定义）中生成样本时，理论能力值和模拟能力水平实际上是无法分清的。

不过，对于重尾分布（H 型分布），当样本量较小时，模拟能力曲线将显著高于理论能力曲线。这些重尾分布包括自由度为 5 的 t 分布、Laplace 分布、自由度为 3 的卡方分布、指数分布以及两个受污染的正态分布、CN(0.9, 3) 和 CN(0.8, 3)。因此，在规划样本量以供研究且样本呈重尾分布时，通过理论功效估计的样本量可能比实现给定目标能力所需的实际样本量大。

附录 E：正态与重尾分布的 SJ 检验

附录 C 中模拟研究的结果表明，当呈重度重尾分布时，Bonett 的置信区间的模拟覆盖概率需要较大的样本量，以接近目标水平。但是，偏度似乎没有表现出对模拟覆盖概率产生显著的影响。

因此，我们需要制定一个标准，以根据两个样本量和从中提取样本的分布的尾部重量来评估 Bonett 方法的有效性。幸运的是，Gel et al. (2007) 提供了相当有效的检验方法，以根据重尾分布的另类假设直接检验正态尾部分布的零假设。我们将该检验称为 SJ 检验，将围绕以下统计数据展开：

$$\hat{R} = \frac{S}{\hat{j}}$$

其中 S 是样本标准差， \hat{j} 是样本平均绝对值与中间值之间的偏差估计值， m 可表示为：

$$\hat{j} = \frac{\sqrt{\pi/2}}{n} \sum_{i=1}^n |X_i - m|$$

在以下情况下，针对重尾分布的另类假设的近似 α 样本量检验将拒绝正态分布尾部的零假设：

$$\frac{\sqrt{n}(\hat{R} - 1)}{\sigma_R} \geq z_\alpha$$

其中 z_α 是标准正态分布和 $\sigma_R = (\pi - 3)/2$ 的上 α 百分位。

Gel et al. (2007) 表明，用自由度为 $(\sqrt{n} + 3)/2$ 的 t 分布的上百分份替代标准正态分布的上 α 百分位，将为中等大小的样本量提供更准确的近似值。因此，当用 SJ 检验来评估检验数据检查的有效性时，我们会将 z_α 替换为 $t_{d,\alpha}$ ，即自由度为 $d = (\sqrt{n} + 3)/2$ 的 t 分布的上 α 百分位。

附录 F：检验的有效性

模拟 F1：使用 SJ 检验的模拟能力确定分布分类

我们进行模拟，以调查 SJ 检验的能力。我们从各种分布中生成了各种大小 ($n = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200$) 的样本。分布包括正态、轻尾、中尾或重尾分布，与模拟 C1 中所述相同（参见附录 C）。对于各个给定样本量，将从每个分布中提取 10,000 个重复样本。我们计算了 SJ 检验的模拟能力，以零假设（即正态母体分布）被拒绝的个案比例表示。此外，我们还计算了 R 平均值 (AveR) 和 p 平均值 (AvePV)。

模拟结果显示在下面的表 5 和表 6 中。

表 5 SJ 检验的模拟能力水平。从轻尾、正态、接近正态或重尾的对称分布生成样本。

分布		轻尾、正态或接近正态的对称分布				重尾的对称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(.8, 3)	CN(.9, 3)	t(5)
n	TrueR	0.921	0.965	1.0	1.032	1.128	1.152	1.118	1.085
10	能力	0.021	0.041	0.075	0.103	0.249	0.264	0.198	0.161
	平均值	1.010	1.036	1.060	1.073	1.129	1.131	1.106	1.096
	P 平均值	0.482	0.401	0.341	0.314	0.219	0.228	0.272	0.278
15	能力	0.009	0.027	0.071	0.121	0.35	0.389	0.283	0.215
	平均值	0.986	1.018	1.043	1.063	1.13	1.14	1.11	1.093
	P 平均值	0.572	0.44	0.357	0.302	0.171	0.181	0.24	0.247
20	能力	0.002	0.016	0.066	0.144	0.428	0.465	0.331	0.253
	平均值	0.966	1.001	1.030	1.054	1.127	1.137	1.104	1.086
	P 平均值	0.669	0.503	0.382	0.311	0.147	0.161	0.236	0.244
25	能力	0.002	0.011	0.065	0.153	0.50	0.55	0.397	0.293
	平均值	0.959	0.995	1.025	1.050	1.128	1.141	1.107	1.086
	P 平均值	0.721	0.535	0.391	0.305	0.12	0.128	0.208	0.223
30	能力	0.001	0.010	0.060	0.17	0.561	0.603	0.431	0.334
	平均值	0.951	0.989	1.019	1.046	1.127	1.141	1.106	1.084
	P 平均值	0.773	0.57	0.409	0.304	0.103	0.112	0.197	0.209
40	能力	0.00	0.006	0.058	0.19	0.665	0.709	0.513	0.401
	平均值	0.944	0.984	1.015	1.043	1.126	1.145	1.109	1.084
	P 平均值	0.84	0.616	0.42	0.287	0.073	0.076	0.162	0.179

