Capability Assessment SnapStat

Summary

The **Capability Assessment** SnapStat creates a one-page summary of a capability analysis for variables data. There are two SnapStats: one for individuals data collected one at a time, and one for data collected in subgroups. The analysis performed is the same as for the *Capability Analysis* procedures for variable data. However, the output here is formatted to fit on a single page.

Sample StatFolio: *capabilitysnapstat.sgp*

Sample Data:

The file *bottles.sgd* contains the measured bursting strength of $n = 100$ glass bottles, similar to a dataset contained in Montgomery (2005). Each row consists of a sample tested at 10 minute intervals. The table below shows a partial list of the data from that file:

strength	time
255	0:10
232	0:20
282	0:30
260	0:40
255	0:50
233	1:00
240	1:10
255	1:20
254	1:30
259	1:40
235	1:50
262	2:00

Bottles are required to have a bursting strength between 200 psi and 300 psi.

Data Input

There are two SnapStats that perform a capability analysis for variables data, one for individuals data and one for grouped data.

Case #1: Individuals
The data to be analyzed consist of a single numeric column containing *n* observations. The data are assumed to have been taken one at a time.

- **Data:** numeric column containing the data to be analyzed.
- **Date/Time/Labels:** optional labels for each observation.
- **USL:** upper specification limit, if any.
- **Nominal**: optional nominal or target value for the variable. If not supplied, certain capability indices will not be calculated.
- **LSL**: lower specification limit, if any.
- **Select:** subset selection.

At least one of the specification limits must be entered.
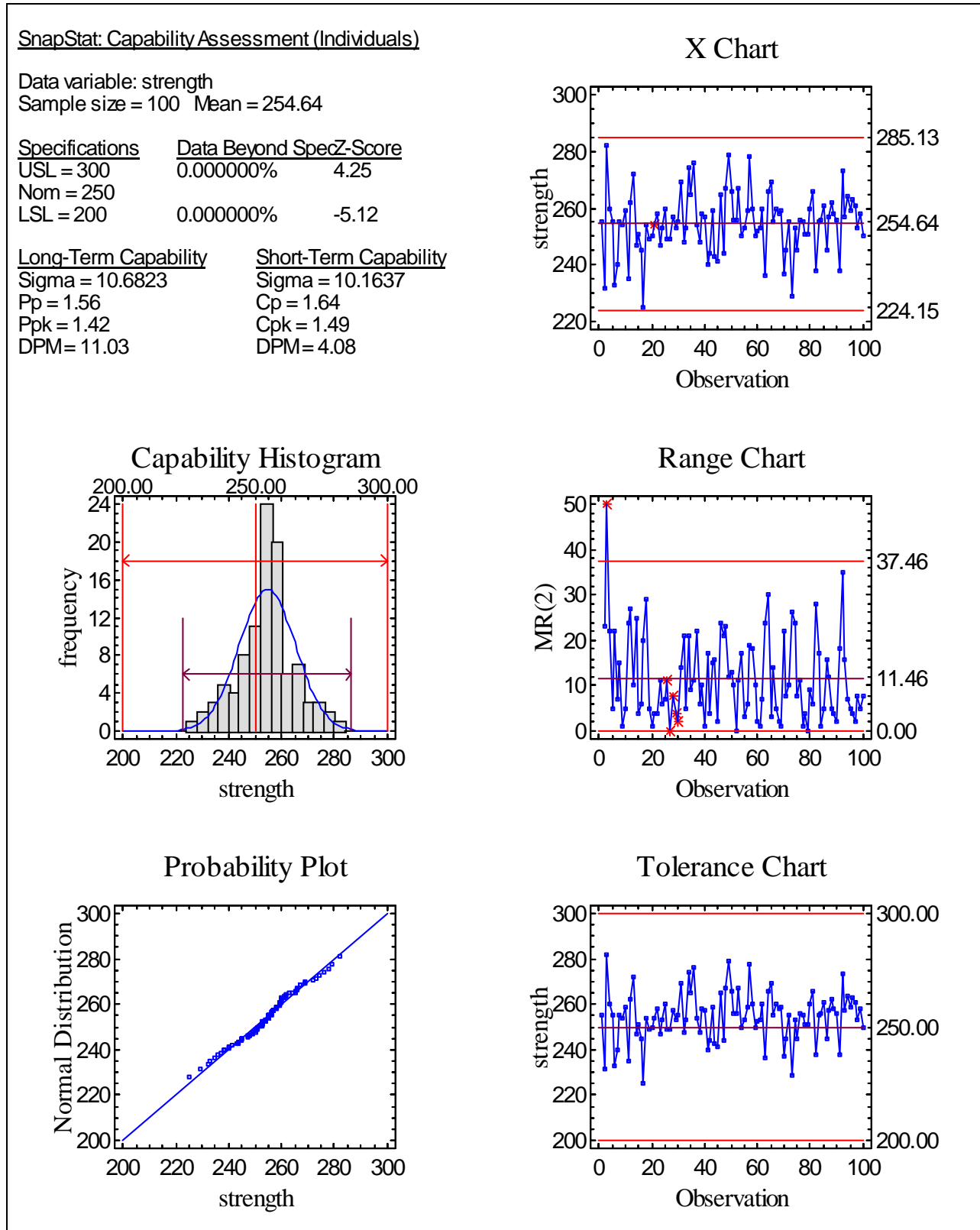
Case #2: Grouped Data

