



4 改善您的新方法 单个种群的估计 Minitab 中的比例



基本统计学中的一个常见问题是估计人口中具有某种感兴趣特征的个体的比例。例如，质量工程师可能想要估计给定日期大批量生产单元中的缺陷比例；医学科学家可能想调查某些社区中接种了特定病原体疫苗但仍然患有相关疾病的个体的比例；竞选经理可能对打算投票给其候选人的登记选民的比例感兴趣。

对于这个问题，最著名的区间估计方法是教科书式的正态近似方法，称为 **Wald 置信区间 (CI)** 和 **Clopper-Pearson 精确 (1934) CI**。一方面，**Wald CI** 非常宽松，因为 **CI** 的实际置信水平（或覆盖概率）远低于目标名义水平，特别是当真实比例接近 0 或 1 时（见图 1）。另一方面，确切的 **Clopper-Pearson** 置信区间过于保守，因为置信区间的实际置信水平（或覆盖概率）远高于目标名义水平。这两种方法都不应再用于任何实际应用（见 **Agresti-Coull, 1998; Brown 等人, 2001 年**）。

然而，近年来，它们在开发具有更好中间覆盖概率的更好的 **CI** 方法方面发挥了重要作用。

例如，**Agresti-Coull** 近似置信区间是对 **Wald** 置信区间的调整；**Blaker (2000, 2001)** 精确置信区间使用 **Clopper-Pearson** 置信边界作为迭代数值算法中的起始估计值。考虑到这些新改进的方法，**Minitab** 更新了用于估计单个总体比例的统计工具，以包括以下 4 种方法：校正后的 **Blaker** 置信区间和检验方法、**Wilson/score** 置信区间和检验方法（带和不带连续性校正）以及 **Agresti-Coull** 置信区间和检验方法。此外，对于这些方法中的每一种，**Minitab** 都会确保置信区间和检验产生一致的结果。

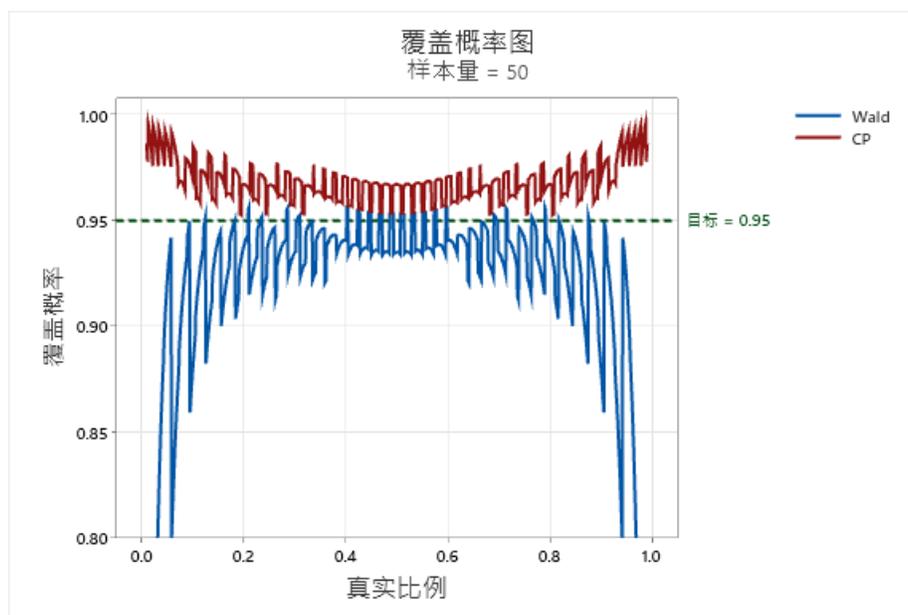


图 1: 当样本量为 50 时，比较 **Wald** 置信区间和 **Clopper-Pearson (CP)** 置信区间作为真实比例函数的覆盖概率。该图表明，**Wald CI** 和 **Clopper Pearson CI** 分别过于自由和保守，尤其是当真实比例接近 0 或 1 时。假设真实比例均匀分布在区间 (0, 1) 中，则基于大小为 50 的样本的 **Wald CI** 和 **Clopper-Pearson CI** 的平均覆盖概率分别为 0.901 和 0.969。