分布		轻尾、正态或接近正态的对称分布				重尾的对称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(.8, 3)	CN(.9, 3)	t(5)
n	TrueR	0.921	0.965	1.0	1.032	1.128	1.152	1.118	1.085
50	能力	0.00	0.004	0.058	0.208	0.746	0.785	0.59	0.462
	平均值	0.939	0.98	1.012	1.040	1.126	1.146	1.111	1.084
	P 平均值	0.886	0.654	0.427	0.279	0.053	0.055	0.131	0.156
60	能力	0.00	0.002	0.060	0.231	0.813	0.836	0.647	0.518
	平均值	0.936	0.978	1.010	1.039	1.127	1.146	1.112	1.084
	P 平均值	0.913	0.686	0.43	0.267	0.039	0.039	0.109	0.134
70	能力	0.00	0.002	0.054	0.247	0.863	0.879	0.702	0.554
	平均值	0.934	0.975	1.009	1.037	1.127	1.147	1.112	1.083
	P 平均值	0.935	0.716	0.437	0.259	0.028	0.029	0.091	0.123
80	能力	0.00	0.001	0.054	0.265	0.896	0.912	0.729	0.591
	平均值	0.933	0.974	1.007	1.037	1.128	1.147	1.111	1.083
	P 平均值	0.95	0.74	0.44	0.241	0.021	0.021	0.079	0.105
90	能力	0.00	0.001	0.054	0.281	0.933	0.934	0.771	0.633
	平均值	0.932	0.973	1.007	1.036	1.128	1.148	1.113	1.083
	P 平均值	0.962	0.759	0.445	0.237	0.014	0.016	0.067	0.093
100	能力	0.00	0.001	0.057	0.301	0.947	0.954	0.805	0.661
	平均值	0.93	0.972	1.006	1.036	1.127	1.148	1.113	1.083
	P 平均值	0.971	0.779	0.446	0.224	0.012	0.011	0.055	0.083
120	能力	0.00	0.00	0.052	0.334	0.974	0.974	0.852	0.732
	平均值	0.929	0.971	1.005	1.035	1.128	1.149	1.114	1.083
	P 平均值	0.982	0.809	0.452	0.206	0.006	0.007	0.041	0.064
140	能力	0.00	0.00	0.052	0.336	0.986	0.988	0.894	0.785
	平均值	0.928	0.971	1.004	1.034	1.127	1.15	1.116	1.084
	P 平均值	0.989	0.834	0.454	0.192	0.004	0.003	0.027	0.048
160	能力	0.00	0.00	0.054	0.402	0.993	0.992	0.916	0.819
	平均值	0.927	0.97	1.004	1.034	1.128	1.15	1.114	1.084
	P 平均值	0.993	0.858	0.457	0.177	0.002	0.002	0.021	0.040
180	能力	0.00	0.00	0.052	0.416	0.998	0.996	0.934	0.853
	平均值	0.926	0.969	1.003	1.034	1.128	1.149	1.115	1.084
	P 平均值	0.995	0.874	0.461	0.167	0.001	0.001	0.016	0.033

