Capability Analysis (Defects Per Unit)

This procedure is designed to estimate the mean number of defects per unit in a population based on samples of items from that population. For each item, the number of defects is counted.

The data for this analysis consist of *m* samples from a population detailing:

n_i = number of items in sample *i*

d_i = number of defects summed over all items in sample *i*

Sample StatFolio: *attcap2.sgp*

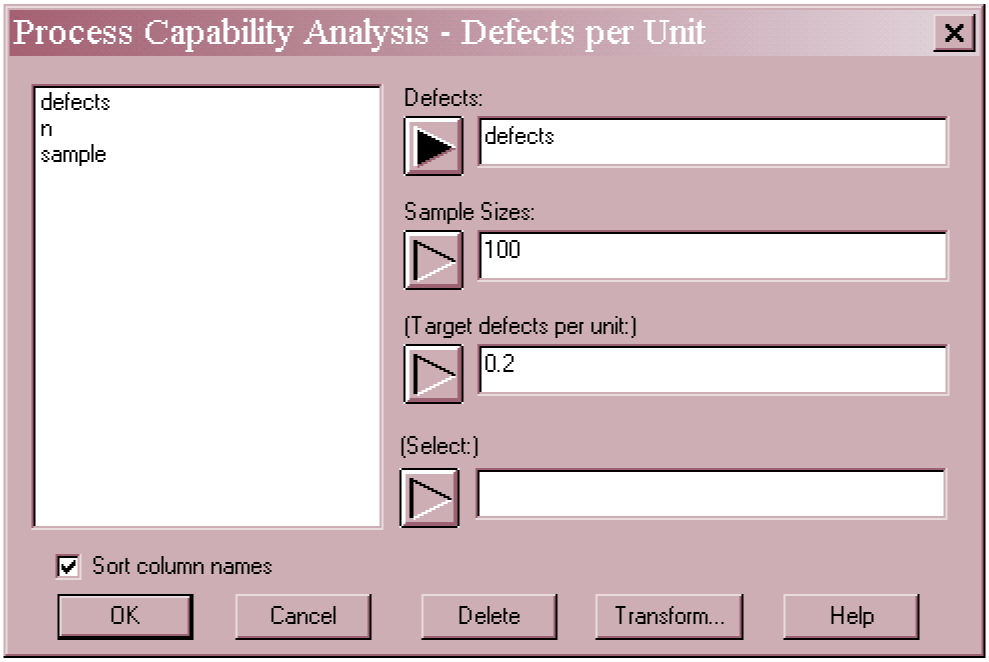
Sample Data:

The file *boards.sgd*contains the information on *m* = 26 samples, each consisting of *n* = 100 printed circuit boards. The data is taken from Montgomery (2005). Each board was inspected and the number of defects on the board was tabulated. The table below shows a partial list of the data in that file:

Sample	n	Defects
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
13	100	16
14	100	19
15	100	10

Data Input

The data input dialog requests information about each sample.



- **Number of Defects**: a numeric column containing the total number of defects d_i found on the items in sample *i*, one row for each sample.

- **Sample Sizes**: the sample sizes n_i. Enter either the name of a numeric column or a single number if all sample sizes are equal.

- **Target defects per unit**: the process target percent mean defects per item, if any. If entered, this value will be indicated on the graphs.

- **Select:** subset selection.

Analysis Summary

The *Analysis Summary* summarizes the input data and displays estimates of the process capability.

Process Capability Analysis - 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 for sample of average size		12	29

Important items in the output include:

- **Distribution**: the assumed distribution for the data. The default is *Poisson*, which is appropriate if the variance is approximately equal to the mean. If the variance is larger than the mean, then the *negative binomial* distribution may be selected instead using *Analysis Options*.

- **Number of samples**: the number of samples *m*.

