

# 单样本不良品率检验

## 概述

单比率检验用于确定某个比率是否不同于目标值。在产品或服务属于不良品还是良品的质量分析中，检验通常用于确定不良品率是否显著不同于目标不良品率。

Minitab 的“协助”中提供单样本不良品率检验。收集的检验数据是样本中不良品的数量，被认定为是二项式随机变量的观测值。“协助”使用准确的方法来计算假设检验结果和置信区间；因此，实际的 I 类错误概率应接近为该检验指定的显著性水平（ $\alpha$  值），因而无需展开进一步的研究。然而，单样本不良品率检验的功效和样本量分析都基于一个近似值，我们需要评估其准确性。

在本文中，我们探讨了用于评估单样本不良品率检验的功效和样本量的方法，对近似方法的理论功效与精确检验的实际功效进行了比较。

我们还介绍了如何确立一个指导方针，帮助您评估样本量是否足够大，以确定不良品率是否不同于目标不良品率。“协助”会自动检查样本量并在“报告卡”中报告研究结果。

单样本不良品率检验还取决于其他假设。有关详细信息，请参见附录 A。

# 单样本不良品率检验方法

## 理论功效水平

“协助”将使用准确的（概率比）方法，对单个 Bernoulli 总体率（不良品率）进行假设检验。然而，由于这种精确检验的功效不容易获得，我们将使用相应正态近似检验的理论功效来获得此功效的近似值。

### 目标

我们想要确定是否可以使用基于正态近似检验的理论功效，来评估“协助”中单样本不良品率检验的功效和样本量要求。为此，我们需要评估这一理论功效能否准确地反映精确（概率比）检验的实际功效。

### 方法

检验统计值、p 值和精确（概率比）检验的置信区间都在附录 B 中定义。基于正态近似检验的理论功效在附录 C 中定义。根据这些定义，我们进行了使用精确检验估计实际功效水平的模拟（我们称之为“模拟功效水平”）。

为进行模拟，我们从若干 Bernoulli 总体中生成了各种大小的随机样本。对于每一个 Bernoulli 总体，我们对 10,000 个重复样本中的每一个样本都进行了精确检验。对于每个样本量，我们分别计算了检验的模拟功效，以检测 10,000 个对检验意义重大的样本中的给定差值。为了便于比较，我们还基于正态近似检验计算了相应的理论功效。如果近似检验很有效，则理论功效水平应接近模拟功效水平。有关详细信息，请参见附录 D。

### 结果

我们的模拟结果表明，在一般情况下，正态近似检验的理论功效与精确（概率比）检验的模拟功效几乎相同。因此，“协助”使用正态近似检验的理论功效来估计所需的样本量，以确保精确检验具有足够的功效，可检测到不良品率中实际存在的重要差异。

# 数据检查

## 样本量

通常情况下，执行假设检验可收集证据，以否定“无差异”的原假设。如果样本太小，检验的功效可能不足以检测出确实存在的差异，这将导致 II 类错误。因此，一定要保证样本量足够大，以高概率地检测到实际存在的重要差异。

### 目标

如果数据未提供足够的证据来否定原假设，我们希望确定样本量是否足够大，从而检验能够高概率地检测到所需的实际差异。尽管样本量规划的目标是确保样本量足够大，能够以较高的高概率检测到重要差异，但它们也不应太大，使得无意义的差异成为具有重要统计意义的高概率差异。






### 方法

单样本不良品率检验的功效和样本量分析基于使用正态近似值的理论功效，它将有效地估计精确检验的实际功效（参见上面的“单样本不良品率检验方法”部分）。给定目标不良品率时，理论功效取决于样本量和要检测的差异。

### 结果

当数据未提供足够的原假设证据时，“协助”将计算用给定样本量 80% 和 90% 的概率检测到的实际差异。另外，如果用户提供所需的特定实际差异，“协助”将计算差异检测概率为 80% 和 90% 的样本量。

为了帮助解释结果，检查功效和样本量时，单样本不良品率检验将在“帮助报告卡”中显示以下状态指标：

状态	条件
	检验发现不良品率和目标值之间存在差异，所以功效不是问题。 或 功效是足够的。检验未发现目标值之间存在差异，但样本量足够大，至少有 90% 的机会检测到给定差异（功效 $\geq .90$ ）。
	功效可能足够。检验未发现与目标值之间存在差异，但样本量足够大，有 80%~90% 的机会检测到给定差异（ $.80 \leq \text{功效} < .90$ ）。报告实现 90% 的功效所需的样本量。
	功效可能不够。检验未发现与目标值之间存在差异，但样本量足够大，有 60%~80% 的机会检测到给定差异（ $.60 \leq \text{功效} < .80$ ）。报告实现 80% 和 90% 的功效所需的样本量。
	功效不够。检验未发现与目标值之间存在差异，样本量不够大，不足以提供至少 60% 的机会检测到给定差异（功效 $< .60$ ）。报告实现 80% 和 90% 的功效所需的样本量。
	检验未发现与目标值之间存在差异。您没有指定要检测的实际差异；因此，该报告将根据样本量和 alpha 值指出检测概率为 80% 和 90% 的差异。

# 参考书

Arnold, S.F. (1990). *Mathematical statistics*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Casella, G., & Berger, R.L. (1990). *Statistical inference*. Pacific Grove, CA: Wadsworth, Inc.