分布		轻尾、正态或接近正态的对称分布				重尾的对称分布			
		U(0, 1)	B(3, 3)	N(0, 1)	t(10)	Lp1	CN(.8, 3)	CN(.9, 3)	t(5)
n	TrueR	0.921	0.965	1.0	1.032	1.128	1.152	1.118	1.085
200	能力	0.00	0.00	0.053	0.448	0.998	0.998	0.954	0.884
	平均值	0.926	0.969	1.003	1.034	1.127	1.15	1.116	1.083
	P 平均值	0.997	0.89	0.461	0.153	0.001	0.001	0.011	0.025

表 6 SJ 检验的模拟能力水平。从接近正态、中度重尾或重尾的偏态分布生成样本。

分布		接近正态或中度重尾的偏态分布			重尾的偏态分布	
		Chi(10)	B(8, 1)	Chi(5)	Chi(3)	期望值
n	TrueR	1.028	1.075	1.059	1.098	1.151
10	能力	0.12	0.213	0.161	0.218	0.283
	平均值	1.072	1.105	1.088	1.108	1.136
	P 平均值	0.326	0.284	0.304	0.279	0.251
15	能力	0.139	0.27	0.205	0.292	0.377
	平均值	1.062	1.105	1.082	1.11	1.141
	P 平均值	0.32	0.261	0.286	0.245	0.209
20	能力	0.152	0.295	0.223	0.328	0.449
	平均值	1.051	1.089	1.070	1.101	1.142
	P 平均值	0.335	0.26	0.296	0.242	0.186
25	能力	0.16	0.336	0.255	0.374	0.515
	平均值	1.043	1.084	1.068	1.101	1.144
	P 平均值	0.337	0.236	0.281	0.219	0.156
30	能力	0.171	0.37	0.285	0.414	0.564
	平均值	1.043	1.084	1.065	1.097	1.142
	P 平均值	0.329	0.228	0.274	0.206	0.139
40	能力	0.193	0.44	0.331	0.49	0.651
	平均值	1.039	1.085	1.064	1.098	1.143
	P 平均值	0.321	0.188	0.246	0.171	0.106
50	能力	0.215	0.484	0.37	0.556	0.72
	平均值	1.037	1.081	1.064	1.10	1.143
	P 平均值	0.314	0.173	0.22	0.14	0.080

分布		接近正态或中度重尾的偏态分布			重尾的偏态分布	
		Chi (10)	B(8, 1)	Chi (5)	Chi (3)	期望值
<i>n</i>	TrueR	1.028	1.075	1.059	1.098	1.151
60	能力	0.224	0.527	0.395	0.607	0.778
	平均值	1.035	1.079	1.062	1.099	1.146
	P 平均值	0.303	0.152	0.208	0.119	0.062
70	能力	0.241	0.568	0.438	0.648	0.822
	平均值	1.034	1.079	1.061	1.098	1.146
	P 平均值	0.292	0.134	0.191	0.104	0.048
80	能力	0.259	0.612	0.474	0.689	0.855
	平均值	1.034	1.079	1.062	1.098	1.148
	P 平均值	0.28	0.115	0.17	0.089	0.036
90	能力	0.284	0.643	0.501	0.733	0.89
	平均值	1.034	1.079	1.060	1.099	1.148
	P 平均值	0.27	0.104	0.163	0.075	0.028
100	能力	0.285	0.675	0.527	0.757	0.912
	平均值	1.032	1.078	1.060	1.098	1.147
	P 平均值	0.267	0.094	0.151	0.067	0.022
120	能力	0.323	0.728	0.572	0.816	0.942
	平均值	1.032	1.077	1.060	1.098	1.149
	P 平均值	0.246	0.074	0.129	0.050	0.014
140	能力	0.344	0.769	0.621	0.852	0.963
	平均值	1.031	1.077	1.060	1.099	1.148
	P 平均值	0.232	0.060	0.112	0.036	0.009
160	能力	0.363	0.815	0.666	0.887	0.978
	平均值	1.031	1.077	1.060	1.098	1.15
	P 平均值	0.217	0.047	0.093	0.027	0.005
180	能力	0.385	0.843	0.692	0.91	0.986
	平均值	1.031	1.077	1.059	1.099	1.148
	P 平均值	0.209	0.039	0.083	0.021	0.004
200	能力	0.41	0.877	0.727	0.931	0.989
	平均值	1.030	1.077	1.059	1.098	1.149
	P 平均值	0.196	0.030	0.071	0.016	0.003