- **Average sample size**: the average size of the *m* samples:

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

- **Mean defects per unit**: the average number of defects per item, calculated from

$$\bar{u} = \frac{\sum_{i=1}^{m} d_i}{\sum_{i=1}^{m} n_i} \tag{2}$$

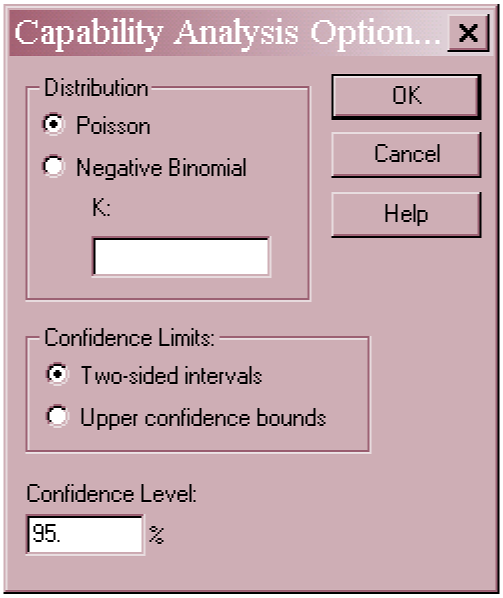
  A confidence interval for the population mean defects per item is also displayed.

- **Tolerance limits for sample of average size:** a 100(1-α)% tolerance interval for the total number of defects in samples of size $\bar{n}$. This interval is calculated by finding the range of X for the estimated distribution leaving no more than α/2 in each tail.

For the sample data, the estimated mean defects per unit $\bar{u}$ = 0.1985. Given the sample size, the margin of error is such that a 95% confidence interval for the true mean defects per unit ranges from 0.1817 to 0.2163.

The tolerance interval states that 95% of all samples of 100 items taken from the population can be expected to contain between 12 and 29 defects.

Analysis Options



- **Distribution**: select the *Poisson* distribution if the population variance is approximately equal to its mean. If the variance is larger than the mean, select the *Negative Binomial* distribution.

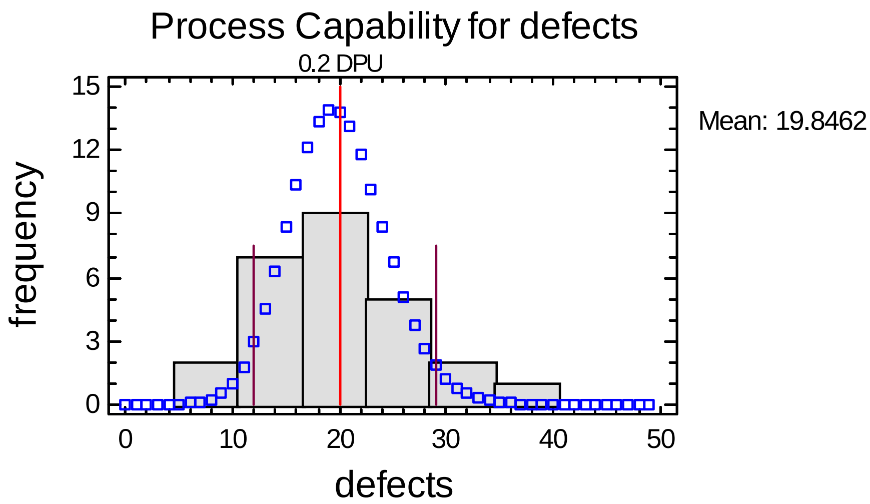
- **K**: if the negative binomial distribution is selected, you can specify the value for the parameter *k* by entering a number greater than 0 or leave the field blank if you want the program to estimate *k* from the data.

- **Confidence Limits**: select *Two-sided intervals* for upper and lower bounds around the mean defects per unit or *Upper confidence bounds* if only an upper limit is desired.

- **Confidence Level**: specify the percentage for the confidence interval and tolerance limits, usually 90, 95, or 99.

Capability Plot

The *Capability Plot* shows a histogram of the data together with the fitted distribution (shown by the point symbols).