# 附录 A：单样本不良品率的其他假设

单样本不良品率检验基于以下假设：

- 数据包括  $n$  个不同的项，每个项都被归类为不良品或良品。
- 样本中每个项的不良品率相同。
- 另一个项是否不良品不会影响不良品率。

这些假设不能在报告卡的数据检查中验证，因为此检验会输入汇总数据，而不是原始数据。

# 附录 B：精确（概率比）检验

假设我们从 Bernoulli 分布中观测到一个随机样本  $X_1, \dots, X_n$ ，成功概率为  $p = \Pr(X_i = 1) = 1 - \Pr(X_i = 0)$ 。

绘制有关  $p$  的推测结果的确切方法如下所述。

## 公式 B1：精确检验和 $p$ 值

考虑针对下述任何替代假设的原假设  $H_0: p = p_0$  检验： $H_A: p > p_0$ 、 $H_A: p < p_0$  或  $H_A: p \neq p_0$ 。

如果  $X = \sum_{i=1}^n X_i$

那么， $X$  是试验次数为  $n$  且成功概率为  $p$  的二项式随机变量。

基于  $X$  的单侧检验是 UMP（一致最大功效）和概率比检验。对于双侧检验，概率比检验也基于  $X$ ，检验统计值是

$$\Lambda(X) = \left(\frac{\hat{p}}{p_0}\right)^X \left(\frac{1-\hat{p}}{1-p_0}\right)^{n-X}$$

（参见 Arnold, 1990）。

单侧检验的  $P$  值可以直接根据确切的  $X$  分布获得。对于双侧检验， $p$  值以基于原假设的概率方式计算得到，此概率则通过观测至少与实际观测到的概率比一样大的概率比（或对数概率比）得到。一般使用数值求根算法来计算此概率。

## 公式 B2：准确的置信区间

$p$  准确的  $100(1-\alpha)\%$  双侧置信区间为

$$\frac{1}{1 + \frac{n-x+1}{x} F_{2(n-x+1), 2x, \alpha/2}} \leq p \leq \frac{\frac{x+1}{n-x} F_{2(x+1), 2(n-x), \alpha/2}}{1 + \frac{x+1}{n-x} F_{2(x+1), 2(n-x), \alpha/2}}$$

其中  $x$  是观测到的成功数目， $F_{\nu_1, \nu_2, \alpha}$  是自由度为  $\nu_1$  和  $\nu_2$  的  $F$  分布的上  $\alpha$  百分位点（参见 Casella and Berger, 1990）。我们采用的约定是，如果  $x = 0$ ，则下限为 0，如果  $x = n$ ，则上限为 1。

# 附录 C：理论功效

精确检验的理论功效太复杂，因而无法获得。因此，我们通过使用基于正态近似检验的理论功效来估计精确检验的功效。此近似检验基于随机变量

$$Z = \frac{n^{1/2}(\hat{p} - p)}{(p(1-p))^{1/2}}$$

渐近分布为标准正态分布的事实进行。此检验的理论功效广为流传并已著书立传。对于双侧替代假设，功效表示为：

$$\pi(n, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) + \Phi\left(\frac{-\delta - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)$$

其中  $p = \delta + p_0$ ， $\Phi(\cdot)$  是标准正态分布的累积分布函数， $z_\alpha$  是标准正态分布的上百分位数。

对于单侧替代  $H_A: p > p_0$ ，功效可以表示为

$$\pi(n, \delta) = 1 - \Phi\left(\frac{-\delta + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)$$

当检验单侧替代  $H_A: p < p_0$  时，功效也可表示为

$$\pi(n, \delta) = \Phi\left(\frac{-\delta - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)$$

# 附录 D：实际功效与理论功效比较

## 模拟 D1：使用精确检验估计实际功效

我们设计了一种模拟方法，以将估计的实际功效水平（称为模拟功效水平）与基于正态近似检验功效的理论功效水平（称为近似功效水平）进行比较。在每个实验中，我们会从给定成功概率为  $p$  的 Bernoulli 总体中生成 10,000 个样本，每个样本的大小为  $n$ 。我们考虑了两种成功概率：（1）中等成功概率， $p$  的值接近 0.5（具体为  $p = 0.45$ ）以及（2）较小或较大的成功概率， $p$  值接近 0 或 1（具体为  $p = 0.85$ ）。我们考虑这两种情况是因为我们知道，当 Bernoulli 样本量大于 10 时，得出正态近似检验的 DeMoivre - Laplace 二项分布正态近似值是准确的，并且成功概率接近 0.5。但是，对于更小或更大的成功概率，只有提供更大的 Bernoulli 样本才能得到准确的近似值。