我们表 5 和 6 中的模拟结果与 Gel et al. (2007) 中发布的结果一致。当样本来自正态总体时，模拟能力水平（在这种情况下代表检验的实际重要性水平）也离目标水平不远了，即使样本量低至 25 也无妨。当样本呈重尾分布时，小样本量的检验能力较低，但当样本量达到 40 时，检验能力至少会增加到 40%。具体来说，样本量为 40 的检验能力约为自由度为 5 的 t 分布的 40.1%、Laplace 分布的 66.5%，以及指数分布的 65.1%。

对于轻尾分布（Beta(3, 3) 和均匀分布），小样本的检验能力接近 0，随着样本量的增加会进一步减小。这也不足为奇，因为这些分布的证据实际支持轻度轻尾分布的另类假设，而不是重度重尾分布的另类假设。

当样本来自轻度重尾分布时，如自由度为 10 的 t 分布或自由度为 10 的卡方分布，中等到大样本量的能力水平较低。对我们而言，这实际上已是不错的结果了，因为对这些分布执行的某个方差（标准差）检验开展顺利，我们不希望将这些分布标记为重尾。不过，当样本量增加时，检验能力会提高，因此这些轻度重尾分布将被检测为重尾分布。

因此，用于评估该检验的分布的尾部重量的规则还必须考虑样本量。这样做的一种方法是计算尾部重量测量值的置信区间；不过，SJ 统计数据的分布对样本的母体分布极为敏感。另一种方法是根据 SJ 检验和样本量的零假设的拒绝力度评估分布的尾部重量。更准确地说，较小的 p 值表示重度重尾，较大的 p 值表示轻度轻尾。然而，大样本的 p 值往往比小样本的更小。因此，根据模拟的能力水平、样本量和表 3 中的 P 平均值，我们设计了一套通用规则，用于使用 SJ 检验评估每个样本的分布尾部。

对于中等到大样本量 ($40 \leq n \leq 100$)，如果 p 值介于 0.01 和 0.05 之间，我们认为零假设的证据不够充分。即样本的分布被划分为中度重尾 (M 型) 分布。另一方面，如果 p 值小于 0.01，则零假设的证据充分，并且样本的母体分布被划分为重尾 (H 型) 分布。

对于大样本 ($n > 100$)，如果 p 值在 0.005 和 0.01 之间，我们将母体分布归类为 M 型分布，如果 p 值非常小（低于 0.005），则归类为 H 型分布。请注意，当样本量小于 40 时，SJ 的检验能力通常太低，从而无法有效确定样本分布。

使用 Bonett 方法的单样本方差检验的有效性的一般分类规则在下面的表 7 中总结。

表 7 用于识别每一个样本的母体分布的分类规则 (p 是 SJ 检验的 p 值)

条件	分布类型
$n < 40$	不确定
$100 \geq n \geq 40$ 和 $p > 0.05$ 。	L 型分布
$n > 100$ 和 $p > 0.01$	L 型分布
$40 \leq n \leq 100$ 和 $0.01 < p \leq 0.05$	M 型分布
$n > 100$ 和 $0.005 < p \leq 0.01$	M 型分布
$40 \leq n \leq 100$ 和 $p \leq 0.01$	H 型分布
$n > 100$ 和 $p \leq 0.005$	H 型分布

如前所述，根据模拟 C1 中表 1 和表 2 的结果，从 L 型、M 型和 H 型分布生成样本时，达到 0.93 的最小覆盖概率所需的近似最小样本量分别为 20、80 和 200。不过，因为小样本的 SJ 检验的能力较低，L 型分布的最小样本量要求被设置为 40。

模拟 F2：验证分布分类的规则

我们从模拟 C1 中所述的一些分布中生成了样本并使用 SJ 检验确定归类到 3 个分布组（L 型、M 型和 H 型）中的一个分布的样本比例。模拟结果在表 8 中显示。

