Capability Analysis (Percent Defective)

This procedure is designed to estimate the percentage of defective items in a population based on samples of items from that population that have been classified as either *defective* or *non-defective*. Note: some practitioners prefer the term "nonconforming" over "defective".

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 defective items in sample i

Sample StatFolio: *attcap1.sgp*

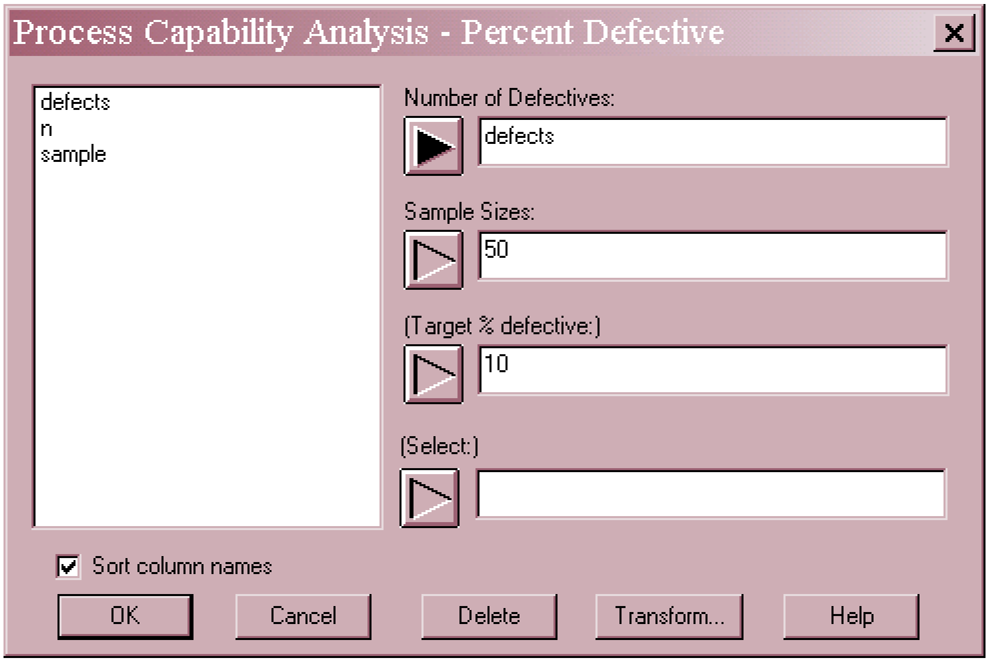
Sample Data:

The file *juice.sgd* contains the measured bursting strength of m = 30 samples, each corresponding n = 50 cardboard orange juice cans. The data is taken from Montgomery (2005). Each can was inspected and the number of defective cans was tabulated. The table below shows a partial list of the data from that file:

Sample	n	Defects
1	50	12
2	50	15
3	50	8
4	50	10
5	50	4
6	50	7
7	50	16
8	50	9
9	50	14
10	50	10
11	50	5
12	50	6
13	50	17
14	50	12
15	50	22

Data Input

The data input dialog requests information about each sample.



- **Number of Defectives**: a numeric column containing the number of defectives d_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 % defective**: the process target percent defective items, 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: 10.0

Distribution: Binomial
number of samples = 30
average sample size = 50.0
mean percent defective = 23.1333

	Estimate	*Lower 95% Limit*	*Upper 95% Limit*
Mean percent defective	23.1333	21.0203	25.3521
Defects per million	231333.0	210203.0	253521.0
Process Z	0.734465	0.80572	0.663453
Sigma quality level	2.23	2.31	2.16
Tolerance limits (average size sample)		6	18

Important items in the output include:

- **Distribution**: the assumed distribution for the data. The default is *binomial*, which is appropriate if the samples are taken from large lots, typically lots at least 10 times larger than the sample size. If the lots are small, then the *hypergeometric* 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 percent defective**: the average percent of defective items:

$$\bar{p} = 100 \frac{\sum_{i=1}^{m} d_i}{\sum_{i=1}^{m} n_i} \% \tag{2}$$

  A confidence interval for the population mean percentage is also displayed.