在每个实验中，我们将样本量固定为单个值  $n$ ，其中  $n = 10, 15, 20, 30, \dots, 100$ 。在所有的实验中，我们将要检测的差值  $\delta = p - p_0$  固定为 0.2，以确保所得到的功效值在样本量增至 100 时既不会太小，也不会太大。要根据每个模拟的结果估计检验的实际功效，我们同时使用单侧和双侧精确检验，计算了 10,000 个重复样本中的一部分样本，针对其开展的精确检验的目标显著性水平  $\alpha = 0.05$  具有重要意义。最后，我们计算了基于正态近似检验的相应理论功效水平以进行比较。结果在下面的表 1 中显示。

表 1 双侧和单侧精确检验的模拟与近似功效水平。目标显著性水平为  $\alpha = 0.05$ 。

$n$	双侧检验				单侧检验			
	$p = 0.45$		$p = 0.85$		$p = 0.45$		$p = 0.85$	
	模拟功效	近似功效	模拟功效	近似功效	模拟功效	近似功效	模拟功效	近似功效
10	0.101	0.333	0.20	0.199	0.257	0.436	0.20	0.335
15	0.339	0.441	0.322	0.327	0.339	0.55	0.322	0.489
20	0.406	0.537	0.409	0.455	0.59	0.643	0.648	0.621
30	0.632	0.69	0.708	0.674	0.632	0.779	0.708	0.808
40	0.781	0.799	0.863	0.822	0.781	0.867	0.863	0.911
50	0.877	0.872	0.874	0.91	0.877	0.921	0.933	0.961
60	0.878	0.92	0.942	0.957	0.922	0.954	0.969	0.984
70	0.925	0.951	0.972	0.981	0.953	0.973	0.987	0.994
80	0.954	0.971	0.986	0.992	0.986	0.985	0.993	0.998
90	0.971	0.982	0.993	0.996	0.991	0.991	0.996	0.999
100	0.989	0.99	0.998	0.999	0.994	0.995	0.999	1.00