The tall vertical line corresponds to the target, if any. The shorter vertical lines correspond to the tolerance limits shown in the *Analysis Summary*.

The right margin of the plot displays the estimated mean number of defects in a sample of average size.

Goodness-of-Fit Tests

The *Goodness-of-Fit Tests* pane performs a chi-squared test to determine whether the data sample may reasonably have come from the assumed distribution.

Goodness-of-Fit Tests for defects
Chi-Squared Test

	Lower Limit	*Upper Limit*	*Observed Frequency*	*Expected 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

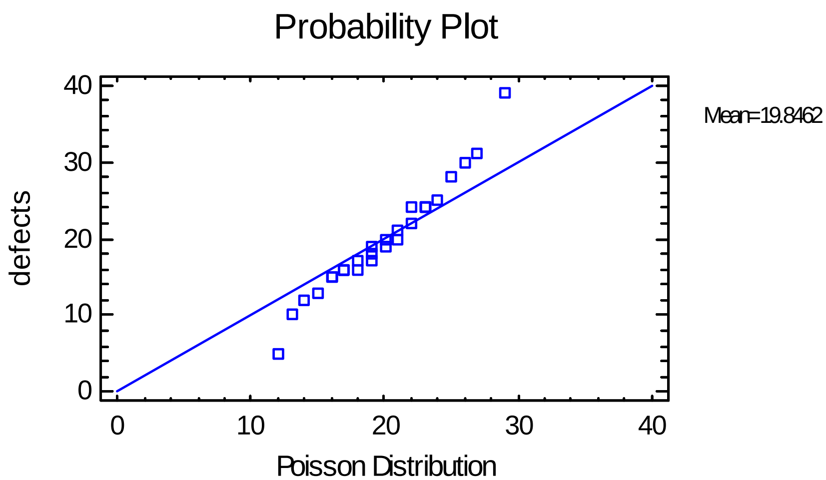
The range of the data is divided into intervals, each representing a range of possible numbers of defects in a sample. Classes are structured such that the expected number of data values in each class is at least 2.

The primary statistic of interest is the *P-Value*. P-values below α indicate significant departures from the assumed distribution at the 100α% significance level. For example, since the P-Value in the table above is well above 0.05, there is not a significant departure from the assumed Poisson distribution at the 5% significance level.

The test is exact if the sample sizes are all equal. It is an approximation if the sample sizes are unequal.

Probability Plot

The *Probability Plot* is used to assess any discrepancy between the data and the assumed distribution.

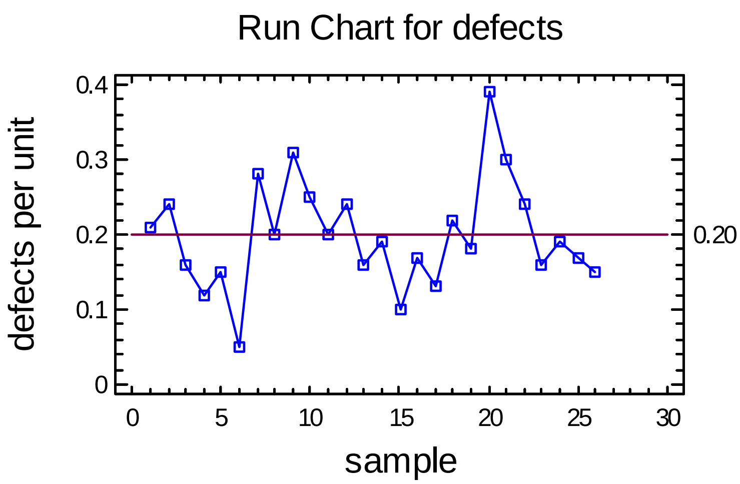


The vertical axis shows the data, sorted from smallest to largest, shows the number of defects in a sample. Each point is plotted versus an equivalent percentile of the fitted distribution. Allowing for the discreteness of the data, the points should fall approximately along the diagonal line.