- **Process Z**: the equivalent Z value that equates to the estimated defects per million. The process Z is calculated by finding the value of a standard normal distribution for which the probability of exceeding |Z| equals the estimated DPM. Larger values of Z are preferable, since they correspond to smaller percentages of defective items. Typical targets for Z are 4 or larger.

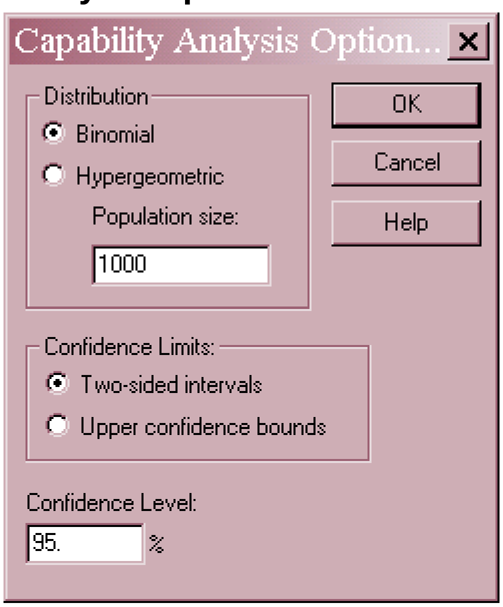
- **Sigma Quality Level:** the Sigma Quality Level as defined by Six Sigma practitioners. This item is only displayed if *Sigma quality level* is selected on the *Capability* tab of the *Preferences* selection on the *Edit* menu. The Sigma Quality Level equals the **Process Z** or the **(Process Z + 1.5),** depending on whether *1.5 sigma shift* is selected on the *Capability* tab of the *Preferences* dialog box.

- **Tolerance limits for sample of average size:** a 100(1-α)% tolerance interval for the number of defective items 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 proportion of defective items $\bar{p}$ = 23.13%. This equates to 231,333 nonconforming items out of every million produced. Given the sample size, the margin of error is such that a 95% confidence interval for the true mean percent defective ranges from 21.02% to 25.35%. The process Z score is a rather dismal 0.73.

The tolerance interval states that 95% of all samples of 50 items taken from the population can be expected to contain between 6 and 18 defective items.

Analysis Options



- **Distribution**: select the binomial distribution if the population size if large compared to the average sample size (at least 10 times as large) or the hypergeometric distribution if the population size is small compared to the sample size.

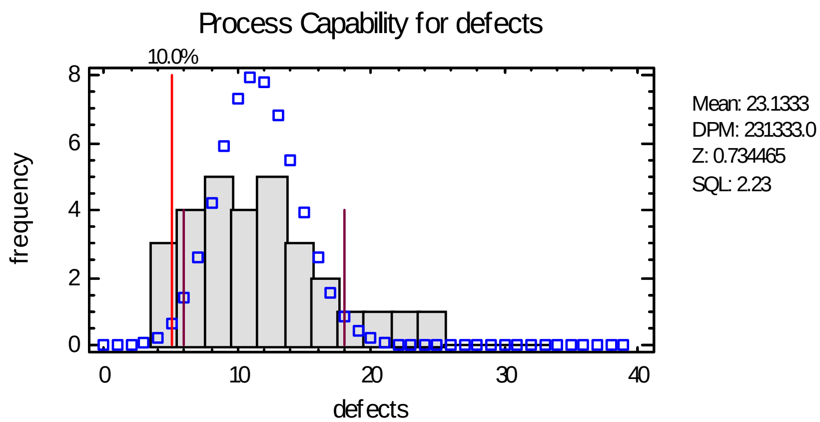
- **Population Size**: if the hypergeometric distribution is selected, specify the size of the population from which the samples are taken.