结果表明，模拟功效水平和近似功效水平通常非常一致。当结果以图形方式显示为功效曲线时，您可以更清楚地看到这种一致性，如下面的图 1 和图 2 中所示。

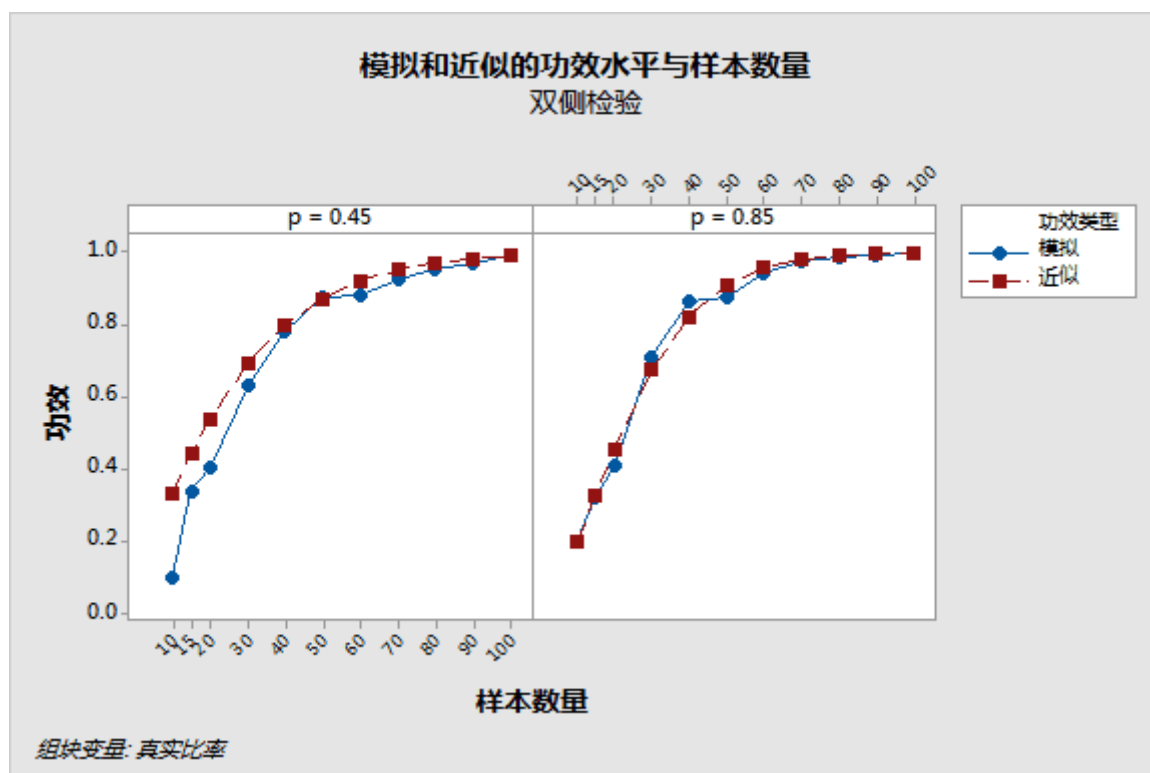


图 1 根据样本量绘制双侧精确检验的模拟和近似功效水平。

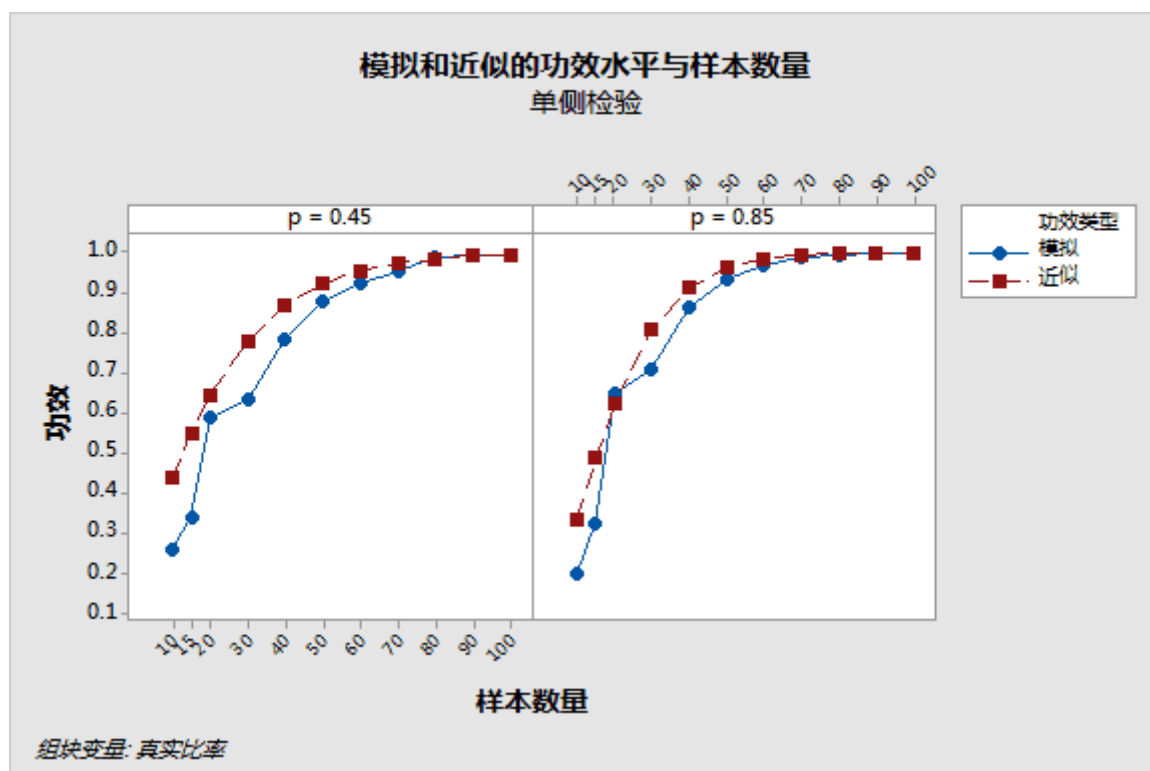


图 2 根据样本量绘制单侧精确检验的模拟和近似功效水平。

当样本量较小时，图 1 和图 2 中的每个面板中显示的两个功效曲线彼此接近，少数情况除外。曲线的接近程度表明，在实际应用精确检验时，近似功效与模拟功效非常接近。因此，可使用近似功效估计样本量。

图 1 和图 2 还表明，理论（近似值）功效曲线一般比模拟功效曲线高。近似功效曲线较高是因为计算理论功效水平时假定目标显著性水平有一个精确值（0.05）。相比之下，精确检验往往比较保守，特别是当样本量较小时，因此获得的实际显著性水平会低于目标水平。因此，当样本量较小时，模拟功效水平也往往较小。

总之，我们的模拟结果表明，正态近似检验的理论功效非常接近精确（概率比）检验的功效。因此，正态近似检验的理论功效将为估计所需的样本量奠定良好的基础，该样本量可确保精确检验的功效足以检测实际存在的重要差异。

© 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](http://minitab.com/legal/trademarks) for more information.