The data to be analyzed consist of one or more numeric columns. The data are assumed to have been taken in groups, in sequential order by rows.

- **Data:** one or more numeric columns. If more than one column is entered, each row of the file is assumed to represent a subgroup with subgroup size m equal to the number of columns entered. If only one column is entered, then the *Date/Time/Labels or Size* field is used to form the groups.
- **Date/Time./Labels or Size**: If each set of m rows represents a group, enter the single value m. For example, entering a 5 as in the example above implies that the data in rows 1-5 form the first group, rows 6-10 form the second group, and so on. If the subgroup sizes are not equal, enter the name of an additional numeric or non-numeric column containing group identifiers. The program will scan this column and place sequential rows with identical codes into the same group.
- **USL:** upper specification limit, if any.
- **Nominal**: optional nominal or target value for the variable. If not supplied, certain capability indices will not be calculated.
- **LSL**: lower specification limit, if any.
- **Select:** subset selection.

At least one of the specification limits must be entered.

Output

The output from the SnapStat consists of a single page of graphs and numerical statistics.

SnapStat: Capability Assessment (Individuals)

Data variable: strength
Sample size = 100   Mean = 254.64

Specifications	Data Beyond Spec	Z-Score
USL = 300	0.000000%	4.25
Nom = 250		
LSL = 200	0.000000%	-5.12

Long-Term Capability
Sigma = 10.6823
Pp = 1.56
Ppk = 1.42
DPM = 11.03

Short-Term Capability
Sigma = 10.1637
Cp = 1.64
Cpk = 1.49
DPM = 4.08

X Chart

Capability Histogram

Range Chart

Probability Plot

Tolerance Chart

Analysis Summary (top left)

The top left section of the output summarizes the data and compares it to the specification limits. Of primary interest are:

1. The sample size n and sample mean $\bar{x}$.
2. **Data Beyond Spec**. – the percentage of the input data that is beyond the specification limits.
3. **Z-score** – for a normal distribution, the distance from the specification limit to the sample mean, divided by the sample standard deviation. For non-normal distributions, an equivalent Z-score is displayed based on the percentage of the fitted distribution that is beyond the specification limit. Consequently, a Z-score of 3, which corresponds to a specification limit that is 3 standard deviations away from the mean in the case of a normal distribution, corresponds to the identical 0.14% beyond the specification limit regardless of what distribution is selected.
4. **Long-term Capability** – a summary of the performance of the process over the entire sampling period. Included are:

    Sigma – an estimate of the process standard deviation $\hat{\sigma}_L$ over the period covered by the data.

    Pp – a two-sided process performance index calculated by

    $$P_P = \frac{USL - LSL}{k\hat{\sigma}_L} \tag{1}$$

    where k is a multiple (usually 6) specified on the *Capability* tab of the *Preferences* dialog box, accessible from the *Edit* menu.

    Ppk – a one-sided process performance index calculated by

    $$P_{PK} = \min\left[\frac{\hat{\mu} - LSL}{(k/2)\hat{\sigma}_L}, \frac{USL - \hat{\mu}}{(k/2)\hat{\sigma}_L}\right] \tag{2}$$

    DPM – defects per million, determined from the area of the fitted distribution that lies outside the specification limits.

5. **Short-term Capability** – a summary of the capability of the process over a short period of time, available for a normal distribution only. Included are:

    Sigma – an estimate of the process standard deviation $\hat{\sigma}_S$ obtained from the within-subgroup variability if the data is arranged in groups or from the differences between consecutive observations if the data are individuals.

