Process Capability: History, Assumptions, and Challenges

Daniel Griffith, Minitab Inc.
Eduardo Santiago, Minitab Inc.

Session M18
ASQ World Conference

The Global Voice of Quality™

Learning Objectives

1. Evaluate the assumptions for Normal Capability Analysis
2. Run a Non-Normal (NN) Capability Analysis using different techniques
3. Determine what makes data Non-Normal
4. Assess the weaknesses of Normality tests
5. Learn how to deal with data coming from a measurement system with low discrimination

History of Capability Analysis

- Popularized by J. Juran in his Quality Control Handbook [1].

$$Process\ Capability = 6\sigma$$

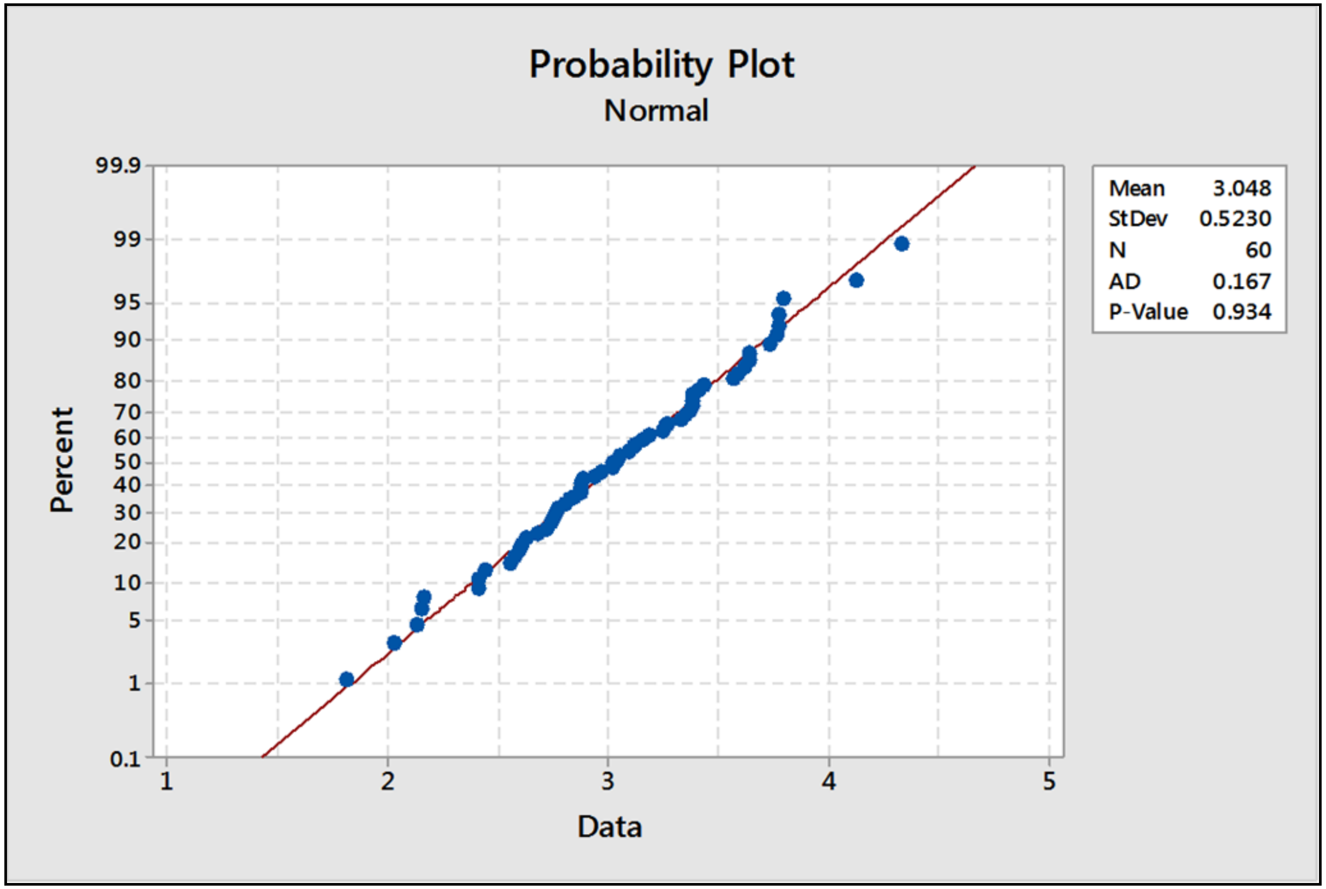
- Capability ratio is the tolerance width divided by the process capability, Cp.

History of Capability Analysis

- Late 1980’s, more capability indices were formally introduced.
- Kane [2] introduced Cpk.
- Capability indices can be translated to a defect or quality rate.

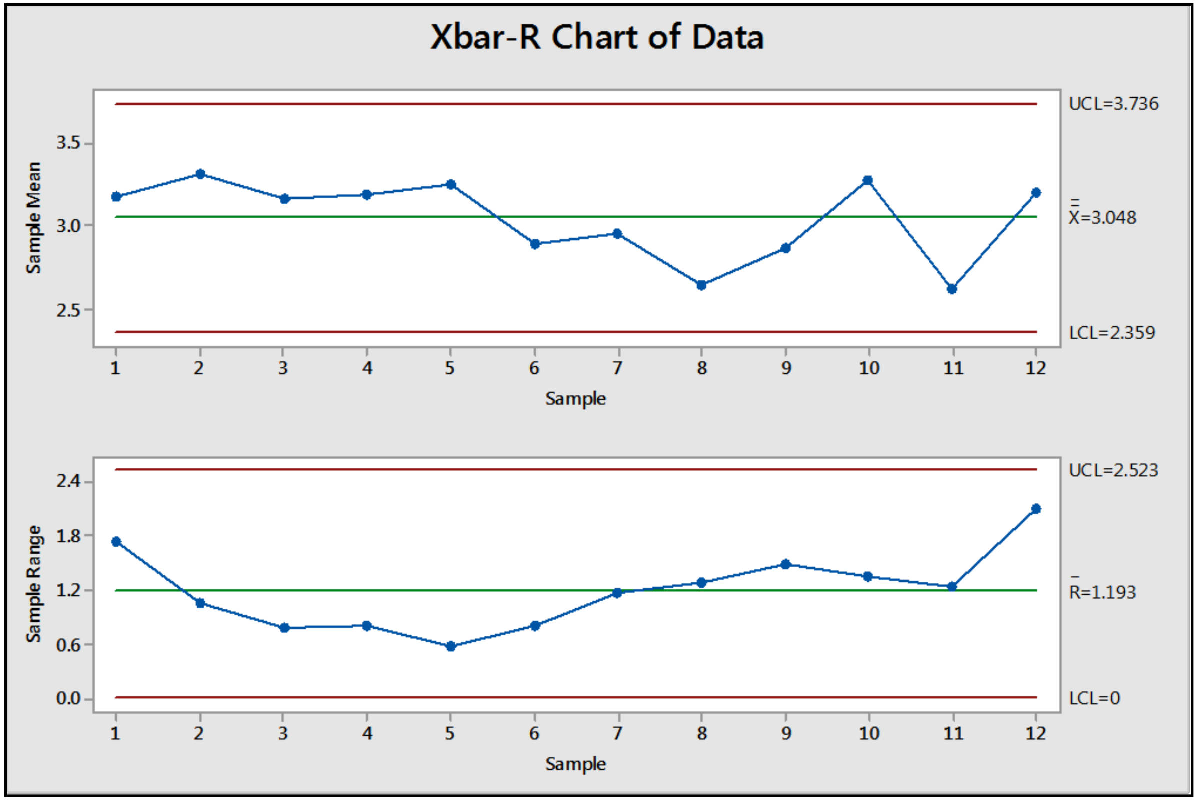
Normal Capability Analysis

- Is the data normally distributed?



Normal Capability Analysis

- Is the data normally distributed?
- Is the data stable over time?



Normal Capability Analysis

- Is the data normally distributed?
- Is the data stable over time?
- Are we confident in our conclusions?