了解 4 种新的估算方法

这 4 种新方法包括 1 个精确置信区间和称为调整后的 **Blaker** 方法的测试方法，以及 3 个近似置信区间和测试方法，包括 **Wilson/score (Wilson)** 方法、**Wilson/score with Yates' continuity correction (Wilson CC)** 方法和 **Agresti-Coull (AC)** 方法。在这种情况下，精确方法意味着在方法的推导中没有使用近似值，而不是使用某些形式的正态近似过程获得的近似方法。

1. 调整后的 Blaker 方法

由于 Klaschka 和 Reiczigel (2021) 而调整后的 Blaker 方法是对 Blaker (2000, 2001) 精确置信区间和测试方法的修改。该修改解决了原始 Blaker 算法的计算密集型性质以及其 CI 和测试结果之间偶尔的不一致。与原始 Blaker CI 一样，生成的调整后的 CI 是精确的、嵌套的，并且包含在 Clopper-Pearson CI 中。因此，调整后的 Blaker CI 不如 Clopper-Pearson CI 保守。置信区间嵌套的意义在于，置信度较高的置信区间始终包含置信水平较低的置信区间。例如，双侧 95% (校正) Blaker 置信区间始终包含相应的双侧 90% 置信区间。嵌套性是从离散分布 (如二项式) 派生的精确置信区间方法的一个吸引人的属性。例如，嵌套了 Clopper-Pearson CI。但是，有一些可用的精确 CI 方法不一定是嵌套的。例如，所谓的 Blyth-Still-Casella CI (Blyth 和 Still, 1983 年; Casella, 1986) 保证是最短的精确置信区间，但不是嵌套的。Crow (1956) CI 也没有嵌套。基于 Blaker 或调整后的 Blaker 方法的置信区间计算比上述经典置信区间方法更复杂，因为它们需要数值算法。然而，随着当前计算技术的创新，我们不应再回避实施产生更好结果的复杂算法。图 2 显示了调整后的 Blaker CI 相对于 Clopper-Pearson CI 的改进。

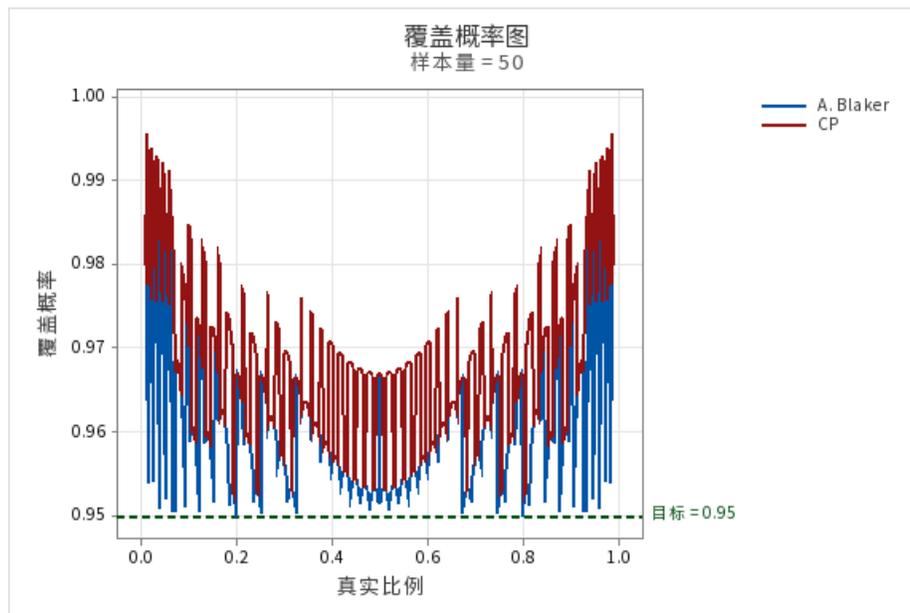


图 2: 当样本量为 50 时，校正后的 Blaker (A. Blaker) 置信区间和 Clopper-Pearson (CP) 置信区间作为真实比例的函数的覆盖概率比较。该图表明，Clopper-Pearson CI 的覆盖概率至少是调整后的 Blaker CI 的覆盖概率。这与调整后的 Blaker CI 包含在 Clopper-Pearson CI 中的事实一致。对于任何大小为 50 的给定样本，调整后的 Blaker CI 和 Clopper-Pearson CI 的平均覆盖概率 (假设真实比例均匀分布在单位区间上) 分别为 0.960 和 0.969。

