Chapter 255

Capability Analysis

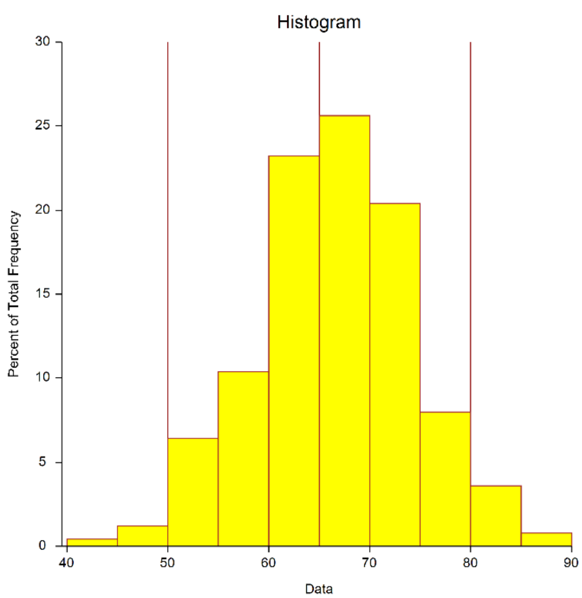
Introduction

This procedure summarizes the performance of a process based on user-specified specification limits. The observed performance as well as the performance relative to the Normal distribution are output. Process capability ratios such as C_p and C_{pk} are produced. C_{pm} and C_{pkm} may also be generated if a specification target is entered. A capability histogram with specification limit lines may also be produced in this procedure. Normality Tests are also given in this procedure. Subgroup data or individual values may be used.

Capability Analysis

Capability analysis, or process capability analysis, is the comparison of the distribution of sample values to the specification limits, and possibly also the specification target. One basic measure of the capability of the process is the proportion of values falling inside (or outside) the specification limits. Another measure of capability is the proportion of values that would fall inside (or outside) the specification limits if the data are assumed to follow the normal distribution. Several capability ratios, or capability indices, have been developed to summarize how well the process yields measurements within the specification limits. Those produced in this procedure are C_p, C_{pk}, C_{pm}, and C_{pkm}. C_{pm} and C_{pkm} additionally take into account the nearness of the process to the specification target.

Process data are typically gathered as samples or individual measurements taken from the process at given times (hours, shifts, days, weeks, months, etc.). If more than one value is taken at a time, the measurements of the samples at a given time constitute a subgroup.



Typically, an initial series of subgroups or individual values is used to estimate the mean and standard deviation of a process. The mean and standard deviation can then be used to estimate the capability of the process.

Because the assumption of normality is integral to the usefulness of the summaries, an important part of capability analysis is determining whether the data follow a Normal distribution, at least approximately. Normality tests and the capability histogram can be useful for investigating this assumption.

Other Procedures for Process Capability

Some of the other procedures in NCSS that may be useful for analyzing process capability are X-bar and R (or s) charts, IM-R Charts, Descriptive Statistics, Stem-and-Leaf Plots (for smaller samples), Normality Tests, Outlier Tests, Distribution Fitting, Box-Cox Transformation, and the Data Simulation Tool.

Process Capability Formulas

The formulas for estimating the mean and sigma depend on whether the data is subgroup data or individual value data.

Estimating the Mean – Subgroup Data

Suppose we have k subgroups, each of size n. Let x_{ij} represent the measurement in the j^{th} sample of the i^{th} subgroup.

The i^{th} subgroup mean is calculated using

$$\bar{x}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n},$$

The formula for the overall mean is

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}}{\sum_{i=1}^{k} n_i}.$$

If the subgroups are of equal size, the above equation for the grand mean reduces to

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_k}{k}.$$

Estimating the Mean – Individual Values Data

Suppose we have k individual values. The estimate of the overall mean is given by

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i}{k}.$$

Estimating Sigma – Subgroup Data

In this procedure, sigma can be entered directly, or there are three options for estimating sigma from subgroup data: sample ranges, sample standard deviations, and the weighted approach. Suppose we have *k* subgroups, each of size *n*. Let x_{ij} represent the measurement in the j^{th} sample of the i^{th} subgroup.

