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:

$$n_i = \text{number of items in sample } i$$

$$d_i = \text{number of defects summed over all items in sample } i$$

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:

<table>
<thead>
<tr>
<th>Sample</th>
<th>n</th>
<th>Defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>31</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td>100</td>
<td>24</td>
</tr>
<tr>
<td>13</td>
<td>100</td>
<td>16</td>
</tr>
<tr>
<td>14</td>
<td>100</td>
<td>19</td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>10</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Process Capability Analysis - defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data variable: defects</td>
</tr>
<tr>
<td>Target: 0.2</td>
</tr>
<tr>
<td>Distribution: Poisson</td>
</tr>
<tr>
<td>number of samples = 26</td>
</tr>
<tr>
<td>average sample size = 100.0</td>
</tr>
<tr>
<td>mean defects per unit = 0.198462</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Lower 95% Limit</th>
<th>Upper 95% Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean defects per unit</td>
<td>0.198462</td>
<td>0.181705</td>
<td>0.216348</td>
</tr>
<tr>
<td>Tolerance limits for sample of average size</td>
<td></td>
<td>12</td>
<td>29</td>
</tr>
</tbody>
</table>

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:

  $$\bar{n} = \frac{\sum_{i=1}^{m} n_i}{m}$$  \hspace{1cm} (1)

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

  $$\bar{d} = \frac{\sum_{i=1}^{m} d_i}{\sum_{i=1}^{m} n_i}$$  \hspace{1cm} (2)

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

- **Tolerance limits for sample of average size**: a $100(1-\alpha)%$ 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 $\alpha/2$ in each tail.

For the sample data, the estimated mean defects per unit $\bar{d} = 0.1985$. 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.

<table>
<thead>
<tr>
<th>Goodness-of-Fit Tests for defects</th>
<th>Chi-Squared Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Limit</td>
</tr>
<tr>
<td>at or below</td>
<td>14.0</td>
</tr>
<tr>
<td>15.0</td>
<td>16.0</td>
</tr>
<tr>
<td>17.0</td>
<td>17.0</td>
</tr>
<tr>
<td>18.0</td>
<td>18.0</td>
</tr>
<tr>
<td>19.0</td>
<td>19.0</td>
</tr>
<tr>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>21.0</td>
<td>21.0</td>
</tr>
<tr>
<td>22.0</td>
<td>23.0</td>
</tr>
<tr>
<td>24.0</td>
<td>25.0</td>
</tr>
<tr>
<td>at or above</td>
<td>26.0</td>
</tr>
</tbody>
</table>

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 $\alpha$ indicate significant departures from the assumed distribution at the 100$\alpha$% 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.

![Probability Plot](image)

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.

![Run Chart for defects](image)

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

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

![u Chart for defects](image)

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

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$K$</th>
<th>Log Likelihood</th>
<th>Chi-Squared P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>-94.6698</td>
<td>0.608556</td>
<td></td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>12.5081</td>
<td>-87.2331</td>
<td>0.610756</td>
</tr>
</tbody>
</table>

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:

![Probability Plot](image)

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

![u Chart for defects](image)
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:

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

(3)

where

\[ x = \sum_{i=1}^{m} d_i \]  

(4)

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

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

(5)

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

\[
\hat{p} = \sqrt{\frac{\bar{u} n}{s_d^2}} \quad \text{where } s_d^2 \text{ is the sample variance of the } d_i \text{'s}
\]

(6)

\[
\hat{k} = \frac{\hat{p} \bar{u} n}{1 - \hat{p}}
\]

(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

\[
\hat{p} = \frac{k}{d + k}
\]

(8)
Control Chart

Centerline:

\[
\overline{d} = \frac{\sum_{i=1}^{m} d_i}{\sum_{i=1}^{m} n_i}
\]  \hspace{1cm} (9)

Control limits:

\[
\overline{u} \pm 3 \sqrt{\frac{\overline{d}}{n}} \quad \text{if Poisson} \hspace{1cm} (10)
\]

\[
\overline{u} \pm 3 \sqrt{\frac{k(1-\hat{p})}{n^2 \hat{p}^2}} \quad \text{if negative binomial} \hspace{1cm} (11)
\]