



能力分析（单位缺陷数 / Defects per Unit）



修订日期：2020/8/10

摘要

本过程用于根据该总体的样本估计总体中单位缺陷数的均值。对每个样本项，统计其缺陷数量。

本分析的数据由来自某总体的 m 个样本构成，详述如下：

n_i = 第 i 个样本中的样本项数

d_i = 第 i 个样本中所有样本项的缺陷合计数量

样本 StatFolio: attcap2.sgp

样本数据：

boards.sgd 文件包含 $m = 26$ 个样本的信息，每个样本由 $n = 100$ 块印刷电路板组成。数据取自 Montgomery (2005)。对每块电路板进行了检验，并统计了板上的缺陷数量。下表列出了该文件中的部分数据：

样本	n	缺陷数
1	100	21
2	100	24
3	100	16
4	100	12
5	100	15
6	100	5
7	100	28
8	100	20
9	100	31
10	100	25
11	100	20
12	100	24

数据输入

数据输入对话框用于录入每个样本的相关信息。

- **缺陷数 (Number of Defects)**：一个数值列，包含样本 i 中样本项上发现的总缺陷数 d_i ，每个样本对应一行。
- **样本量 (Sample Sizes)**：样本量 n_i 。若所有样本量相等，可输入数值列名称或一个单独的数值。
- **单位缺陷数目标 (Target defects per unit)**：过程目标平均每个样本项的缺陷数（如有）。如输入，该值将在图上标出。
- **LSL、标称值 (Nominal) 和 USL**：可添加到图表中的可选规格限。
- **选择 (Select)**：子集选择。

分析摘要

分析摘要 (Analysis Summary) 汇总输入数据并显示过程能力的估计值。

过程能力分析（单位缺陷数 / Defects per Unit） - 缺陷数

Process Capability Analysis (Defects per Unit) - defects

Data variable: defects

Target: 0.2

Distribution: Poisson

number of samples = 26

average sample size = 100.0

mean defects per unit = 0.198462

	Estimate	Lower 95% Limit	Upper 95% Limit
Mean defects per unit	0.198462	0.181705	0.216348
Tolerance limits (average size sample)		12	29

数据变量: defects

目标: 0.2

分布: Poisson (泊松)

样本数 = 26

平均样本量 = 100.0

单位缺陷数均值 = 0.198462

	估计值	95% 下限	95% 上限
单位缺陷数均值	0.198462	0.181705	0.216348
容差限 (平均样本量)		12	29

输出中的重要内容包括:

- **分布 (Distribution)** : 数据的假定分布。默认为 *Poisson (泊松)* , 适用于方差近似等于均值的情况。若方差大于均值, 则可改用 *negative binomial (负二项分布)* , 通过 *分析选项 (Analysis Options)* 选择。
- 样本数: 样本数 m 。
- **平均样本量 (Average sample size)** : m 个样本的平均大小:

$$\bar{n} = \sum_{i=1}^m n_i / m$$

- 单位缺陷数均值: 每个样本项的平均缺陷数, 由下式计算:

$$\bar{u} = \frac{\sum_{i=1}^m d_i}{\sum_{i=1}^m n_i}$$

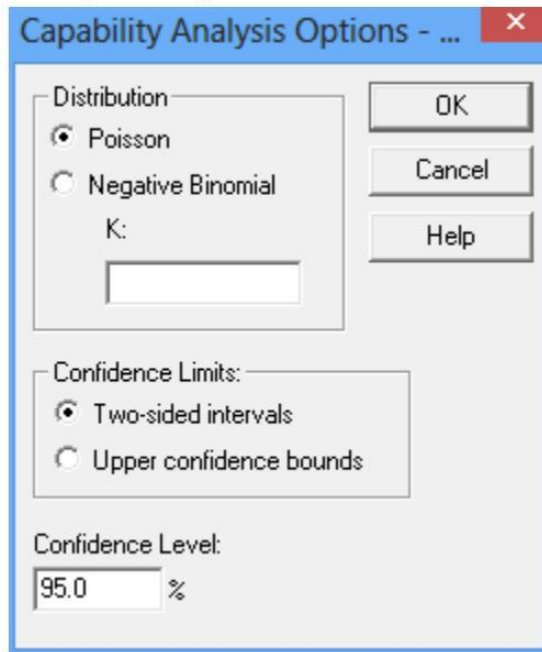
同时显示总体单位缺陷数均值的置信区间。

- 平均大小样本的容差限: 对于大小为 $\bar{n}$ 的样本中缺陷总数的 $100(1 - \alpha)\%$ 容差区间。该区间通过寻找估计分布中 X 的范围来计算, 使每侧尾部不超过 $\alpha/2$ 。