Estimating Sigma – Subgroup Data – Sample Ranges

If the standard deviation (sigma) is to be estimated from the ranges, R_i, it is estimated as

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

where

$$\bar{R} = \frac{\sum_{i=1}^{k} R_i}{k}$$

$$d_2 = \frac{E(R)}{\sigma} = \frac{\mu_R}{\sigma}$$

The calculation of *E(R)* requires the knowledge of the underlying distribution of the x_{ij}'s. Making the assumption that the x_{ij}'s follow the normal distribution with constant mean and variance, the values for d_2 are derived through the use of numerical integration. It is important to note that the normality assumption is used and that the accuracy of this estimate requires that this assumption be valid.

In the procedure, this calculation is performed if Sigma Estimation is set to From Data – R-bar or s-bar Estimate, and Range or SD Estimation is set to Range.

Estimating Sigma – Subgroup Data – Sample Standard Deviations

If the standard deviation (sigma) is to be estimated from the standard deviations, it is estimated as

$$\hat{\sigma} = \frac{\bar{s}}{c_4}$$

where

$$\bar{s} = \frac{\sum_{i=1}^{k} s_i}{k}$$

$$c_4 = \frac{E(s)}{\sigma} = \frac{\mu_s}{\sigma}$$

The calculation of $E(s)$ requires the knowledge of the underlying distribution of the x_{ij}'s. Making the assumption that the x_{ij}'s follow the normal distribution with constant mean and variance, the values for c_4 are obtained from

$$c_4 = \sqrt{\frac{2}{n-1}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

In the procedure, this calculation is performed if Sigma Estimation is set to From Data – R-bar or s-bar Estimate, and Range or SD Estimation is set to SD.

Estimating Sigma – Subgroup Data – Weighted (SD) Approach

When the sample size is variable across subgroups, a weighted approach is recommended for estimating sigma (Montgomery, 2013):

$$\hat{\sigma} = \left[\frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{\sum_{i=1}^{k} n_i - k}\right]^{1/2}$$

In the procedure, this calculation is performed if Sigma Estimation is set to From Data – SD Approach.

Estimating Sigma – Individual Values Data

Suppose we have k individual values. There are two methods in this procedure for estimating sigma: moving ranges and overall sample standard deviation.

Estimating Sigma – Individual Values Data – Moving Ranges

If there is only one observation per time point, a moving range may be calculated by finding the range of each value with its previous value:

$$R_i = |x_i - x_{i-1}|$$

Then the standard deviation (sigma) is estimated from the ranges, R_i, in the same manner as for subgroup data, namely,

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

where

$$\bar{R} = \frac{\sum_{i=1}^{k} R_i}{k}$$

$$d_2 = \frac{E(R)}{\sigma} = \frac{\mu_R}{\sigma}$$

In the procedure, this calculation is performed if the data are individual values data, and Sigma Estimation is set to From Data – R-bar or s-bar Estimate, and Range or SD Estimation is set to Range.

Estimating Sigma – Individual Values Data – Overall Standard Deviation

If there is only one observation per time point, and the process is assumed to be in control, sigma may be estimated using the sample standard deviation

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2}{k-1}}$$

In the procedure, this calculation is performed if the data are individual values data, and if Sigma Estimation is set to From Data – SD Approach.

Process Capability Ratios

Several capability ratio formulas are presented below. Further details may be found in Montgomery (2013) and Ryan (2011).

C_p

The process capability ratio C_p is given by

$$C_p = \frac{USL - LSL}{6\sigma}$$

where USL and LSL are the upper and lower specification limits, respectively. An estimate of C_p is produced by substituting a suitable estimate of σ, namely $\hat{\sigma}$.

Confidence intervals for Cp are given as

$$C_{p_{lower}} = C_p\sqrt{\frac{\chi^2_{n-1,\alpha/2}}{n-1}}$$

$$C_{p_{upper}} = C_p\sqrt{\frac{\chi^2_{n-1,\,1-\alpha/2}}{n-1}}$$

C_{pl} and C_{pu}

The one-sided capability ratios C_{pl} and C_{pu} are defined as

$$C_{pl} = \frac{\mu - LSL}{3\sigma}$$

and

$$C_{pu} = \frac{ULS - \mu}{3\sigma}$$

These are estimated by substituting mean and standard deviation estimates.