- **Confidence Limits**: select *Two-sided intervals* for upper and lower bounds around the mean percent defective 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:

- *Mean*: the estimated mean percent defective items
- *DPM*: the estimated defects per million.
- *Z*: the calculated Z value.
- *SQL*: the calculated Sigma Quality Level.

The SQL is displayed only if the associated checkbox is selected on the *Capability* tab of the *Preferences* dialog box, accessible from the *Edit* menu.

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		7.0	7	2.43	8.56
	8.0	8.0	2	2.10	0.00
	9.0	9.0	3	2.95	0.00
	10.0	10.0	3	3.64	0.11
	11.0	11.0	1	3.98	2.23
	12.0	12.0	3	3.89	0.20
	13.0	13.0	2	3.42	0.59
	14.0	14.0	1	2.72	1.09
at or above	15.0		8	4.86	2.02

Chi-Squared = 14.8194 with 7 d.f. P-Value = 0.0383857

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

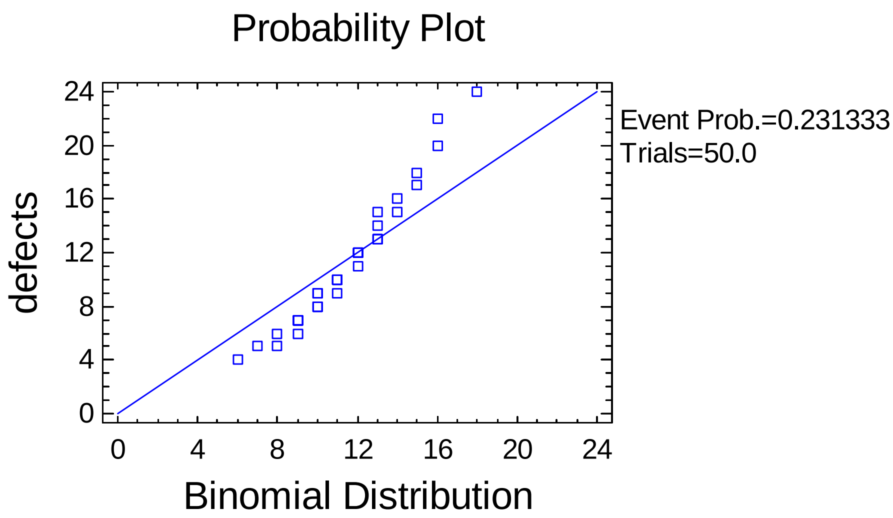
The primary statistic of interest is the *P-Value*. P-values below α indicate significant departures from the assumed distribution at the 100α% significance level. For example, since the P-Value in the table above is less than 0.05, there is a significant departure from the assumed binomial distribution at the 5% significance level.

If the test indicates a significant departure, any large values in the column labeled *Chi-Squared* show the location of greatest discrepancy. For example, the above table indicates that 7 samples had less than or equal to 7 defectives, while only 2.43 would have been expected to be that small given a binomial distribution with a mean of 23.1% defectives. One possible explanation for the failure of the binomial distribution to fit the data well is that the process may not be in a state of statistical control, which can be examined by selecting *Control Chart* from the list of graphical options.

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.