Potential (Within) Capability	
Cp	1.04
CI for Cp	(0.85, 1.22)
CPL	1.07
CPU	1.00
Cpk	1.00
CI for Cpk	(0.80, 1.20)

Normal Capability Analysis

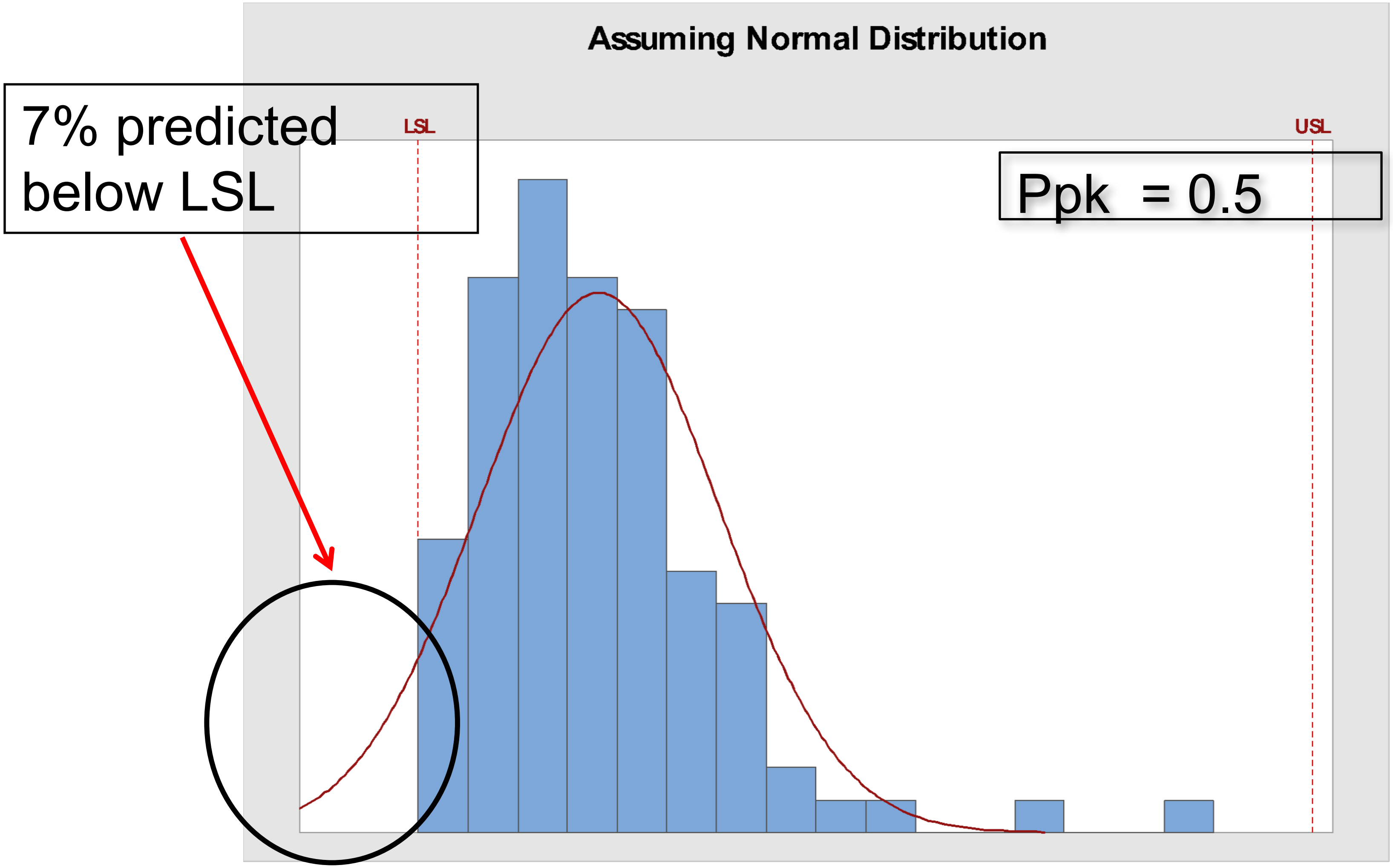
- Is the data normally distributed?
- Is the data stable over time?
- Are we confident in our conclusions?
- How can we improve our process?

Overall Capability	
Pp	1.13
PPL	0.66
PPU	1.60
Ppk	0.66

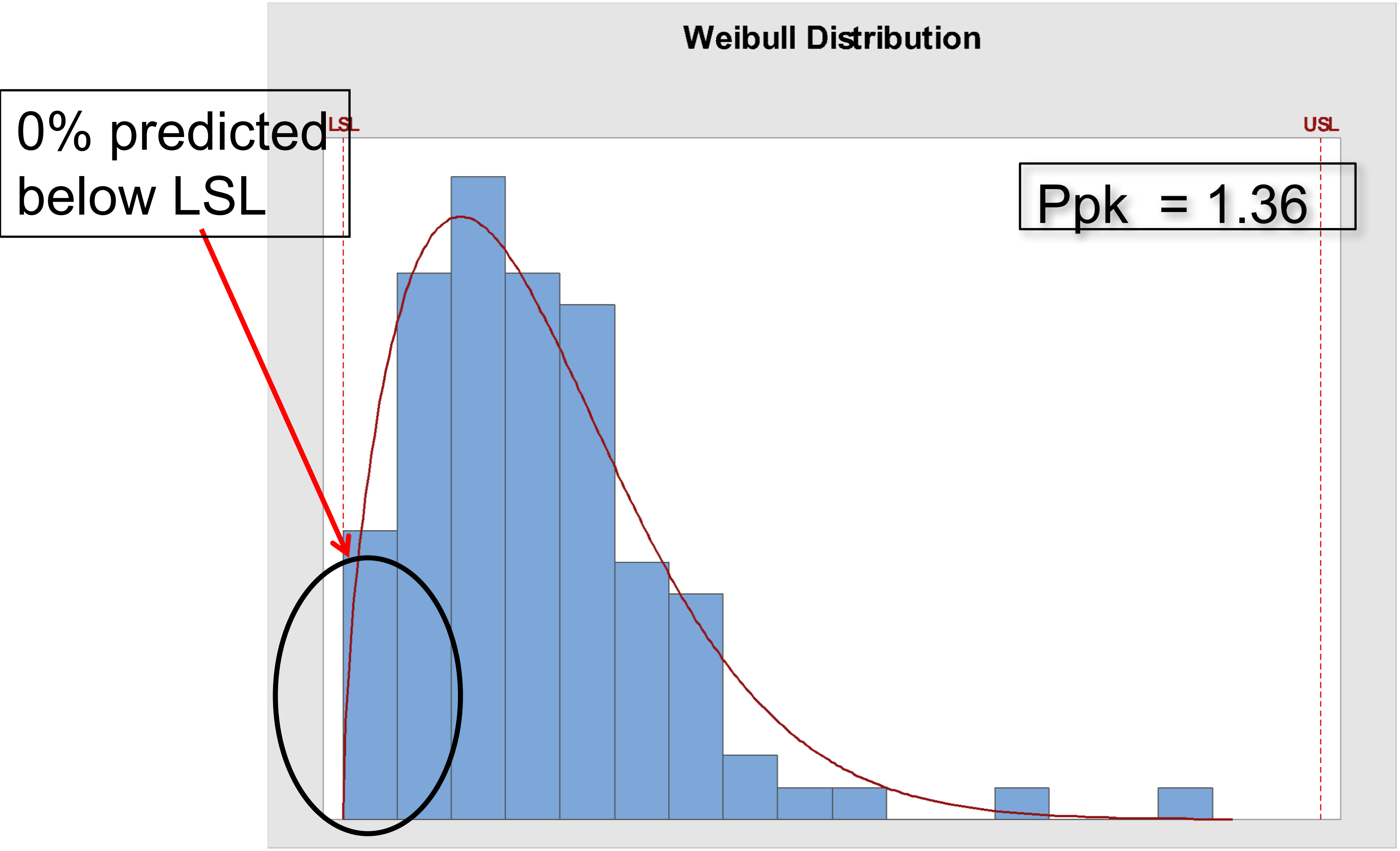
Is the Data Normally Distributed?

- Why does it matter?
- Let’s test how robust Capability Analysis is to Normality
- We will generate data from a Weibull distribution, and see how well Normality works in this case.