C_{pk}

C_{pk} is the lesser of C_{pl} and C_{pu}, or

$$C_{pk} = \min(C_{pl}, C_{pu})$$

The lower and upper confidence limits for C_{pk} reported in **NCSS** are given by

$$C_{pk_{lower}} = C_{pk} - z_{1-\alpha/2}\sqrt{\frac{n-1}{9n(n-3)} + \left(\frac{{C_{pk}}^2}{2n-6}\right)\left(1 + \frac{6}{n-1}\right)}$$

$$C_{pk_{upper}} = C_{pk} + z_{1-\alpha/2}\sqrt{\frac{n-1}{9n(n-3)} + \left(\frac{{C_{pk}}^2}{2n-6}\right)\left(1 + \frac{6}{n-1}\right)}$$

C_{pm}

A capability ratio which incorporates the nearness to the specification target is defined as

$$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}}$$

where T refers to the specification target. A suitable estimate of C_{pm} is made by substituting estimates of the mean and standard deviation.

C_{pmk}

Similarly, to C_{pm}, C_{pmk} also accounts for nearness to the specification target:

$$C_{pmk} = \frac{C_{pk}}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}$$

where T refers to the specification target. A suitable estimate of C_{pmk} is made by substituting estimates of the mean and standard deviation.

Data Structure

In this procedure, the data may be in any of three formats. The first data structure option is to have the data in several columns, with one subgroup per row.

Example Dataset

S1	S2	S3	S4	S5
2	6	3	8	5
8	8	7	7	9
6	2	2	4	3
5	6	7	6	10
48	2	6	5	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

The second data structure option uses one column for the response data, and either a subgroup size or a second column defining the subgroups.

Alternative Example Dataset

Response	Subgroup
2	1
6	1
3	1
8	1
5	1
8	2
8	2
7	2
7	2
9	2
6	3
2	3
.	.
.	.
.	.

In the alternative example dataset, the Subgroup column is not needed if every subgroup is of size 5 and the user specifies 5 as the subgroup size. If there are missing values, the Subgroup column should be used, or the structure of the first example dataset.

If there are no subgroups (individual values only), the only input needed is a single column of values.

Response
2
6
3
8
5
8
8
7
7
9
6
2
.
.
.

Example 1 – Capability Analysis for Subgroup Data

This section presents an example of how to run a capability analysis for subgroups. The data represent 50 subgroups of size 5 that are assumed to be in control. The specification limits for the process are 50 and 80, with a specification target of 65. The data used are in the QC dataset. We will analyze the variables D1 through D5 of this dataset.

Setup

To run this example, complete the following steps:

1 Open the QC example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **QC** and click **OK**.

2 Specify the Capability Analysis procedure options

- Find and open the **Capability Analysis** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Data Variables **D1-D5**

Limits & Estimation Tab

Lower Limit **50**
Upper Limit **80**
Target Specification Value.............................. **65**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Mean Estimation Section

Mean Estimation Section for Subgroups 1 to 50

Number of Subgroups: 50

Estimation Type	Estimate
Estimated Grand Average	67.12

This section displays the estimated mean to be used in all calculations.

Estimated Grand Average

This value is the average of all the observations. If all the subgroups are of the same size, it is also the average of all the X-bars.

Sigma Estimation Section

Sigma Estimation Section for Subgroups 1 to 50

Estimation Type	Estimated Value	Estimated Sigma
Ranges (R-bar)*	18.14	7.798796
Standard Deviations (s-bar)	7.365443	7.835698
Weighted Approach (s-bar)	7.902911	7.902911

* Indicates the estimation type used in this report.

This report gives the estimation of the population standard deviation (sigma) based on three estimation techniques. The estimation technique used for the calculations in this procedure is based on the ranges.

Estimation Type

Each of the formulas for estimating sigma is shown earlier in this chapter in the Process Capability Formulas section.

Estimated Value

This column gives the R-bar and s-bar estimates based on the corresponding formulas.

Estimated Sigma

This column gives estimates of the population standard deviation (sigma) based on the corresponding estimation type.

Capability Analysis Section

Capability Analysis Section

Data Summary	
Number of Values	250
Sigma (Estimated)	7.798796
Mean (Estimated)	67.12

Number of Values

This is the number of observations in the capability analysis. While subgroups were used in the estimation of sigma, they are no longer distinguished in the remainder of the capability analysis.

Sigma

This is the sigma that will be used for the capability analysis calculations.

Mean

This is the mean to be used for the capability analysis calculations.

Specification Summary

Specification Summary

Specification	Value	Corresponding Z-Value
Lower Limit	50	-2.195211
Upper Limit	80	1.651537
Target Value	65	-0.271837

This report lists the user-specified specification values, as well as the corresponding Z-value.

