Capability Analysis with Non-Normal Data
Not All Data is Normal

- Some data is not normally distributed.

- Non-normal data leads to incorrect capability indices since the distribution is not symmetrical about the mean.

- Furthermore, the estimates for the mean and standard deviation will have a bias.

- There are several alternatives to calculate capability indices for non-normal data.

- Depending on the transformation, Cpk might not be calculated.
Box-Cox Transformation

• The Box-Cox transformation is a family of power functions that transform the data based on the following function.

\[ x(\lambda) = \frac{x^\lambda - 1}{\lambda} \quad \lambda \neq 0 \]

\[ x(\lambda) = \ln(\lambda) \quad \lambda = 0 \]

• The selection of \( \lambda \) is such that it minimizes the standard deviation of the resultant transformation.
Another Variation

- Another variation on the Box-Cox transformation is a more simple power function.

\[ y(\lambda) = y^\lambda \quad \lambda \neq 0 \]

\[ x(\lambda) = \ln(\lambda) \quad \lambda = 0 \]

- In a similar fashion, the selection of \( \lambda \) is such that it minimizes the standard deviation of the resultant transformation.

- Some common values for \( \lambda \) are 0.5 (square root), 2 (squared) and 0 which is the \( \log_e \).
Example of Box-Cox ($\lambda=0$)
Johnson Transformation

• The Johnson transformation optimally selects one of the three families of distribution; bounded (B), lognormal (L) or unbounded (U).

• The algorithm for the Johnson transformation is:

1. Considers almost all potential transformation functions from the Johnson system.
2. Estimate the parameters in the function.
3. Transforms the data with the transformation function.
4. Calculates Anderson-Darling statistics and the corresponding p-value for the transformed data.
5. Selects the transformation function with the largest p-value that is greater than the p-value criterion (default is 0.10). Otherwise, no transformation is appropriate.
Capability for Non-normal Data

- $P_p$ is calculated using the parameter values from the distribution of the data.

\[
P_p = \frac{USL - LSL}{X_{0.99865} - X_{0.00135}}
\]
Capability for Non-normal Data

- PpL is calculated using the parameter values from the distribution of the data.

\[ PpL = \frac{X_{0.5} - LSL}{X_{0.5} - X_{0.00135}} \]
Capability for Non-normal Data

- $PpU$ is calculated using the parameter values from the distribution of the data.

\[
PpU = \frac{USL - X_{0.5}}{X_{0.99865} - X_{0.5}}
\]
Alternative Approach

• An alternative approach to transforming the data is to find a different distribution that fits the data well and use it.

• This approach is more difficult unless you know from experience how the data is distributed.

• Many distribution (i.e. lognormal) are a member of the family of transformations presented early.

• Any transformation should include transformation of the specifications prior to capability analysis.
Example

Process Capability of Original

Transformed (λ = 0)

Mean  1.46  -0.071
SD (within)  1.72  0.976
LSL  0.25  -1.386
USL  6.0  1.792
Cp  0.56  0.54
Cpk  0.23  0.45

© Statistical Outsourcing Services 2009
Effect on Capability

- Non normal distributions give capability estimates that are based on the percentiles and not the standard ±3S limits.

- Mis-specifying the distribution of the data gives misleading estimates of the product expected to be out of specification.

- Utilizing normal statistics on non-normal data can hamper a continuous improvement effort by giving false signals on the process behavior.
Conclusions

- Always check your assumption of normality.

- There are techniques such as data transformations that can help to make non-normal data normally distributed.

- Any transformation to the data also needs to be performed on the specifications.