对于样本数据, 估计的单位缺陷数均值 $u = 0.1985$ 。给定该样本量, 误差幅度使得真实单位缺陷数均值的 95% 置信区间介于 0.1817 与 0.2163 之间。

容差区间表明, 从总体中抽取的 100 个样本项的样本, 预期有 95% 包含 12 至 29 个缺陷。

分析选项

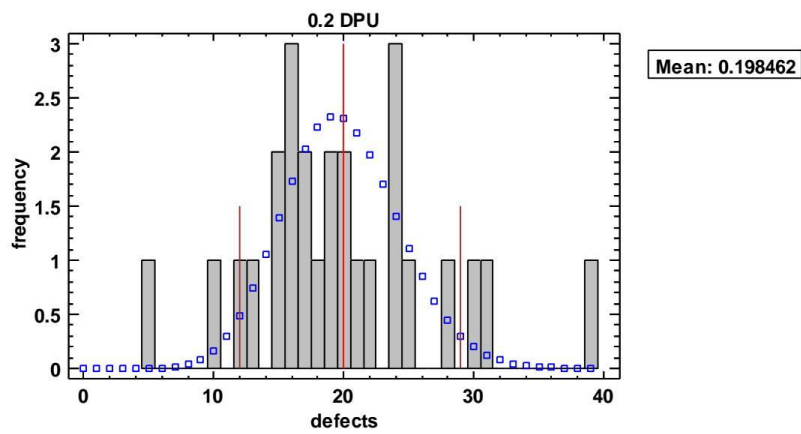


- **分布 (Distribution)**：若总体方差近似等于其均值，选择 *Poisson* (泊松) 分布；若方差大于均值，选择 *Negative Binomial* (负二项) 分布。
- **K**：若选择了负二项分布，可输入参数 k 的值（输入大于 0 的数），或将该字段留空由程序根据数据估计 k 。
- **置信限 (Confidence Limits)**：选择 *双侧区间* (*Two-sided intervals*) 以给出单位缺陷数均值上下界，或选择 *置信上限* (*Upper confidence bounds*) 仅给出上限。
- **置信水平 (Confidence Level)**：指定置信区间和容差限的百分比，通常为 90、95 或 99。

能力图

能力图 (*Capability Plot*) 显示数据的直方图以及拟合分布（由点符号表示）。

缺陷数的过程能力



高竖线对应于目标值（如有）。较短的竖线对应于 分析摘要 中显示的容差限。

图的右边缘显示平均大小样本中缺陷数的估计均值。

拟合优度检验

拟合优度检验 (Goodness-of-Fit Tests) 面板执行卡方检验，以确定数据样本是否合理来自假定分布。

Goodness-of-Fit Tests for defects					
Chi-Squared Test					
	Lower	Upper	Observed	Expected	
	Limit	Limit	Frequency	Frequency	Chi-Squared
at or below		14.0	4	2.89	0.43
	15.0	16.0	5	3.13	1.12
	17.0	17.0	2	2.02	0.00
	18.0	18.0	1	2.23	0.68
	19.0	19.0	2	2.33	0.05
	20.0	20.0	2	2.31	0.04
	21.0	21.0	1	2.18	0.64
	22.0	23.0	1	3.67	1.94
	24.0	25.0	4	2.52	0.87
at or above	26.0		4	2.74	0.58