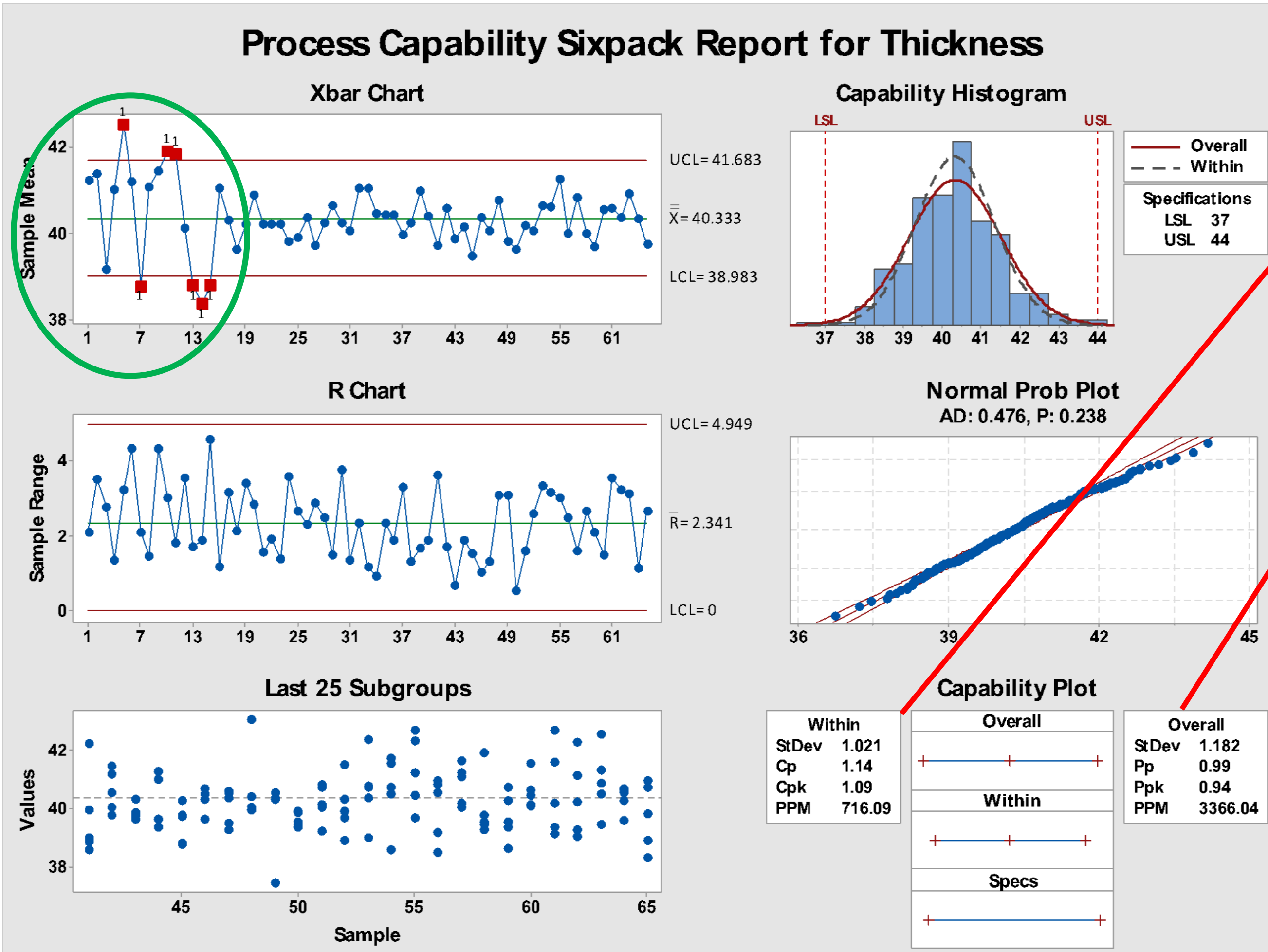
Is the Data Normally Distributed?



Is the Data Normally Distributed?



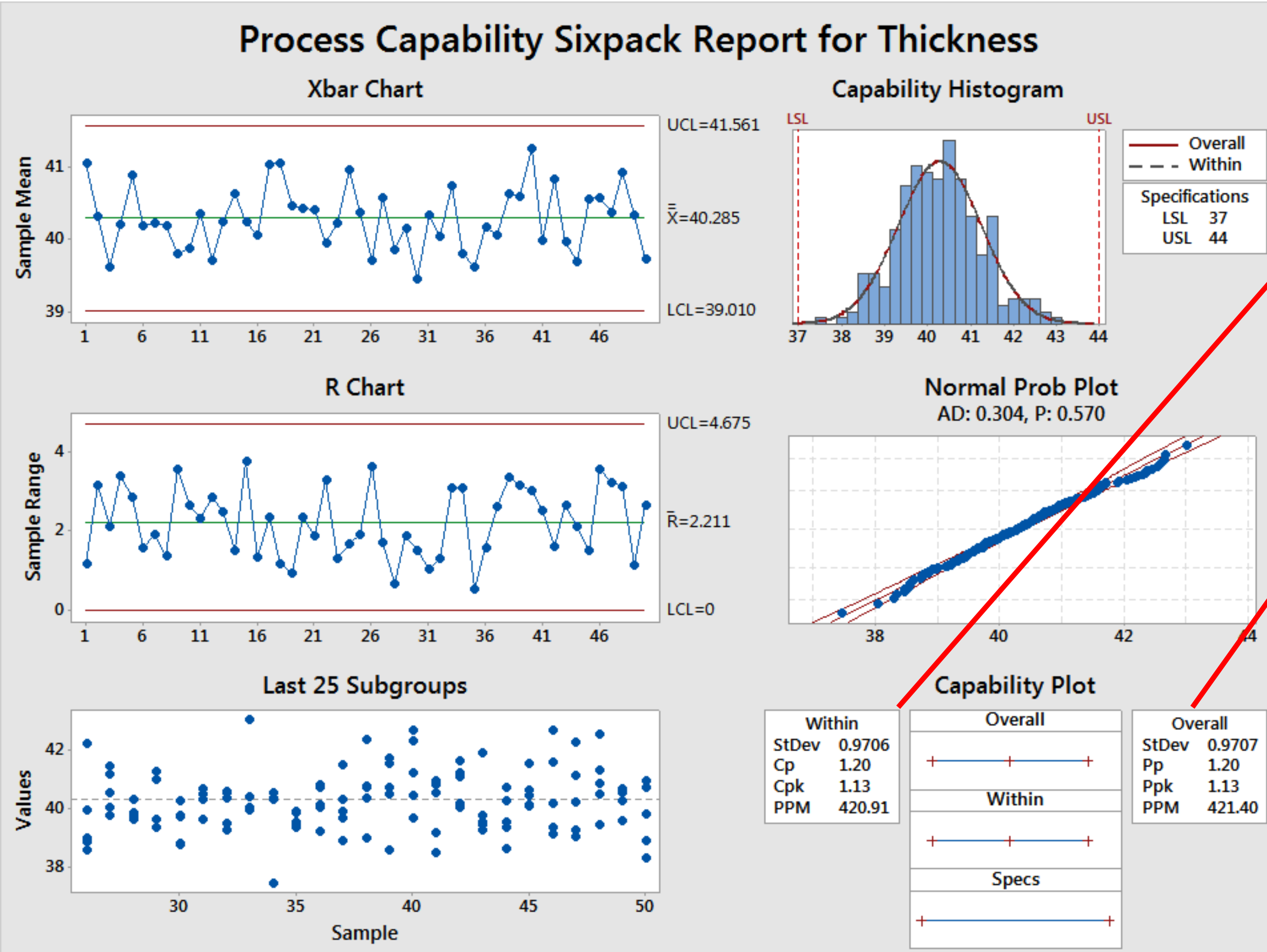
Is the Data Stable Over Time?



Within	
StDev	1.021
Cp	1.14
Cpk	1.09
PPM	716.09

Overall	
StDev	1.182
Pp	0.99
Ppk	0.94
PPM	3366.04

Is the Data Stable Over Time?



Within	
StDev	0.9706
Cp	1.20
Cpk	1.13
PPM	420.91

Overall	
StDev	0.9707
Pp	1.20
Ppk	1.13
PPM	421.40

Are we confident in our conclusions?

Would I rather see this?

Cpk	1.29

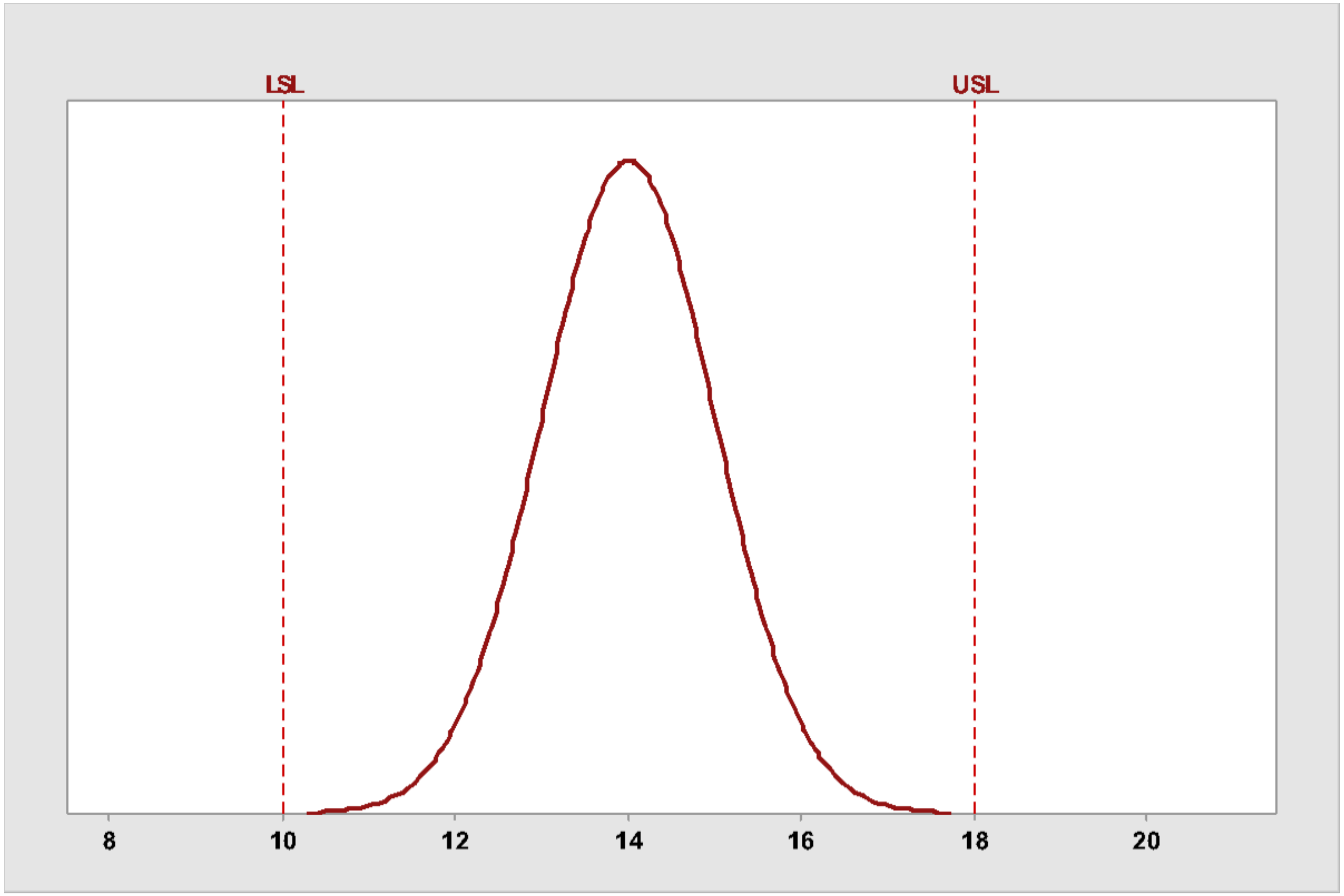
Or this?