In the above plot, the data show what appears to be a significant departure from the Poisson distribution, with the data extending out further in both directions than would be expected for a Poisson distribution.

Run Chart

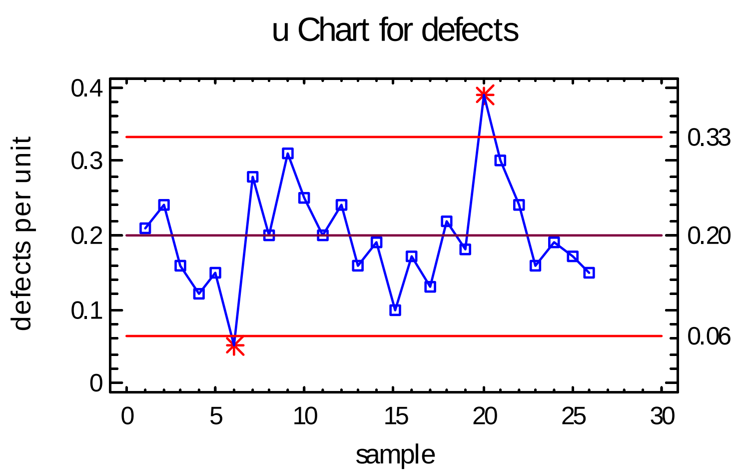
This plot shows the data in sequential order with a horizontal line drawn at the target value, if any.



Ideally, the points will vary randomly around the target line.

Control Chart

This pane shows a u-chart for the sample data.



Each data value if plotted together with a centerline and control limits. In this case, the 2 points outside the control limits are signals that the process is *not* in a state of statistical control, which probably accounts for the apparent long tails.

Comparison of Alternative Distributions

This pane displays the results of goodness-of-fit tests for the Poisson distribution and negative binomial distribution.

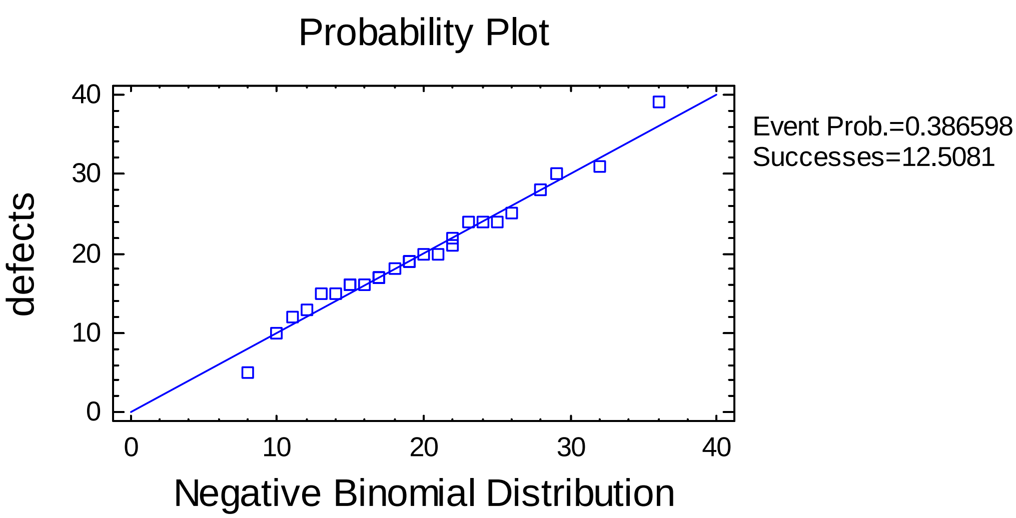
Comparison of Alternative Distributions

Distribution	*K*	*Log Likelihood*	*Chi-Squared P*
Poisson		-94.6698	0.608556
Negative Binomial	12.5081	-87.2331	0.610756