Specification

This column identifies the specification value type.

Value

This is the user-input specification value. The target value is only required for the C_{pm} and C_{pkm} capability ratios.

Corresponding Z-Value

These are the z-values of the specification limits and target value, calculated using the formula

$$z_{spec} = \frac{spec - \hat{\mu}}{\hat{\sigma}}$$

Performance Summary

Performance Summary (Lower Limit)

		Observed Performance			Normal Distribution Performance	
Specification	LL Value	# < LL	% < LL	PPM < LL	% < LL	PPM < LL
Lower Limit (LL)	50	4/250	1.6000%	16000.00	1.4074%	14074.25

Performance Summary (Upper Limit)

		Observed Performance			Normal Distribution Performance	
Specification	UL Value	# > UL	% > UL	PPM > UL	% > UL	PPM > UL
Upper Limit (UL)	80	11/250	4.4000%	44000.00	4.9314%	49314.49

Performance Summary (Outside Both Limits)

	Observed Performance			Normal Distribution Performance	
Specification	**# Outside**	**% Outside**	**PPM Outside**	**% Outside**	**PPM Outside**
Outside Both Limits	15/250	6.0000%	60000.00	6.3389%	63388.74

Performance Summary (Between Limits)

		Observed Performance			Normal Distribution Performance	
Specification	**Range (UL - LL)**	**LL ≤ # ≤ UL**	**LL ≤ % ≤ UL**	**LL ≤ PPM ≤ UL**	**LL ≤ % ≤ UL**	**LL ≤ PPM ≤ UL**
Between Limits	30	235/250	94.0000%	940000.00	93.6611%	936611.26

This report gives the percentage of values inside or outside the specification limits. In this example, the observed performance is similar to the Normal distribution (expected) performance.

Specification

This identifies the region to be examined for performance.

LL, UL, and Range

The LL and the UL values are the user-specified lower and upper specification limits. The range is the lower limit subtracted from the upper limit.

Observed Performance –

This gives the actual number of observed values in the corresponding region.

Observed Performance – %

This gives the percent of observed values in the corresponding region.

Observed Performance – PPM

This gives the parts per million number of observed values in the corresponding region.

Normal Distribution Performance – %

If the values are assumed to follow a normal distribution with mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$, this is the percent of values that would fall in the corresponding region. This is sometimes called the expected performance.

Normal Distribution Performance – PPM

If the values are assumed to follow a normal distribution with mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$, this is the parts per million number of values that would fall in the corresponding region. This is sometimes called the expected performance.

Process Capability Ratios

Process Capability Ratios

Confidence Level: 95.00%

Capability Ratio	Value	Confidence Interval Lower Limit	Confidence Interval Upper Limit
Cp (with C.I.)	0.641125	0.584820	0.697364
Cpk (with C.I.)	0.550512	0.486211	0.614813
Cpl	0.731737		
Cpu	0.550512		
Cpm	0.618673		
Cpkm	0.531234		

This report gives the values of the various capability ratios. Confidence intervals are given for C_p and C_{pk}. We refer the reader to Montgomery (2013) or Ryan (2011) for interpretation details of each ratio.

Capability Ratio

This identifies the capability ratio of each line.

Value and Confidence Interval Limits

The formulas for each of these values are given earlier in this chapter in the Process Capability Ratios section under Process Capability Formulas.

3- to 6-Sigma Limits

3- to 6-Sigma Limits

Limit Type	Mean	Lower Limit	Upper Limit
3-Sigma Limits	67.12	43.72361	90.51639
4-Sigma Limits	67.12	35.92482	98.31519
5-Sigma Limits	67.12	28.12602	106.114
6-Sigma Limits	67.12	20.32722	113.9128

Limits

The formulas for the limits are

$$LL = \hat{\mu} - m\hat{\sigma}$$

$$UL = \hat{\mu} + m\hat{\sigma}$$

where *m* is the multiplier 3, 4, 5, or 6.

Normality Tests Section

Normality Tests Section

User-Specified Alpha Level: 0.05

Test Results

Normality Test	Test Statistic	Prob Level	Conclusion
Shapiro-Wilk	0.995	0.668324	Do Not Reject Normality Assumption
Anderson-Darling	0.455	0.268189	Do Not Reject Normality Assumption
Chi-Square	1.637	0.441089	Do Not Reject Normality Assumption

The details of the Shapiro-Wilk and Anderson-Darling (and other) Normality tests are discussed in the Normality Tests procedure. The Chi-Square goodness of fit test for normality is obtained by dividing the data into bins, and then comparing the observed counts to the expected counts for each bin using

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

The individual observed and expected counts are detailed in the Chi-Square Test Frequency Distribution Details section.