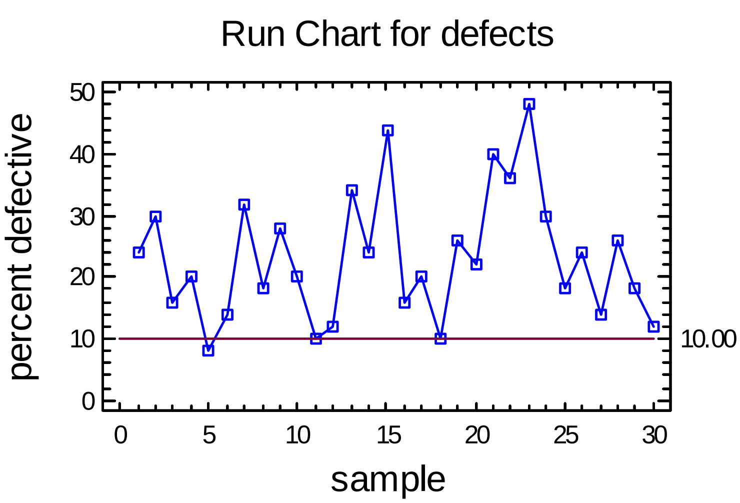


The vertical axis shows the data, sorted from smallest to largest. Each point is plotted versus an equivalent percentile of the fitted distribution. Allowing for the discreteness of the data, the points should fall approximately along the diagonal line.

In the above plot, the data show obvious curvature, seriously questioning the assumption of a single binomial distribution.

Run Chart

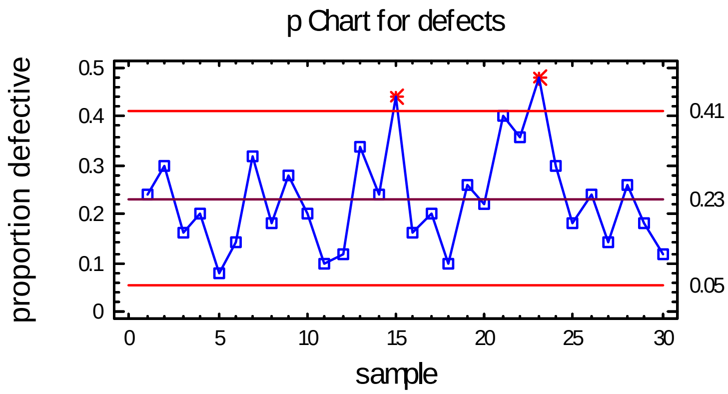
This plot shows the data in sequential order with a horizontal line drawn at the target value, if any.



Ideally, the points will vary randomly around the target line. In this case, the data are consistently above the target.

Control Chart

This pane shows a p-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 upper control limit are signals that the process is *not* in a state of statistical control.

Comparison of Alternative Distributions

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

Comparison of Alternative Distributions

Distribution	*Specified N*	*Log Likelihood*	*Chi-Squared P*
Binomial		-101.304	0.0383857
Hypergeometric	1000	-102.787	0.0201058

The lot size assumed for the hypergeometric distribution is specified on the *Analysis Options* dialog box.

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 binomial distribution, suggesting that switching to the hypergeometric distribution would not improve the fit.

Calculations

Confidence Interval for Mean Percent Defective

For binomial distribution:

$$\left[\frac{100v_1 F_{1-\alpha/2,v_1,v_2}}{v_2 + v_1 F_{1-\alpha/2,v_1,v_2}}\%, \frac{100v_3 F_{\alpha/2,v_3,v_4}}{v_4 + v_3 F_{\alpha/2,v_3,v_4}}\%\right] \tag{3}$$

where

$$v_1 = 2x,\ v_2 = 2(n - x + 1),\ v_3 = 2(x + 1),\ v_4 = 2(n - x),\ n = \sum_{i=1}^{n} n_i \tag{4}$$

$$x = \frac{n\bar{p}}{100} \tag{5}$$

If a hypergeometric distribution is selected the above interval is shortened by a factor of $\sqrt{(mN - mn)/(mN - 1)}$, where N is the size of the populations from which each of the m samples was taken.

Control Chart

Centerline:

$$\bar{p} = \frac{\sum_{i=1}^{m} d_i}{\sum_{i=1}^{m} n_i} \tag{6}$$

Binomial control limits:

$$\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}}} \tag{7}$$

Hypergeometric control limits:

$$\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}} \frac{(N-\bar{n})}{(N-1)}} \tag{8}$$