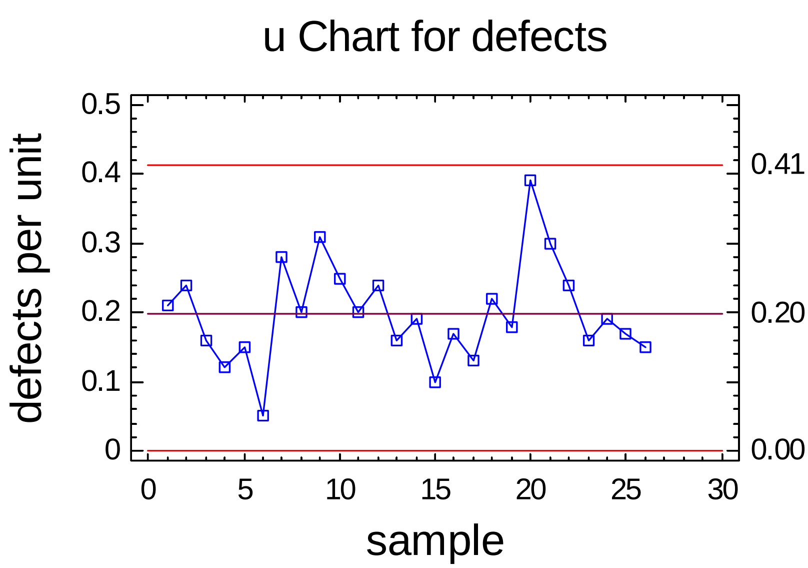
Better fitting distributions will have:

- larger values of the log likelihood function.
- larger values of the Chi-Squared *P-value*.

In the above table, the statistics are larger for the negative binomial distribution, suggesting that switching to it might give a better fit. The fitted negative binomial distribution is shown below:



The control limits are somewhat wider than for the Poisson distribution, since the variance of the fitted distribution is greater:



The biggest impact of using the negative binomial distribution rather than the Poisson distribution is on the 95% tolerance interval, which now extends from 8 to 36 instead of from 12 to 29. Whether the distribution is really negative binomial or the better fit is a manifestation of an out-of-control process would require further investigation.

Calculations

Confidence Interval for Mean Defects per Unit

For Poisson distribution:

$$\left[\frac{X^2_{1-\alpha/2,2x}}{2\sum_{i=1}^{m} n_i}, \frac{X^2_{\alpha/2,2(x+1)}}{2\sum_{i=1}^{m} n_i} \right] \quad (3)$$

where

$$x = \sum_{i=1}^{m} d_i \quad (4)$$

If the data are assumed to followed a negative binomial distribution, then the confidence interval is calculated using a normal approximation:

$$\bar{u} \pm z_{\alpha/a} \sqrt{\frac{\hat{k}}{m\bar{n}^2} \frac{1-\hat{p}}{\hat{p}^2}} \quad (5)$$

If the field for *k* on the **Analysis Options** dialog box is left blank, then both parameters will be estimated from the data:

$$\hat{p} = \sqrt{\frac{\bar{u}\bar{n}}{s_d^2}} \qquad \text{where } s_d^2 \text{ is the sample variance of the } d_i\text{'s} \quad (6)$$

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

An error message is issued if $\hat{p}$ is greater than 1, which will occur if the variance of the data is less than the mean.

If a value for *k* is supplied on the *Analysis Options* dialog box, then only *p* is estimated from the data according to

$$\hat{p} = \frac{k}{\bar{d} + k} \quad (8)$$

Control Chart

Centerline:

$$\bar{u} = \frac{\sum_{i=1}^{m} d_i}{\sum_{i=1}^{m} n_i} \tag{9}$$

Control limits:

$$\bar{u} \pm 3\sqrt{\frac{\bar{u}}{\bar{n}}} \quad \text{if Poisson} \tag{10}$$

$$\bar{u} \pm 3\sqrt{\frac{\hat{k}(1-\hat{p})}{\bar{n}^2 \hat{p}^2}} \quad \text{if negative binomial} \tag{11}$$