表 8 确定为 L 型、M 型和 H 型的各个分布中的 10,000 个不同大小的样本的一部分

n	分布	L 型				M 型	H 型	
		B(3, 3)	N(0, 1)	t(10)	Chi(10)	Chi(5)	Lp1	期望值
40	%L 型	99.6	94.0	81.5	80.3	66.6	33.0	34.4
	%M 型	0.4	5.5	14.0	14.0	20.0	31.9	22.9
	%H 型	0.0	0.5	4.5	5.7	13.4	35.1	42.8
50	%L 型	99.7	94.4	78.7	79.1	64.0	25.1	28.0
	%M 型	0.3	5.1	15.6	14.2	20.0	29.9	20.7
	%H 型	0.0	0.5	5.7	6.7	16.0	45.0	51.3
60	%L 型	99.7	94.5	77.3	77.3	59.1	18.5	22.6
	%M 型	0.3	5.1	16.4	15.0	22.0	27.4	19.2
	%H 型	0.0	0.5	6.3	7.7	18.9	54.1	58.2
70	%L 型	99.8	94.4	74.5	75.2	55.9	14.0	18.1
	%M 型	0.2	5.0	18.1	16.0	22.2	24.0	17.5
	%H 型	0.0	0.6	7.4	8.8	21.9	62.0	64.4
80	%L 型	99.9	94.3	74.1	74.4	53.0	10.0	13.9
	%M 型	0.1	5.1	17.8	16.7	22.8	21.0	15.5
	%H 型	0.0	0.6	8.2	8.9	24.2	69.0	70.6
90	%L 型	99.9	94.4	71.2	72.1	49.5	7.5	11.1
	%M 型	0.1	5.0	19.1	17.2	22.6	16.5	13.7
	%H 型	0.0	0.6	9.7	10.7	27.9	76.0	75.3
100	%L 型	99.9	94.5	70.8	70.3	47.3	4.8	8.9
	%M 型	0.1	4.9	19.5	17.9	22.7	14.3	11.8
	%H 型	0.0	0.6	9.7	11.8	30.0	80.9	79.4
120	%L 型	100.0	99.4	87.4	87.2	64.8	12.0	14.4
	%M 型	0.0	0.4	5.0	4.5	7.9	7.8	5.6
	%H 型	0.0	0.2	7.6	8.4	27.4	80.4	80.0
140	%L 型	100.0	99.3	86.0	85.1	60.5	7.0	9.9
	%M 型	0.0	0.5	5.2	5.0	8.6	5.6	4.1
	%H 型	0.0	0.2	8.8	9.9	30.9	87.4	86.0

n	分布	L 型				M 型	H 型	
		B(3, 3)	N(0, 1)	t(10)	Chi(10)	Chi(5)	Lp1	期望值
160	%L 型	100.0	99.4	83.4	83.0	55.6	4.0	6.9
	%M 型	0.0	0.5	6.3	5.8	9.5	3.5	3.0
	%H 型	0.0	0.1	10.4	11.2	34.9	92.5	90.1
180	%L 型	100.0	99.3	81.1	81.7	51.0	2.5	4.6
	%M 型	0.0	0.5	6.8	5.9	9.4	1.9	2.2
	%H 型	0.0	0.2	12.1	12.4	39.6	95.6	93.2
200	%L 型	100.0	99.5	79.0	80.5	47.2	1.3	3.0
	%M 型	0.0	0.4	7.6	6.1	9.4	1.6	1.7
	%H 型	0.0	0.1	13.4	13.4	43.4	97.1	95.3

表 8 中的结果表明，当样本呈轻尾（L 型）和重尾（H 型）分布时，将对较高比例的样本进行正确的分类。例如，当从 Beta(3, 3) 分布生成 40 个样本量时，99.6% 的样本被正确地归类为轻度轻尾分布；当从 Laplace 分布生成 90 个样本量时，76.0% 的样本被正确地归类为重尾分布。因此，当样本真正来自轻度轻尾分布时，Report Card 中不会正确地发出有关检验的有效性的警告消息，而当样本来自重尾分布且不满足最小样本量要求时，会正确地发出警告消息。此外，对于来自诸如卡方（5）分布等中度重尾（M 型）分布的样本，当样本量较小时，较高的样本比例会被错误地归类为轻尾（L 型）（对于 40 个样本量，66% 的样本会被错误地归类为轻尾分布）。因此，对于这些情况，Report Card 中的警告消息可能无法发出，即使母态分布呈中度重尾也如此。不过，当样本量大于 80 时，即使错误地归类为 L 型分布也无妨，因为最小样本量要求已满足。

© 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.