CI for Cpk	(1.10, 1.48)

Both!

Confidence Intervals provide a better understanding of how much trust we can place in our results.

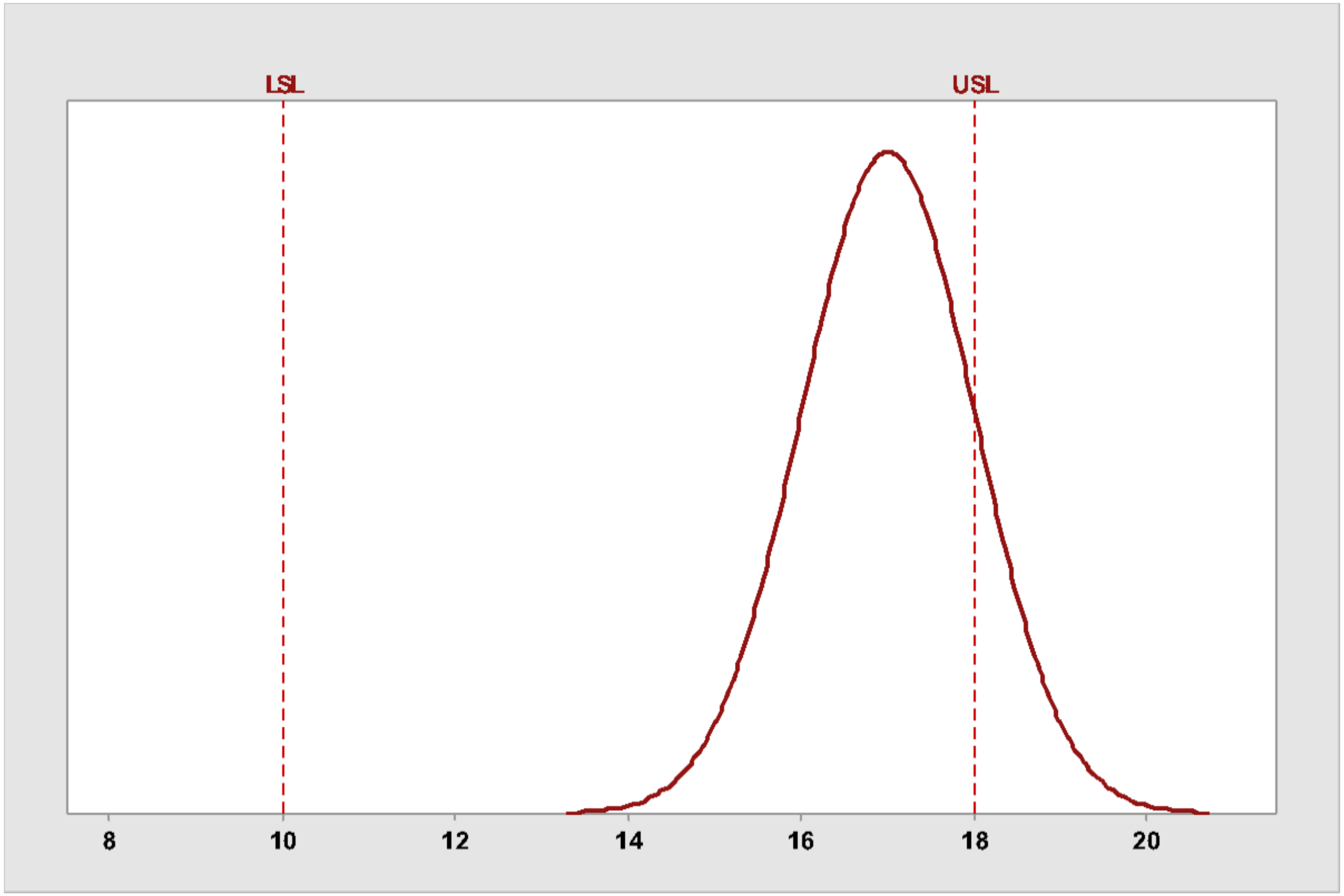
How can we Improve our Process?



Mean = 14
StDev = 1

Pp = 1.33
Ppk = 1.33

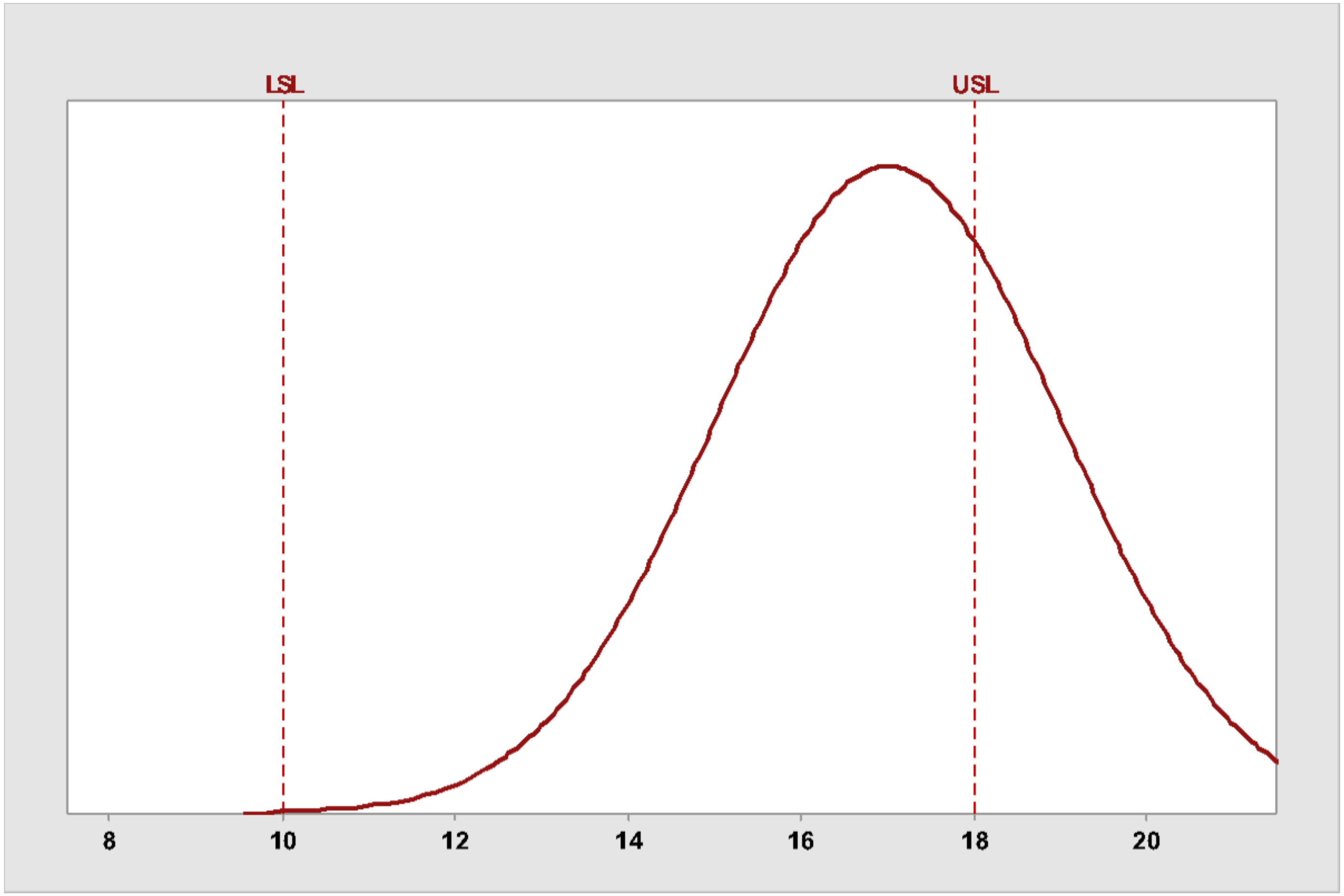
How can we Improve our Process?