Chi-Square Test Frequency Distribution Details

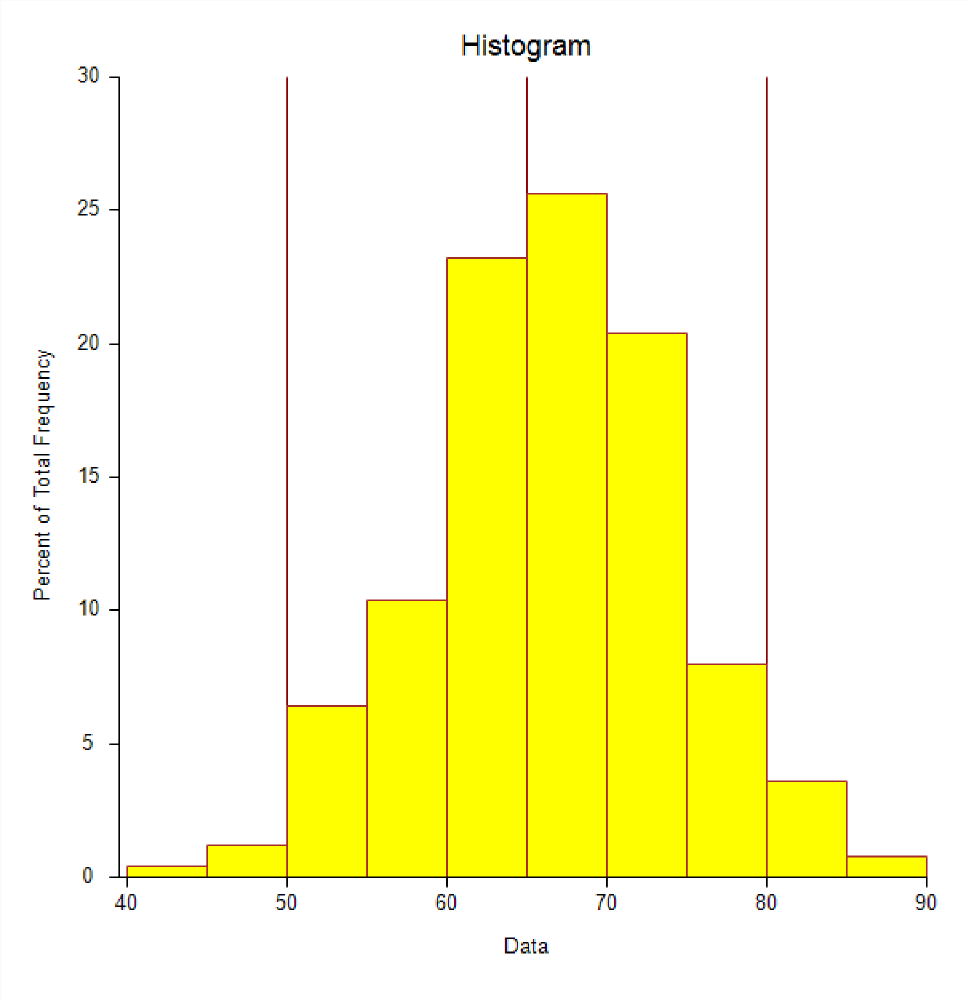
Chi-Square Test Frequency Distribution Details

Bin Boundaries								
Lower Boundary	Upper Boundary	Actual Count	Normal Count	Diff. Count	Actual Percent	Normal Percent	Diff. Percent	Chi-Sqr Amount
	47.62301	1.0	1.6	-0.6	0.4	0.6	-0.2	0.00
47.62301	55.42181	19.0	15.1	3.9	7.6	6.1	1.5	0.65
55.42181	63.2206	59.0	60.4	-1.4	23.6	24.2	-0.6	0.03
63.2206	71.0194	101.0	95.7	5.3	40.4	38.3	2.1	0.29
71.0194	78.81819	55.0	60.4	-5.4	22.0	24.2	-2.2	0.49
78.81819	86.61699	14.0	15.1	-1.1	5.6	6.1	-0.5	0.17
86.61699		1.0	1.6	-0.6	0.4	0.6	-0.2	0.00
Total		250.0	250.0	0.0	100.0	100.0	0.0	1.64

This section summarizes the contribution of each bin to the Chi-Square goodness of fit test statistic.