2. Wilson 和 Wilson CC 方法

Wilson (1927) CI 方法派生为与分数检验相对应的置信区间，该检验使用零标准误差 $\sqrt{p_0(1-p_0)/n}$ ，而不是经典标准误差 $\sqrt{\hat{p}(1-\hat{p})/n}$ ，在检验统计量的分母上。因此，它也被称为威尔逊/分数置信区间。其实际覆盖概率可以小于或大于目标标称水平，但保持接近目标水平，除非真实比例接近 0 或 1 (见图 3)。可以通过使用 Yates 的连续性校正进行调整以使 Wilson/score CI 保守。Minitab 提供 CI 版本 (带或不带 Yates 连续性校正) 及其匹配假设检验。

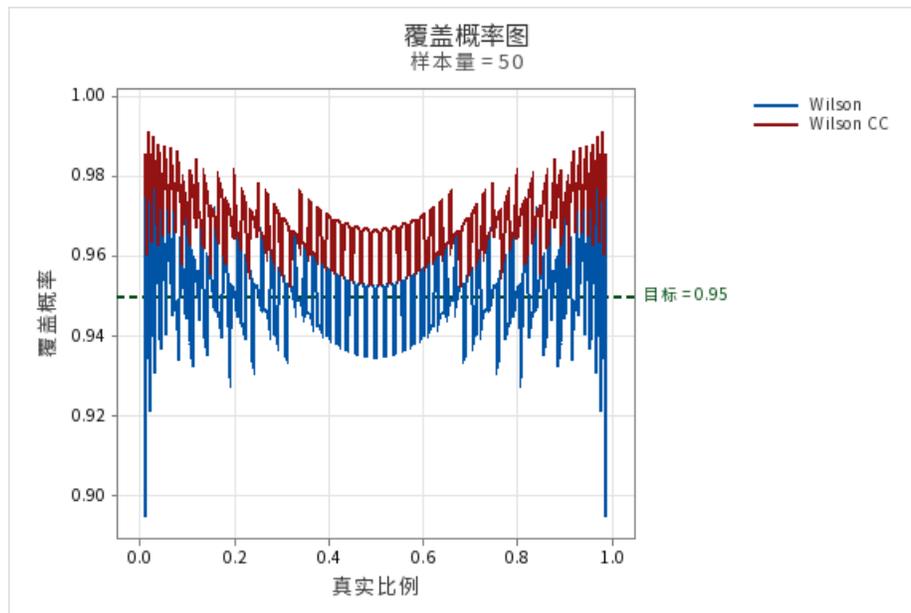


图 3: 当样本数量为 50 时，威尔逊/分数（Wilson）置信区间和 Wilson/score 与 Yates 连续性校正（Wilson CC）的覆盖概率作为真实比例的函数的比较。该图显示，Wilson CC CI 始终是保守的，而 Wilson CI 是保守的和自由的，具体取决于真实比例的大小。特别是，当真实比例非常接近 0 或 1 时，Wilson CI 往往过于宽松。对于任何大小为 50 的给定样本，Wilson CI 和 Wilson CC CI 的平均覆盖概率分别为 0.952 和 0.969。

3. Agresti-Coull 方法

Agresti-Coull CI 是从对过度自由的经典 Wald CI 的调整中获得的。生成的置信区间具有与威尔逊置信区间相似的覆盖率属性，但总体上更保守一些。此外，两种类型的置信区间具有相同的中心点，但威尔逊置信区间始终包含在 Agresti-Coull 置信区间中。如图 4 所示，当真实比例适中时，它们的覆盖概率基本相同。然而，当真实比例接近 0 或 1 时，Agresti Coull CI 通常不那么自由。如图 4 所示，对于大小为 50 的样本，当真实比例接近 0 或 1 时，Agresti-Coull 置信区间变得保守。Agresti-Coull CI 的另一个吸引力是其从 Wald CI 继承而来的实现简单性。此外，它很容易教学和记忆，尤其是当置信度为 95% 时。对于这个置信水平，它通常被称为“添加 2 个成功和 2 个失败”CI 方法，作为对 Wald CI 进行调整以得出它的备忘录。