Mean = 17
StDev = 1

Pp = 1.33
Ppk = 0.167

How can we Improve our Process?



Mean = 17
StDev = 2

Pp = 0.67
Ppk = 0.083

How can we Improve our Process?

So first how would we maximize Pp?

Let’s look at the formula:

$$Pp = \frac{\boxed{USL - LSL}}{6\sigma}$$

Can we change the specs?

Can we change the Standard Deviation?

How can we Improve our Process?

Let's throw out a scenario:

USL = 18 LSL = 10 Target Ppk= 1.33

Standard Deviation(σ) = 1.5

$$Pp = \frac{USL - LSL}{6\sigma} = \frac{18 - 10}{6 * 1.5} = 0.88$$

How can we Improve our Process?

USL = 18 LSL = 10 Target Ppk= 1.33

Standard Deviation(σ) = $\boxed{1.0}$

Reduction of Variation by 33%

$$Pp = \frac{USL - LSL}{6\sigma} = \frac{18 - 10}{6 * 1} = 1.33$$

What have we learned so far? (Obj. 1)

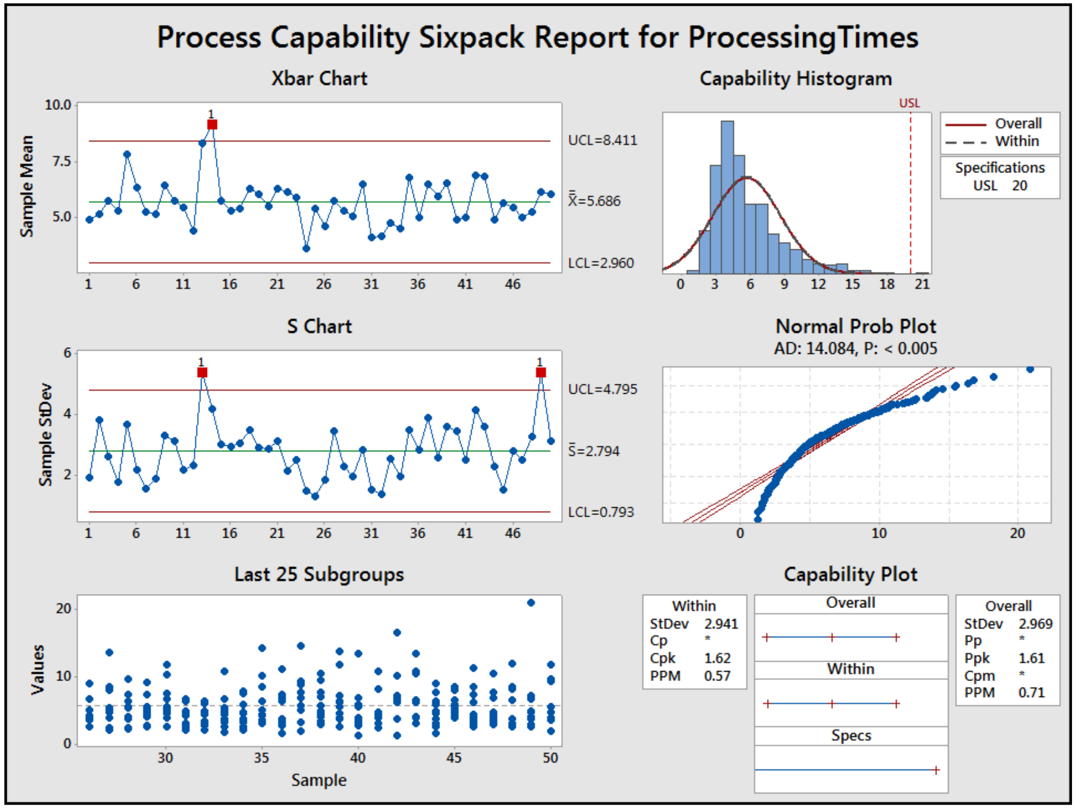
- Capability analysis is sensitive to the following assumptions.
  - Distribution assumption
  - Stability of the process
- Look at confidence intervals.
- To improve the capability of a process you can:
  - Center the mean
  - Reduce the variation
  - Re-evaluate the specs

What to do when Data are not Normal?

1. Use a transformation
2. Find a Non-Normal (NN) distribution to model the data
3. Use a nonparametric method
4. Options above do not apply or are not feasible, ask yourself why is the data not normal?

NN Capability – Transformations

- Let’s analyze the processing times (min) to complete a specific task.