Chi-Squared = 6.34581 with 8 d.f. P-Value = 0.608556

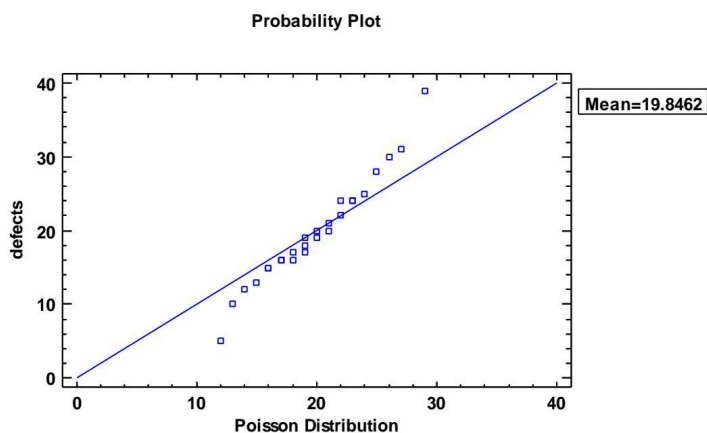
数据的范围被划分为若干区间，每个区间代表样本中可能的缺陷数范围。分组的构造使得每个组中的期望数据值至少为 2。

关注的主要统计量是 *P 值 (P-Value)*。低于该值的 *P* 值表明在 100% 显著性水平下与假定分布存在显著偏离。例如，由于上表中的 *P* 值为 0.608556，远高于 0.05，因此在 5% 显著性水平下，与假定 Poisson（泊松）分布不存在显著偏离。

若样本量全部相等，则该检验是精确的；若样本量不等，则为近似检验。

概率图

概率图 (Probability Plot) 用于评估数据与假定分布之间的任何差异。



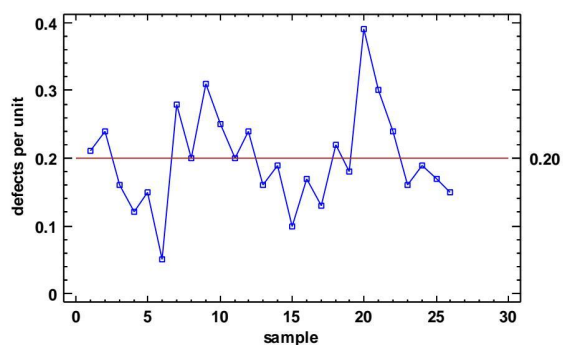
纵轴显示数据，从小到大排序，表示样本中的缺陷数。每个点相对于拟合分布的等效百分位绘制。考虑到数据的离散性，点应大致落在对角线上。