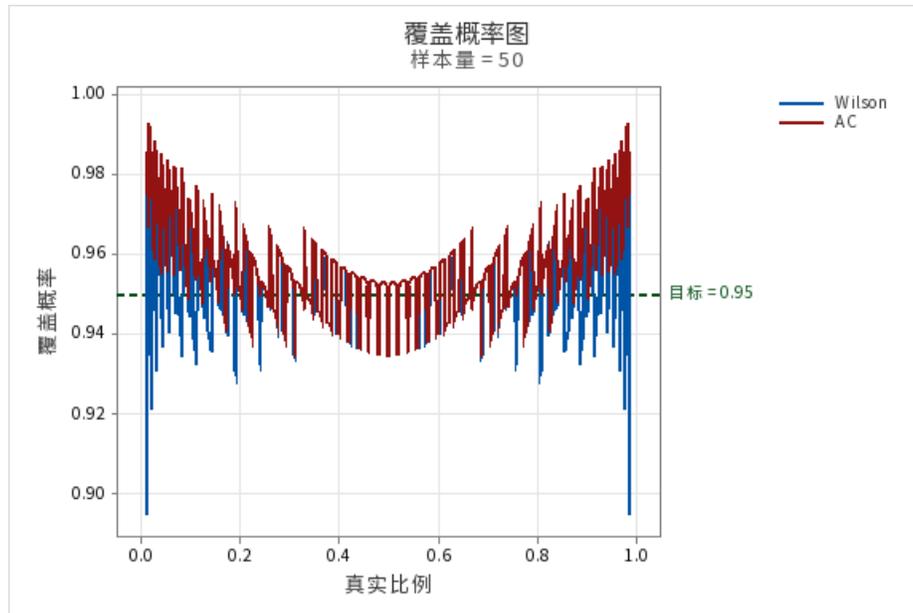


图 4: 当样本量为 50 时，威尔逊/分数 (Wilson) 置信区间和 Agresti-Coull (AC) 置信区间作为真实比例函数的覆盖概率比较。该图表明，对于区间内真实比例的中等值 (0.3, 0.75)，这两种方法产生的覆盖概率基本相同。然而，对于接近 0 或 1 的真实比例值，Wilson CI 是自由的，而 Agresti-Coull CI 是保守的。这与威尔逊置信区间包含在 Agresti-Coull 置信区间中的事实一致。对于任何大小为 50 的给定样本，Wilson CI 和 Agresti-Coull CI 的平均覆盖概率分别为 0.952 和 0.958。

一些简单的说明性例子

一家批量生产工厂的质量工程师在给定的一天随机抽取了 1465 种批量生产的商品。在对 1465 件物品进行独立测试后，发现 53 件有缺陷。工程师想知道在给定日期生产的缺陷品的比例是否与 2.5% 有显著差异。

尽管这个问题被定义为统计假设检验问题，但越来越鼓励应用统计学家在分析结果中包括点估计值和置信区间以及检验的 p 值。Minitab 尽可能遵循此约定，尤其是在基本统计模块中。例如，使用 Minitab，基于调整后的 Blaker 方法，上述问题的分析结果如下。

单比率检验和置信区间

方法

p 事件比率

方法：调整了 Blaker 的确切方法

描述性统计量

N 事件样本 p 的 95% 置信区间

| | | | |
|------|----|----------|----------------------|
| 1465 | 53 | 0.036177 | (0.027353, 0.046822) |
|------|----|----------|----------------------|