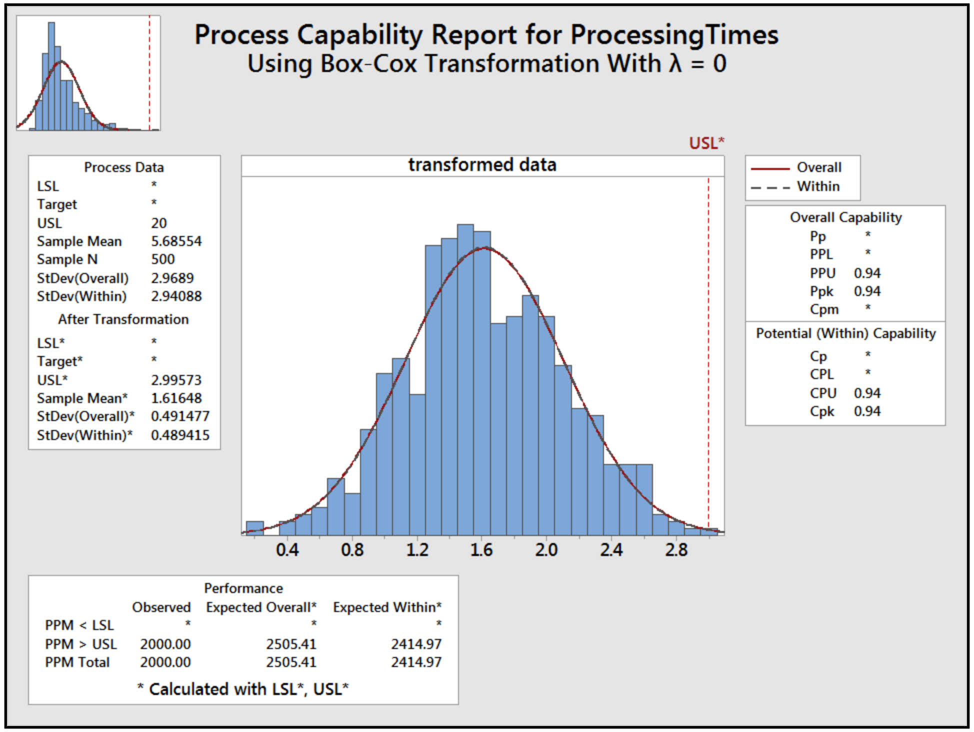


ASQ

NN Capability – Transformations

- The Box-Cox transformation

$$Y^* = Y^\lambda \ \ where\ \lambda \in [-5, 5]$$

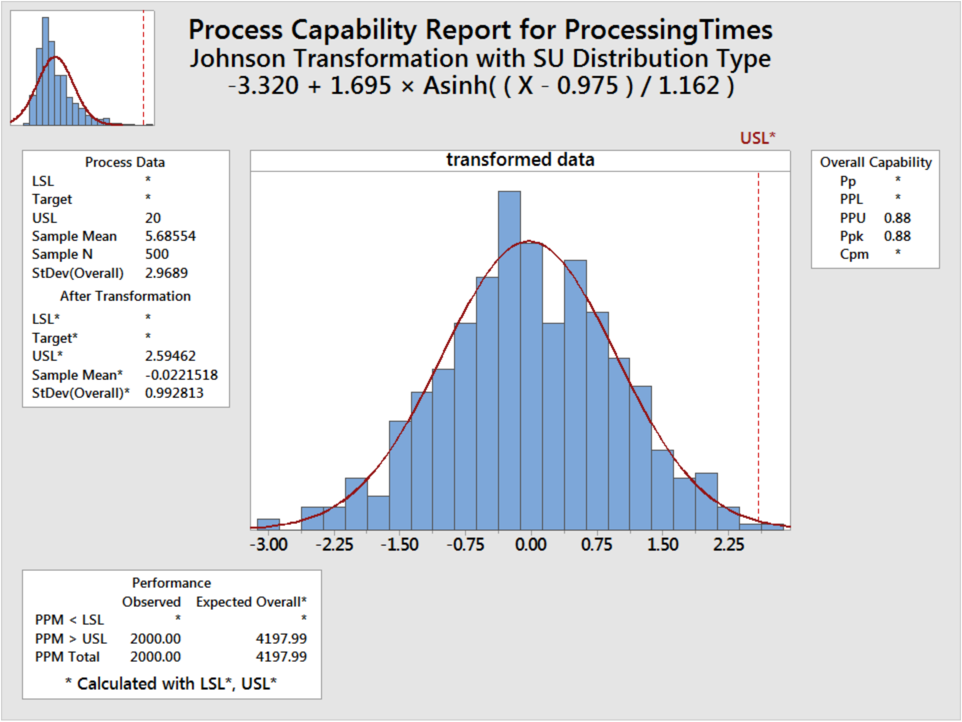


Note: Except when $\lambda = 0$, $Y^* = \ln(Y)$

NN Capability – Transformations

- The Johnson transformation

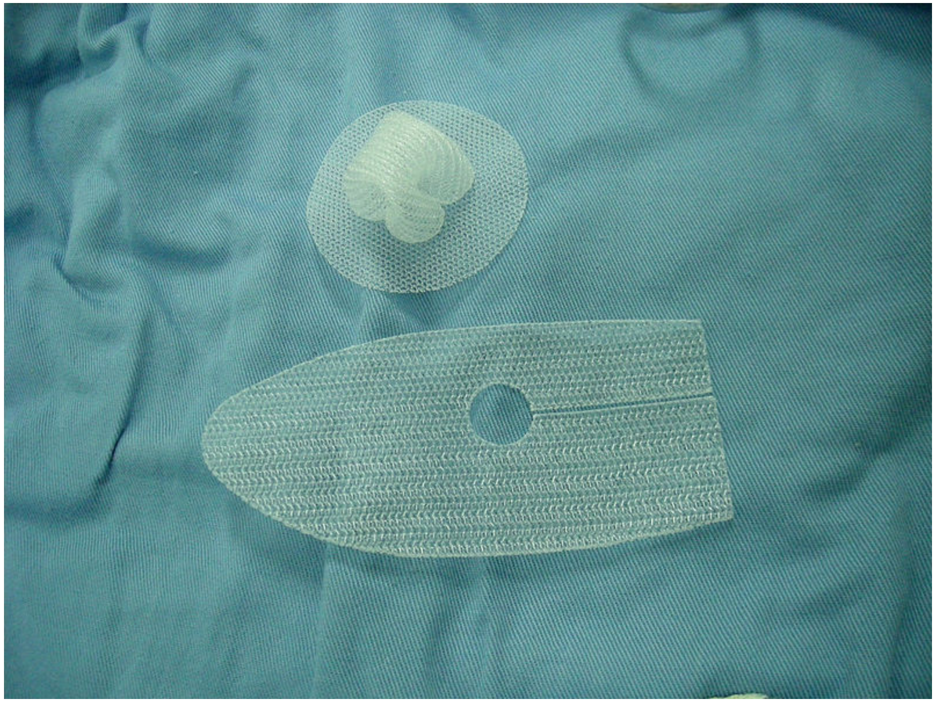
$$Y^* = a + b \cdot \ln(f(Y))$$



Note: $\mathrm{Asinh}(z) = \ln(z + \sqrt{z^2 + 1})$

NN Capability – Alternative Distribution

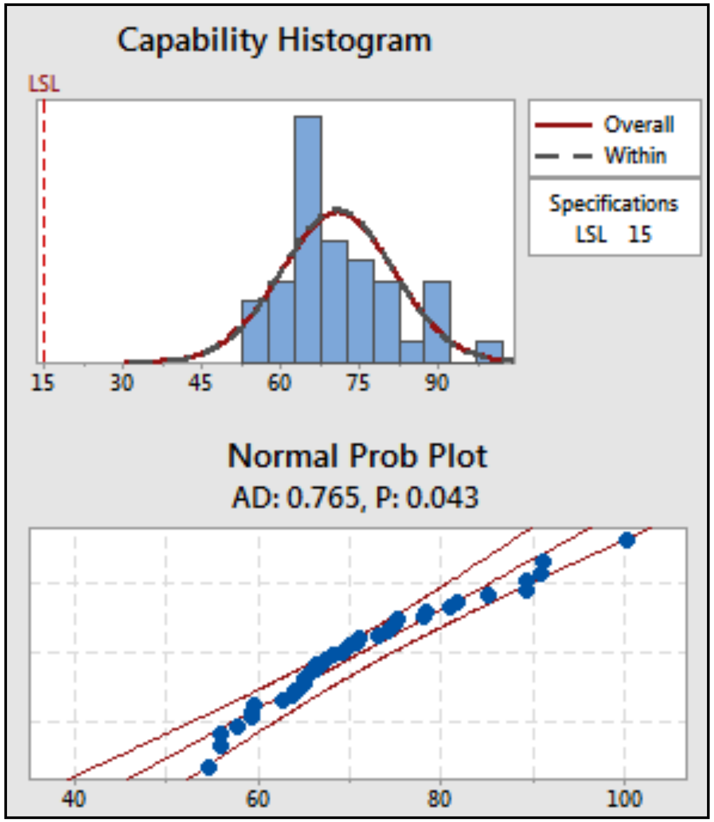
- Inguinal hernia surgery case study.
- The company wants to assess the process capability.
- LSL = 15 lbs. on a ball burst strength test.

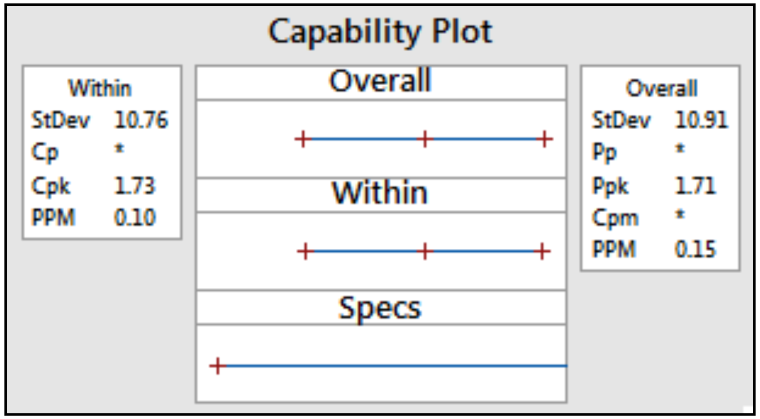


Source: http://en.wikipedia.org/wiki/Inguinal_hernia_surgery

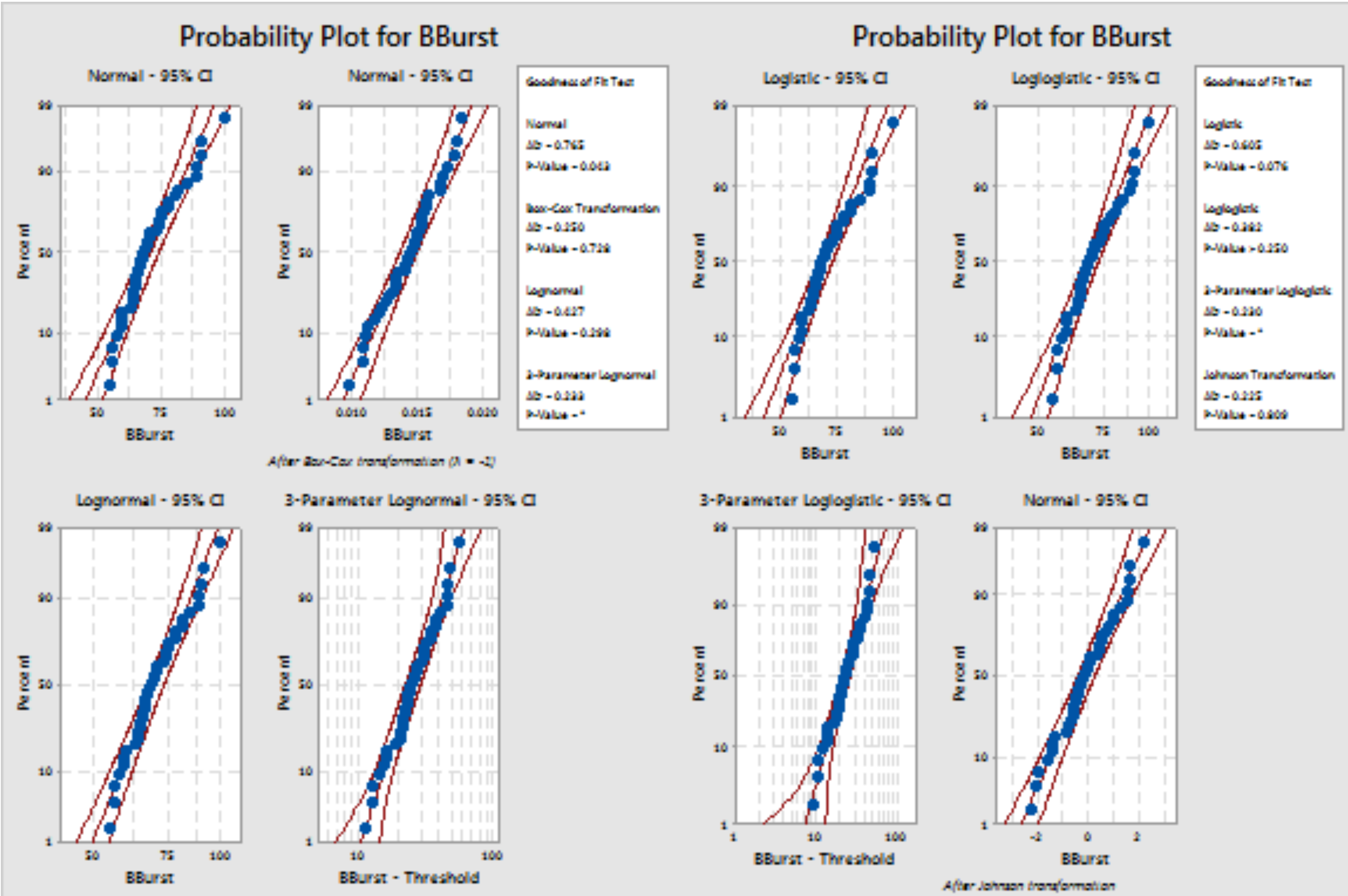
NN Capability – Alternative Distribution

- Underlying data is not normally distributed.
- Capability estimates are invalid.