    Cp – a two-sided process capability index calculated by

    $$C_P = \frac{USL - LSL}{k\hat{\sigma}_S} \tag{3}$$

Cpk – a one-sided process capability index calculated by

$$C_{PK} = \min\left[\frac{\hat{\mu} - LSL}{(k/2)\hat{\sigma}_S}, \frac{USL - \hat{\mu}}{(k/2)\hat{\sigma}_S} \right] \qquad (4)$$

DPM – defects per million, determined from the area of the fitted distribution that lies outside the specification limits.

The key distinction between short-term and long-term indices is the method used to estimate σ, the standard deviation of the process. The long-term sigma is estimated from the sample standard deviation of the entire data set. This could potentially include variation caused by drifts in the process during the period in which the data was collected, so that a large estimate could be a sign of either large inherent variation or poor process control. The short-term estimate is obtained from either the moving range, the mean squared successive differences, or the subgroup standard deviations, depending on the settings on the *Capability* tab of the *Preferences* dialog box, accessible from the *Edit* menu. Since this estimate is taken from observations close together in time, it is much less influenced by lack of control over the process. However, it does not necessarily represent the current performance of the process.

X or X-bar Chart (top right)

The control chart in the top right section plots the individual data values or subgroup means, depending upon whether or not the data are grouped. Upper and lower control limits are added at a multiple of sigma determined by the *Control Charts* settings on the *Preferences* dialog box, accessible from the *Edit* menu. Points outside the limits or unusual runs could indicate an out-of-control process. In such cases, the capability indices must be interpreted with special care (if at all).

Capability Histogram (left center)

The left center section of the output displays a tabulation of the data. Included are:

- A frequency histogram in which the height of the bars represents the number of observations in adjacent, non-overlapping intervals of equal width. The number of classes in the histogram depends on the settings on the *EDA* tab of the *Preferences* dialog box, accessible from the *Edit* menu.

- A solid line showing the fitted distribution. The type of distribution is specified on the *Analysis Options* dialog box.

- Short vertical lines at specified percentiles of the fitted distribution. By default, the percentiles are positioned to cover 99.73% of the fitted distribution, which for a normal distribution covers a *6-sigma interval* centered at the sample mean.

- Taller vertical lines at the nominal value and the specification limits.

MR(2) or R Chart (right center)

The control chart in the right center section plots the moving range for individual data or subgroup ranges if the data are grouped. An upper control limit is included. Points beyond the upper limit or unusual runs could indicate an out-of-control process. In such cases, the capability indices must be interpreted with special care (if at all). The sample data show one unusually large value near the beginning as well as a run of smaller values below the centerline later on.

Probability Plot (bottom left)

The plot in the bottom left corner shows the quantiles of the data versus quantiles of the assumed distribution. If the data fall approximately along a straight line, than the currently selected distribution is probably adequate for the data. In the current example, the normal distribution appears to fit the data well. If necessary, a different distribution may be selected using *Analysis Options*.

Tolerance Chart (bottom left)

This chart plots the data values in row order together with the specification limits.

Analysis Options

- **Distribution** – the assumed distribution for the data.

- **Include** – the type of capability indices to be displayed. The choices are:

    Long-term and short-term – calculate both long-term and short-term indices.

Capability Assessment SnapStat - 7

Long-term only (labeled P) – calculate only long-term indices and label them with the letter P, as in Ppk.

Long-term only (labeled C) – calculate only long-term indices and label them with the letter C, as in Cpk.

Short-term only – calculate only short-term indices. These are always labeled with the letter C, as in Cpk.

- **Data Transformation** – the type of transformation applied to the data before fitting the selected distribution. Often, if a normal distribution does not fit the original data well, it will fit some transformation of the data. The choices are:

  None – fits the original data.

  Logarithm – fits the natural logarithms of the data.

  Power – fits the data after raising each value to the specified power.

  Box-Cox (optimized) – fits the data after raising each value to a power determined using the Box-Cox procedure.

  Note that even when a transformation is selected for analysis, most output is displayed in the original (not the transformed) metric.

- **Lower Threshold** – for distributions defined by a lower threshold, the value of that lower limit. This includes all distributions that indicate the number of parameters after their name.

- **Sigma Limits** – the sigma spread used to plot the limits on the *Capability Plot*. This value is usually set to 6.

- **Parameters** – Push this button to specify values for the distribution parameters. Normally, the parameters will be estimated from the data. However, this option allows you to fix the values of those parameters.

Calculations

For references on the calculations, see the documentation for *Capability Analysis (Variable Data)*.