检验

原假设 : $p=0.025$

备择假设 H_1 : $p \neq 0.025$

P 值

0.009

对于这 4 种方法中的每一种，都可以生成类似的输出。下表汇总了每种方法的特征和优点：

| 方法 | 95% | 匹配检验的 P 值 |
|-----------|---------------|-----------|
| 调整后的布莱克 | (2.74%,4.68%) | 0.009 |
| 威尔逊与耶茨的更正 | (2.75%,4.74%) | 0.008 |
| 威尔逊/得分 | (2.78%,4.70%) | 0.006 |
| 阿格雷斯蒂-库尔 | (2.77%,4.74%) | 0.007 |

在此示例中，所有方法都得出相同的结论，即缺陷百分比在 0.05 显著性水平上与 2.5% 不同，因为所有 p 值都小于 0.05。所有方法的置信区间和相应的 p 值相似，部分原因是样本的大小非常大。此外，每种方法的置信区间未涵盖假设比例值（2.5%），这与每个匹配假设检验的 p 值一致。

在上面的例子中，假设现在质量工程师只测试了 50 个项目，发现 2 个有缺陷。此外，假设工程师想知道缺陷品的比例是否与 1.0% 有显著差异。下表汇总了每种方法的特征和优点：

| 方法 | 95% | 相应检验的 P 值 |
|-----------|----------------|-----------|
| 调整后的布莱克 | (0.72%,13.35%) | 0.089 |
| 威尔逊与耶茨的更正 | (0.70%,14.86%) | 0.155 |
| 威尔逊/得分 | (1.10%,13.46%) | 0.033 |
| 阿格雷斯蒂-库尔 | (0.34%,14.22%) | 0.124 |

在这种情况下，只有 Wilson/score 方法得出一个显著性结论，即缺陷百分比在 0.05 显著性水平上与 1.0% 不同。在相同的显著性水平上，所有其他方法都得出相反的结论，即没有足够的证据来确定存在差异。不同方法的结果不一致很大程度上是由于样本大小适中。平均而言，随着样本量的增加，这些方法的覆盖概率接近标称水平（见下图 5）。然而，对于小到中等样本设计，与每种方法相关的平均覆盖概率的差异更为明显，导致相应的置信区间具有明显不同的宽度。然而，一个重要的问题是，向老板报告哪个结果？这个问题没有直接的答案，因为更好的答案可能取决于诸如对真实比例大小甚至应用领域的先验知识等因素。我们将在下一节中提供一些粗略的一般准则。

结论

图 2、3 和 4 显示 4 种置信区间方法，即 Adjusted Blaker、Wilson、Wilson CC 和 Agresti-Coull 具有不同的覆盖概率属性。Wilson CC 是最保守的，其次是调整后的 Blaker。Agresti Coull 和 Wilson 通常是自由派和保守派，具体取决于真实比例的大小。总体而言，Wilson CI 方法是所有 4 种方法中最自由的。此外，图 5 表明，平均而言，所有 4 种方法都是保守的，其中 Wilson CC 是最保守的，其次是调整后的 Blaker、Agresti-Coull 和 Wilson 方法。对于给定的问题，选择适当的方法可能取决于特定的应用、样本量以及是否对真实比例的大小有一些先验知识。

例如，监管机构通常对保守的方法感兴趣，以保护消费者。然而，过于保守的方法可能会产生严格的法规，而过于宽松的方法可能会产生宽松的法规。一般来说，对于中等到大的样品设计，调整后的 Blaker 或 Willson CC 方法可能是不错的选择。对于小样本设计，如果事先了解真实比例值，Wilson 或 Agresti-Coull 可能是合适的。这种先验知识通常基于先前的类似经验或专门设计的试点小型研究，以获得对比例的粗略估计。例如，在缺陷比例通常接近 0 的应用质量控制领域中，选择的方法取决于研究者是否想要保守的结果。保守的结果可以基于 Agresti-Coull 方法，而自由的结果可以基于 Wilson/score 方法。最后但更重要的是，提前计划样本量始终是很好的做法，以防止 II 类误差或控制置信区的宽度。Minitab 也有“一个应用程序”。Minitab 中提供了一些工具，用于确定足够的样本数量，以控制与假设检验相关的 II 类误差或控制置信区的宽度。