在上图中，数据显示出与 Poisson（泊松）分布的显著偏离，数据在两侧都比 Poisson（泊松）分布预期的延伸得更远。

运行图

该图按顺序排列显示数据，并在目标值（如有）处绘制一条水平线。

Run Chart for defects

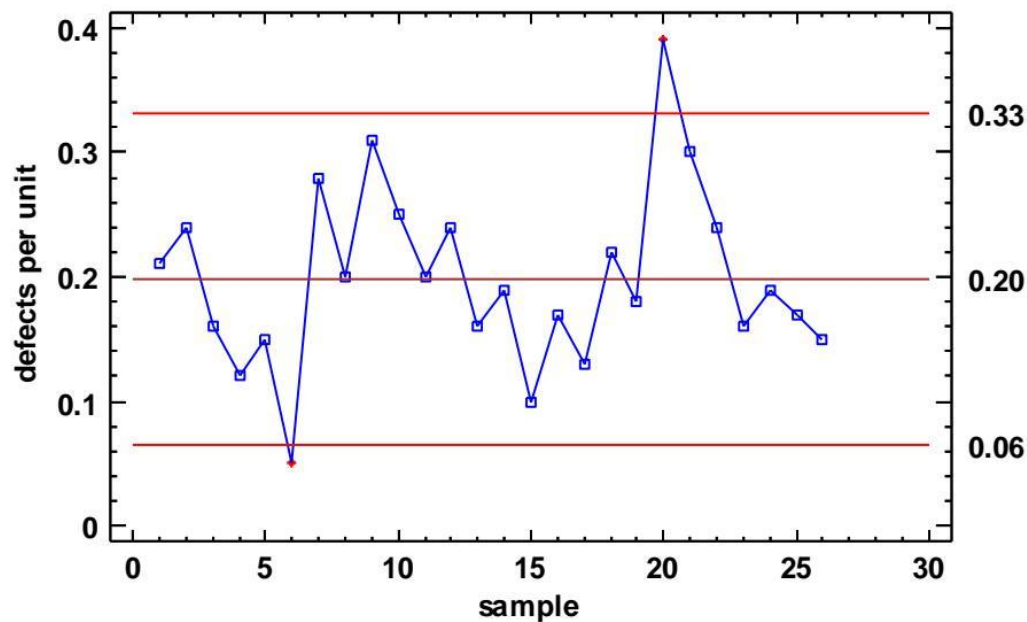


理想情况下，点应围绕目标线随机波动。

控制图

该面板显示样本数据的 u 图 (u-chart) 。

缺陷数的 u 图



每个数据值与中心线及控制限一起绘制。在本例中，控制限之外的 2 个点表明过程未处于统计控制状态，这很可能就是出现明显长尾的原因。

备选分布比较

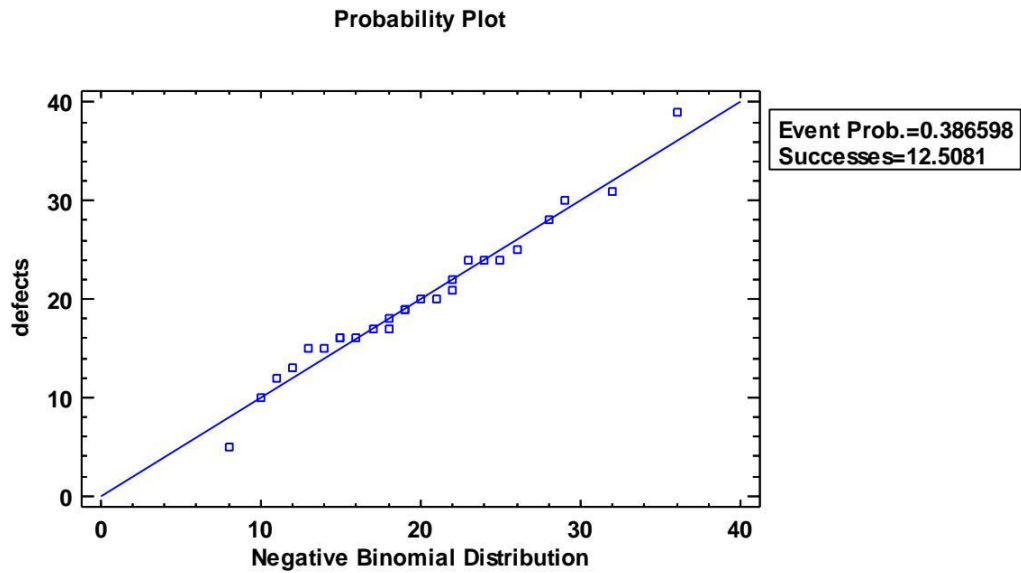
该面板显示 Poisson（泊松）分布和负二项分布的拟合优度检验结果。

Comparison of Alternative Distributions			
Distribution	K	Log Likelihood	Chi-Squared P
Poisson		-94.6698	0.608556
Negative Binomial	12.5081	-87.2331	0.610756