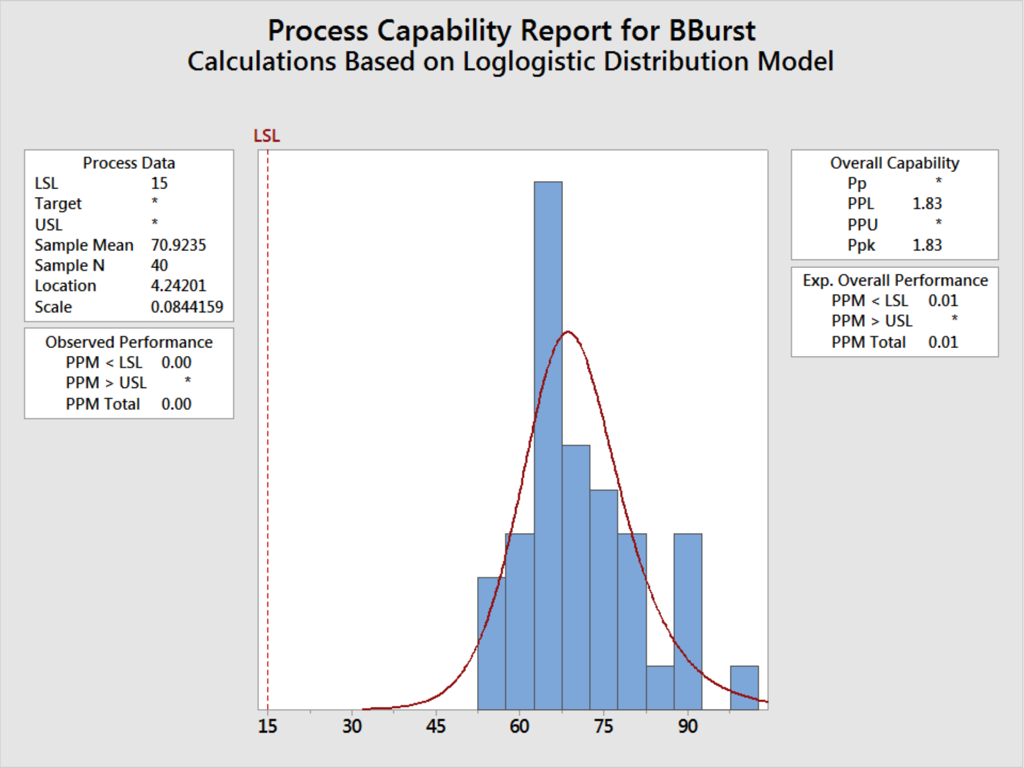
NN Capability – Alternative Distribution



ASQ

NN Capability – Alternative Distribution

- Select the Loglogistic distribution to fit the data and estimate the capability of the process.



NN Capability – Nonparametric

- Use a nonparametric method to estimate the capability of the process
  - See McCormack et al. [3], larger sample sizes required
  - Treat the data as binary and perform a test for proportion

NN Capability – Nonparametric

- **Example.** A process measuring the wet weight of a product has to validate the process is capable.

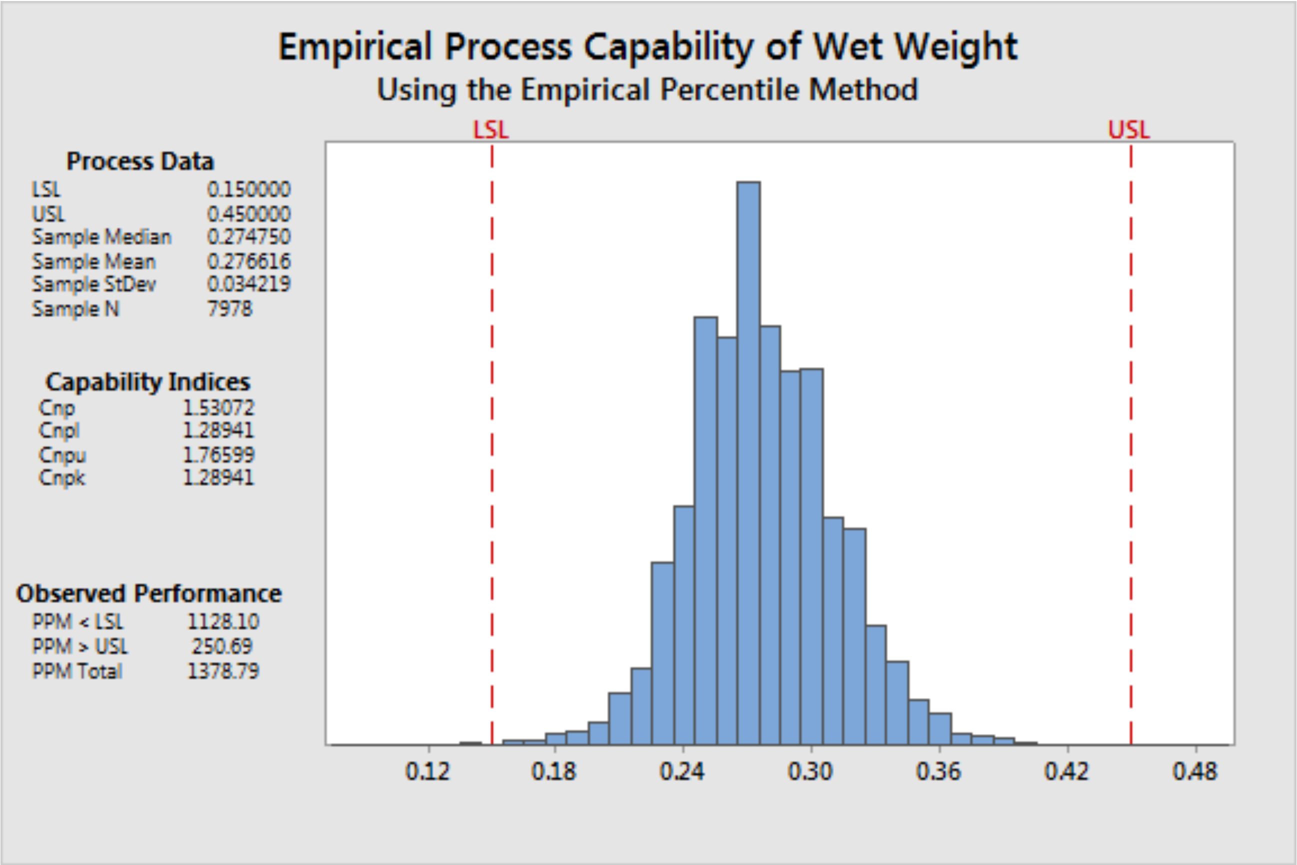
Test and CI for One Proportion: Pass or Fail

Event = P

Variable	**X**	**N**	**Sample p**	**95% Lower Bound**
Pass or Fail	7967	7978	0.998621	0.997719

NN Capability – Nonparametric

- Wet weight analysis using nonparametric capability.



ASQ

What have we learned? (Obj. 2)

- When data are Non-Normal, multiple techniques can be utilized:
  - Use a Johnson or Box-Cox transformation
  - Model the data with a Non-Normal distribution, e.g. Weibull, Lognormal, Smallest Extreme Value
  - With large sample sizes you can utilize nonparametric approaches
- If no approach works for your specific situation, investigate what makes data Non-Normal. How?

Why are the Data not Normal?

Case I. Nature of the beast – process near a boundary, naturally produces data that are skewed.

Case II. Mixture of distributions or a few outliers – process may not be in statistical control.

Why are the Data not Normal?

Case III. Large sample sizes – power of normality tests detects small departures from “perfect” normality.