Histogram Section

Histogram Section for Subgroups 1 to 50



This histogram also displays the (vertical line) specification limits as well as the specification target.

Example 2 – Capability Analysis for Individual Value Data

The capability analysis of individual value data is nearly the same as the analysis of subgroup data. The only difference is the way in which sigma is estimated. This section presents an example of how to run a capability analysis for individual value data. The data represent 200 part widths of a process that is assumed to be in control. The specification limits for the process are 300 and 400, with a specification target of 350. The data used are in the Capability dataset. We will analyze the variable Width of this dataset.

Setup

To run this example, complete the following steps:

1 Open the Capability example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Capability** and click **OK**.

2 Specify the Capability Analysis procedure options

- Find and open the **Capability Analysis** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Input Type... **Response Column with Individual Values (no subgroups)**
Response Variable .... **Width**

Limits & Estimation Tab

Lower Limit **300**
Upper Limit **400**
Target Specification Value.............................. **350**
Sigma Estimation.. **From Data - SD Approach**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Output

Mean Estimation Section for All Individuals

Number of Individuals: 200

Estimation Type	Estimate
Estimated Grand Average	346.79

Sigma Estimation Section for All Individuals

Estimation Type	Estimated Value	Estimated Sigma
Ranges (R-bar)	28.84422	25.57112
Standard Deviation*	25.37945	25.37945

* Indicates the estimation type used in this report.

Capability Analysis Section

Data Summary

Number of Values	200
Sigma (Estimated)	25.37945
Mean (Estimated)	346.79

Specification Summary

Specification	Value	Corresponding Z-Value
Lower Limit	300	-1.843618
Upper Limit	400	2.096578
Target Value	350	0.126480

Performance Summary (Lower Limit)

		Observed Performance			Normal Distribution Performance	
Specification	LL Value	# < LL	% < LL	PPM < LL	% < LL	PPM < LL
Lower Limit (LL)	300	4/200	2.0000%	20000.00	3.2619%	32619.45

Performance Summary (Upper Limit)

		Observed Performance			Normal Distribution Performance	
Specification	UL Value	# > UL	% > UL	PPM > UL	% > UL	PPM > UL
Upper Limit (UL)	400	6/200	3.0000%	30000.00	1.8015%	18015.47

Performance Summary (Outside Both Limits)

	Observed Performance			Normal Distribution Performance	
Specification	**# Outside**	**% Outside**	**PPM Outside**	**% Outside**	**PPM Outside**
Outside Both Limits	10/200	5.0000%	50000.00	5.0635%	50634.91

Performance Summary (Between Limits)

		Observed Performance			Normal Distribution Performance	
Specification	**Range (UL - LL)**	**LL ≤ # ≤ UL**	**LL ≤ % ≤ UL**	**LL ≤ PPM ≤ UL**	**LL ≤ % ≤ UL**	**LL ≤ PPM ≤ UL**
Between Limits	100	190/200	95.0000%	950000.00	94.9365%	949365.09

Process Capability Ratios

Confidence Level: 95.00%

		Confidence Interval	
Capability Ratio	**Value**	**Lower Limit**	**Upper Limit**
Cp (with C.I.)	0.656699	0.592195	0.721123
Cpk (with C.I.)	0.614539	0.537410	0.691669
Cpl	0.614539		
Cpu	0.698859		
Cpm	0.651509		
Cpkm	0.609682		

3- to 6-Sigma Limits

Limit Type	Mean	Lower Limit	Upper Limit
3-Sigma Limits	346.79	270.6516	422.9283
4-Sigma Limits	346.79	245.2722	448.3078
5-Sigma Limits	346.79	219.8927	473.6873
6-Sigma Limits	346.79	194.5133	499.0667