拟合更好的分布将具有：

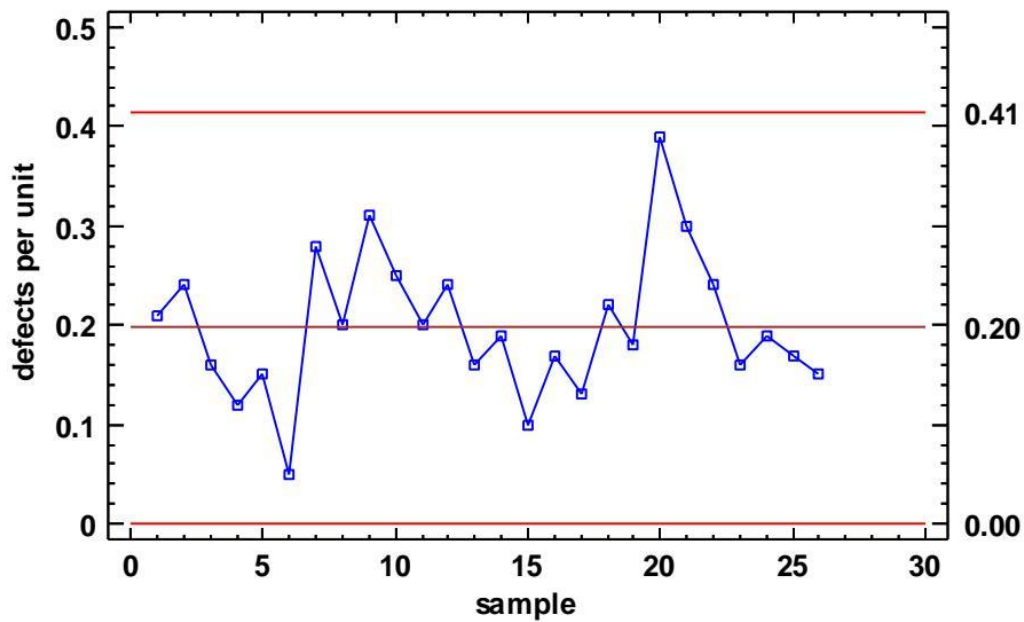
- 更大的对数似然函数值。
- 更大的卡方 P 值。

在上表中，负二项分布的统计量更大，表明改用该分布可能会得到更好的拟合。拟合的负二项分布如下所示：



由于拟合分布的方差更大，控制限比 Poisson（泊松）分布略宽：

缺陷数的 u 图



使用负二项分布而非 Poisson（泊松）分布的最大影响在于 95% 容差区间，该区间现在从 8 到 36，而非从 12 到 29。分布是否为真正的负二项分布，或更好的拟合是失控过程的表现，需要进一步调查。

符合性分析

符合性分析可用于确定过程符合以每包不合格品数量表示的规格的程度。根据 Perakis 和 Xekalaki 的研究，假设被研究的产品以每包含 n 个样本项的包装发货（ n 可等于 1）。进一步假设每个箱子应包含不超过 U 个不合格品（分布于该包的样本项中）。他们建议计算指数

$$C_{pc} = \frac{1 - p_0}{1 - p}$$

其中 p 是包含所有不超过 U 个不合格品的包装的估计比例， p_0 是要求的符合比例。他们还建议 $p_0 = 0.9973$ 是一个合理的值，对应于正态曲线落在均值 3 个标准差范围内的比例。 p 由拟合的 Poisson（泊松）或负二项分布计算。

对于样本数据，生成以下输出：

Conformance Analysis			
Distribution: Poisson			
mean defects per unit = 0.198462			
Package size: 500			
Maximum allowable nonconformities: 100			
Target conformance: 99.73%			
	<i>Estimate</i>	<i>Lower 95% Limit</i>	<i>Upper 95% Limit</i>
Prob. of conformance	0.557194	0.232513	0.844284
Cpc	0.00609748	0.00351798	0.0173392