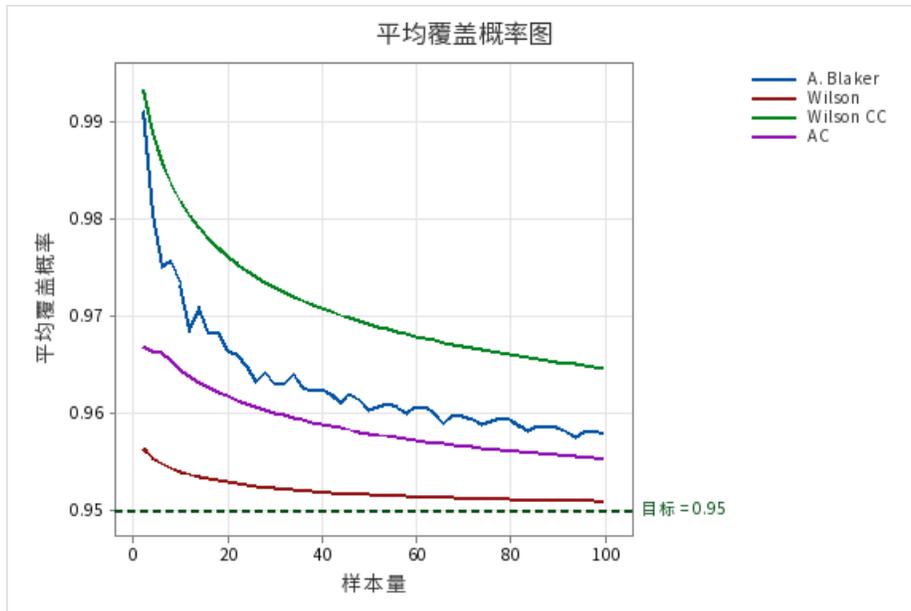


图 5: 所有 4 种 CI 方法的平均覆盖概率作为样本量的函数。平均覆盖概率的计算假设真实比例均匀分布在单位区间上。平均覆盖率曲线表明，平均而言，使用 Yates 连续性校正 (Wilson CC) CI 方法的 Wilson/score 是最保守的，其次是调整后的 Blaker (A. Blaker)、Agresti-Coull (AC) 和 Wilson/score (Wilson) CI 方法。随着样本量的增加，平均覆盖概率曲线接近目标名义覆盖水平。此外，虽然近似方法 (Wilson CC、Wilson、Agresti-Coull) 的平均覆盖曲线是平滑的，但精确调整的 Blaker 的覆盖曲线在接近标称覆盖率时有一些摆动。这也许表明，调整后的 Blaker CI 方法可以进一步改进。

参考值

Agresti, A. and Coull, B. A. (1998). Approximate is better than “Exact” for interval Estimation of Binomial Proportion. *The American Statistician* 52, 119—125

Blaker, H. (2000). Confidence Curves and Improved Exact Confidence Intervals for Discrete Distributions. *The Canadian Journal of Statistics*, 28, 783—798

Blaker, H. (2001). Corrigenda: Confidence curves and improves exact confidence intervals for discrete distributions. The Canadian Journal of Statistics, 29, 681.

Blyth, C. R. and Still, H. A. (1983). Binomial Confidence Intervals. Journal of the American Statistical Association 78, 108—116.

Brown, L. D., Cai, T. and Das Gupta, A. (2001). Interval Estimation for a Binomial Proportion. Statistical Science 16, 101—133.

Casella, G., 1986. Refining binomial confidence intervals. Canad. J. Statist. 14, 113 – 129.

Clopper, C. J. and Pearson, E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of Binomial. Biometrika 26, 404—413

Crow, E.L., 1956. Confidence intervals for a proportion. Biometrika 43, 423 – 435.

Klaschka, J. and Reiczigel, J. (2021). On matching confidence intervals and tests for some discrete distributions: methodological and computational aspects, Computational Statistics, Springer, vol. 36(3), 1775-1790.

Wilson E. B. (1927) Probable Inference, the Law of Successions and Statistical Inference. J. Amer. Statist. Assoc. 22, 209—21

© 2024 Minitab, LLC. All rights reserved. Minitab®, Minitab Connect®, Minitab Model Ops®, Minitab Engage®, Minitab Workspace®, 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.



You Have Data. We Have Solutions Analytics.

Download a Free Trial
minitab.com