Normality Tests Section

User-Specified Alpha Level: 0.05

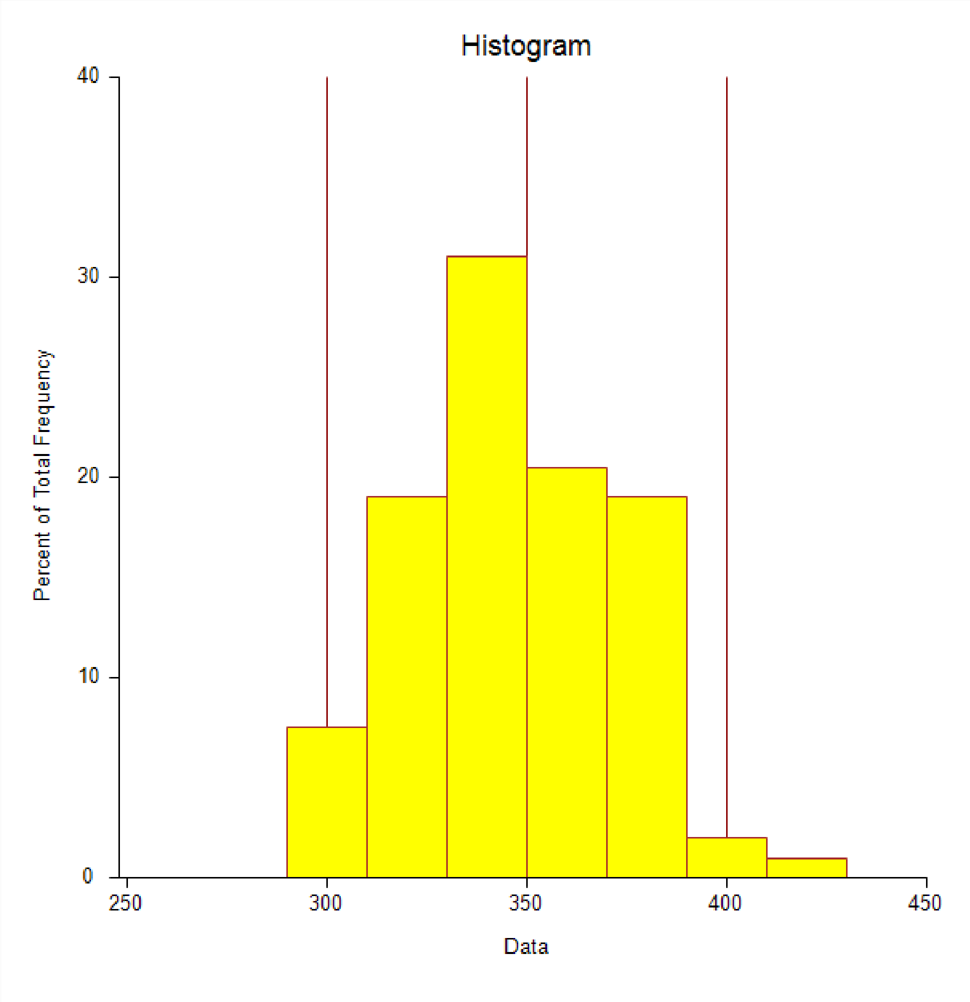
Test Results

Normality Test	Test Statistic	Prob Level	Conclusion
Shapiro-Wilk	0.986	0.039277	Reject Normality Assumption
Anderson-Darling	0.594	0.120995	Do Not Reject Normality Assumption
Chi-Square	2.714	0.257411	Do Not Reject Normality Assumption

Chi-Square Test Frequency Distribution Details

Bin Boundaries								
Lower Boundary	Upper Boundary	Actual Count	Normal Count	Diff. Count	Actual Percent	Normal Percent	Diff. Percent	Chi-Sqr Amount
	283.3414	0.0	1.2	-1.2	0.0	0.6	-0.6	0.00
283.3414	308.7208	12.0	12.1	-0.1	6.0	6.1	-0.1	0.14
308.7208	334.1003	52.0	48.3	3.7	26.0	24.2	1.8	0.28
334.1003	359.4797	77.0	76.6	0.4	38.5	38.3	0.2	0.00
359.4797	384.8592	51.0	48.3	2.7	25.5	24.2	1.3	0.15
384.8592	410.2386	6.0	12.1	-6.1	3.0	6.1	-3.1	2.15
410.2386		2.0	1.2	0.8	1.0	0.6	0.4	0.00
Total		200.0	200.0	0.0	100.0	100.0	0.0	2.71

Histogram Section for All Individuals



The output descriptions for each section of the output are presented in Example 1. The only difference in formulas in Example 2 compared to Example 1 is the difference in the calculation of the sigma estimate. In Example 2, the common sample standard deviation formula using all the individual values is used to calculate the sigma estimate.