符合性分析

分布：Poisson（泊松）

单位缺陷数均值 = 0.198462

包装大小：500

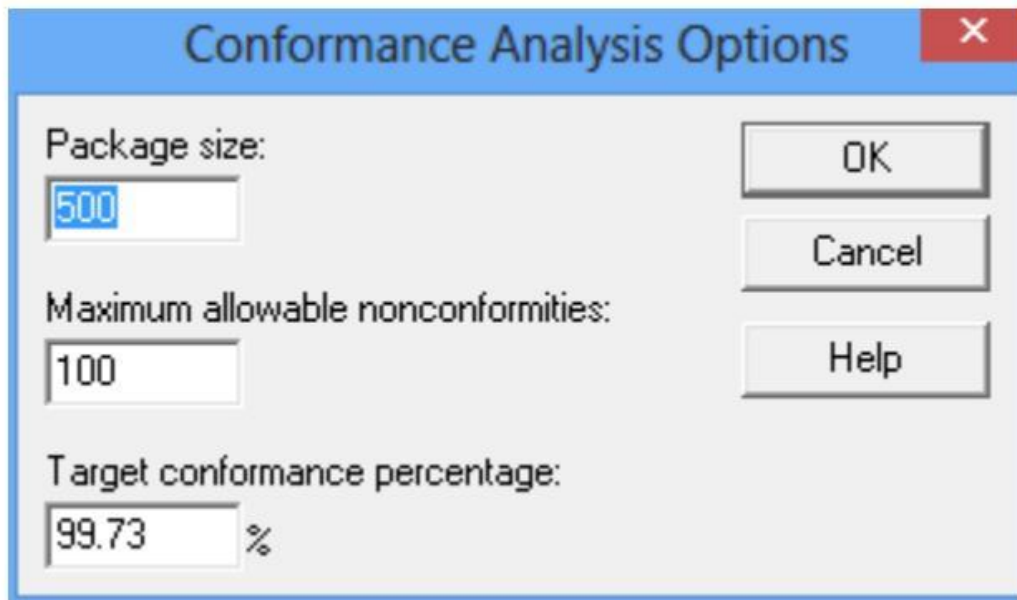
最大允许不合格品数：100

目标符合度：99.73%

	估计值	95% 下限	95% 上限
符合概率	0.557194	0.232513	0.844284
Cpc	0.00609748	0.00351798	0.0173392

给定每个包含 $n = 500$ 个样本项的包装，以及每包允许不合格品数的上限 $U = 100$ ，生产符合要求的包装的概率 $p = 0.557$ 。与 99.73% 符合包装的目标相比，能力指数 $C_{pc} = 0.0061$ 。同时显示了符合概率和 C_{pc} 的 95% 置信限。

面板选项



The image shows a software dialog box titled "Conformance Analysis Options". It contains three input fields and three buttons. The first field is "Package size:" with the value "500". The second field is "Maximum allowable nonconformities:" with the value "100". The third field is "Target conformance percentage:" with the value "99.73 %". To the right of the fields are three buttons: "OK", "Cancel", and "Help".

- 包装大小 (**Package size**) : 每个包装中的样本项数 n 。

statgraphics

- 最大允许不合格品数 (**Maximum allowable nonconformities**) : 包装中允许的最大不合格品数 U 。
- 目标符合百分比 (**Target conformance percentage**) : 必须具有 U 或更少不合格品的包装的目标百分比 T 。

参考文献

Perakis, Michael 和 Xekalaki, Evdokia (2005). 离散过程的过程能力指数 (A process capability index for discrete processes) . Journal of Statistical Computation and Simulation 75, 第 175-187 页。

计算

单位缺陷数均值的置信区间

对于 Poisson (泊松) 分布:

$$\left[\frac{X_{1-\alpha/2, 2x}^2}{2 \sum_{i=1}^m n_i}, \frac{X_{\alpha/2, 2(x+1)}^2}{2 \sum_{i=1}^m n_i} \right]$$

(3)

其中

$$x = \sum_{i=1}^m d_i$$

若假设数据服从负二项分布，则使用正态近似计算置信区间：

$$\bar{u} \pm z_{\alpha/a} \sqrt{\frac{\hat{k}}{m\bar{n}^2} \frac{1-\hat{p}}{\hat{p}^2}}$$

若在 **分析选项 (Analysis Options)** 对话框中 k 字段留空，则两个参数都将根据数据估计：

$$\hat{p} = \sqrt{\frac{\bar{u}\bar{n}}{s_d^2}}$$

其中 s_d^2 是 d_i 的样本方差 (6)

$$\hat{k} = \frac{\hat{p}\bar{u}\bar{n}}{1-\hat{p}} \quad (7)$$

若 $\hat{p}$ 大于 1 则会发出错误消息，这将在数据的方差小于均值时发生。

若在 **分析选项** 对话框中提供了 k 值，则仅根据下式从数据估计 p ：

$$\hat{p} = \frac{k}{\bar{d} + k}$$

控制图

中心线：

$$\bar{u} = \frac{\sum_{i=1}^m d_i}{\sum_{i=1}^m n_i}$$

控制限：

$$\bar{u} \pm 3 \sqrt{\frac{\bar{u}}{\bar{n}}}$$

若为 Poisson (泊松) (10)

$$\bar{u} \pm 3 \sqrt{\frac{\hat{k}(1 - \hat{p})}{\bar{n}^2 \hat{p}^2}}$$

若为负二项 (11)