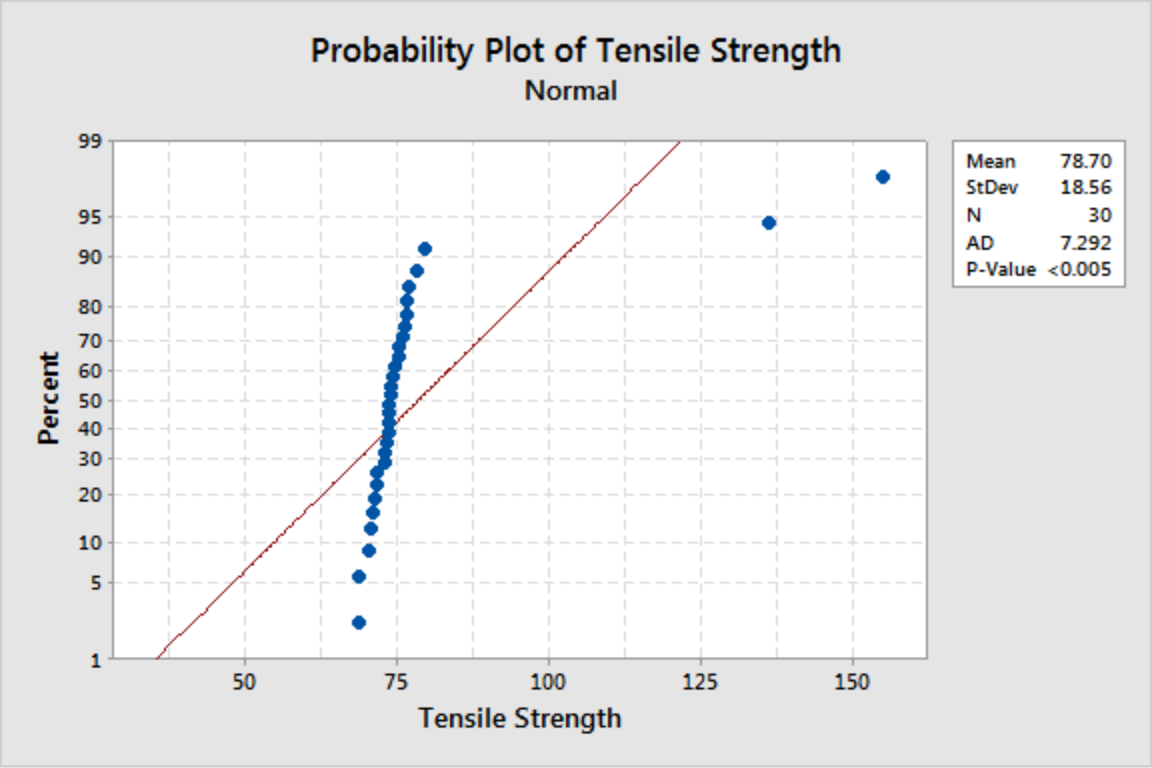
Case IV. The number of significant digits is not sufficient to differentiate between parts, rendering the classic normality tests ineffective.

Normality Test Failed – Case I

- This scenario is typically straightforward, the selection of a non-normal distribution is typically done using a distribution identification tool or scientific knowledge about what distribution models a specific situation.

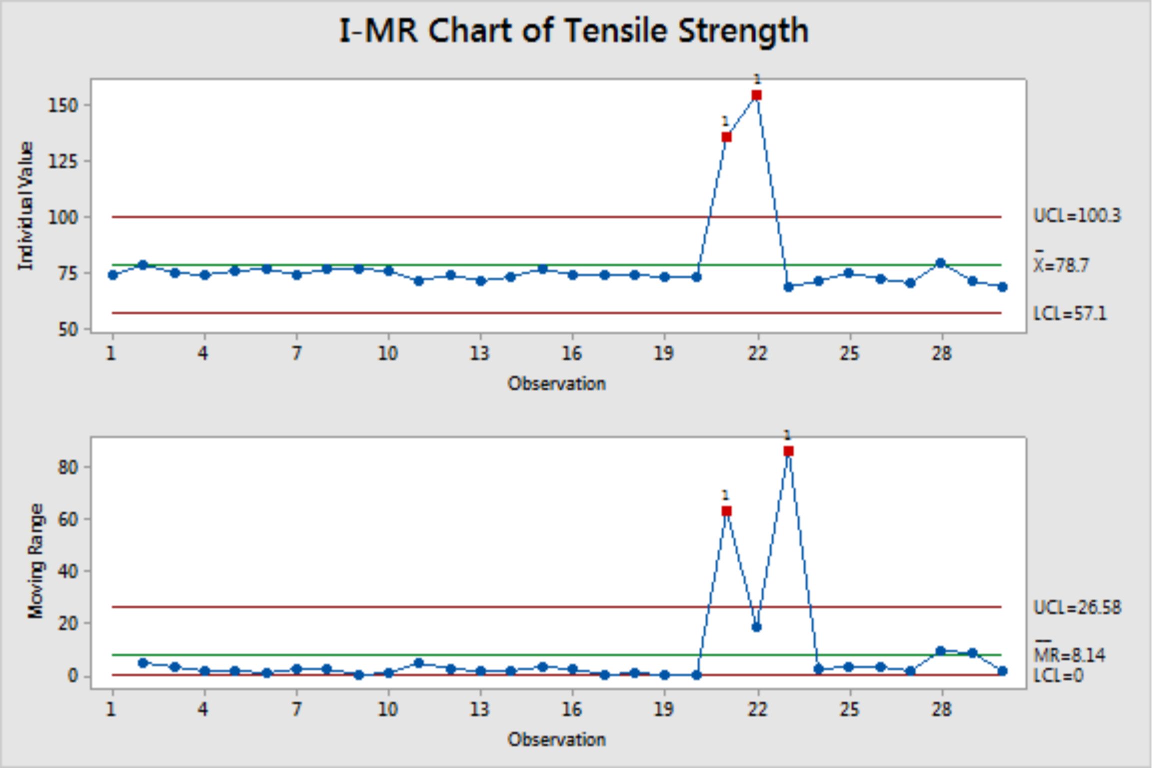
Normality Test Failed – Case II

- The sample obtained to estimate the capability of a process can include data from different distributions.



Normality Test Failed – Case II

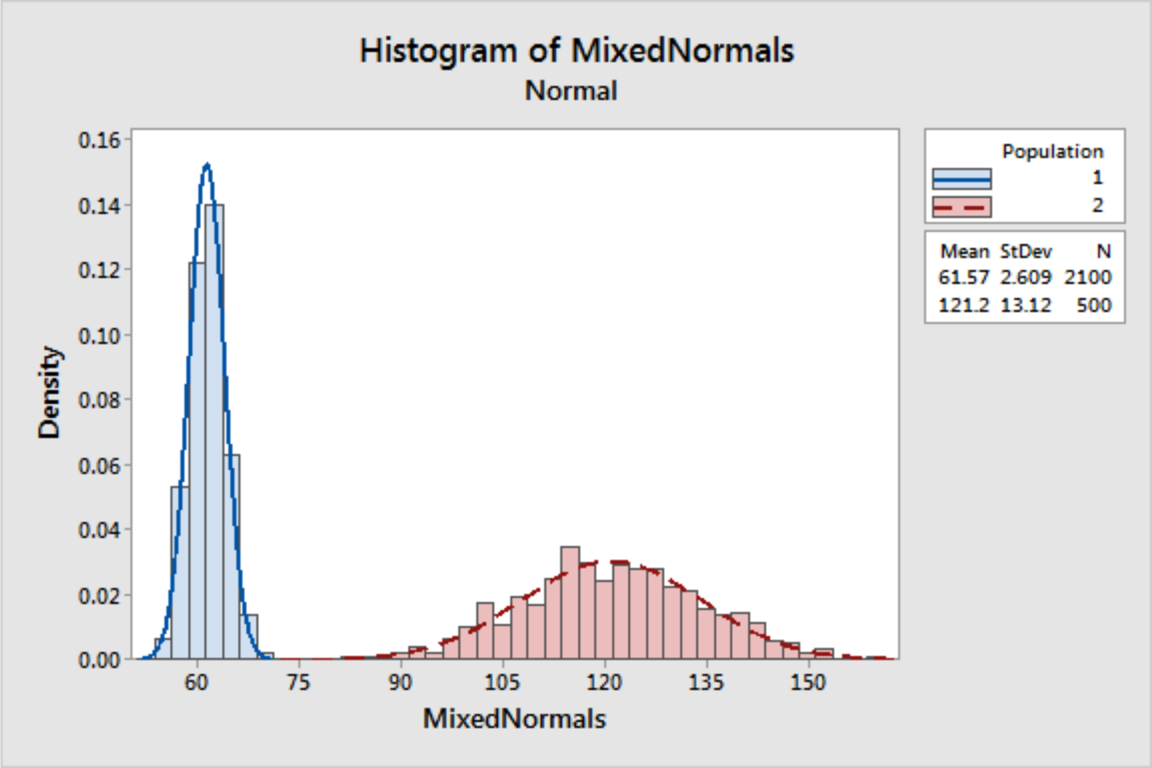
- Detection of multiple distributions can be done with a control chart.



ASQ

Normality Test Failed – Case II

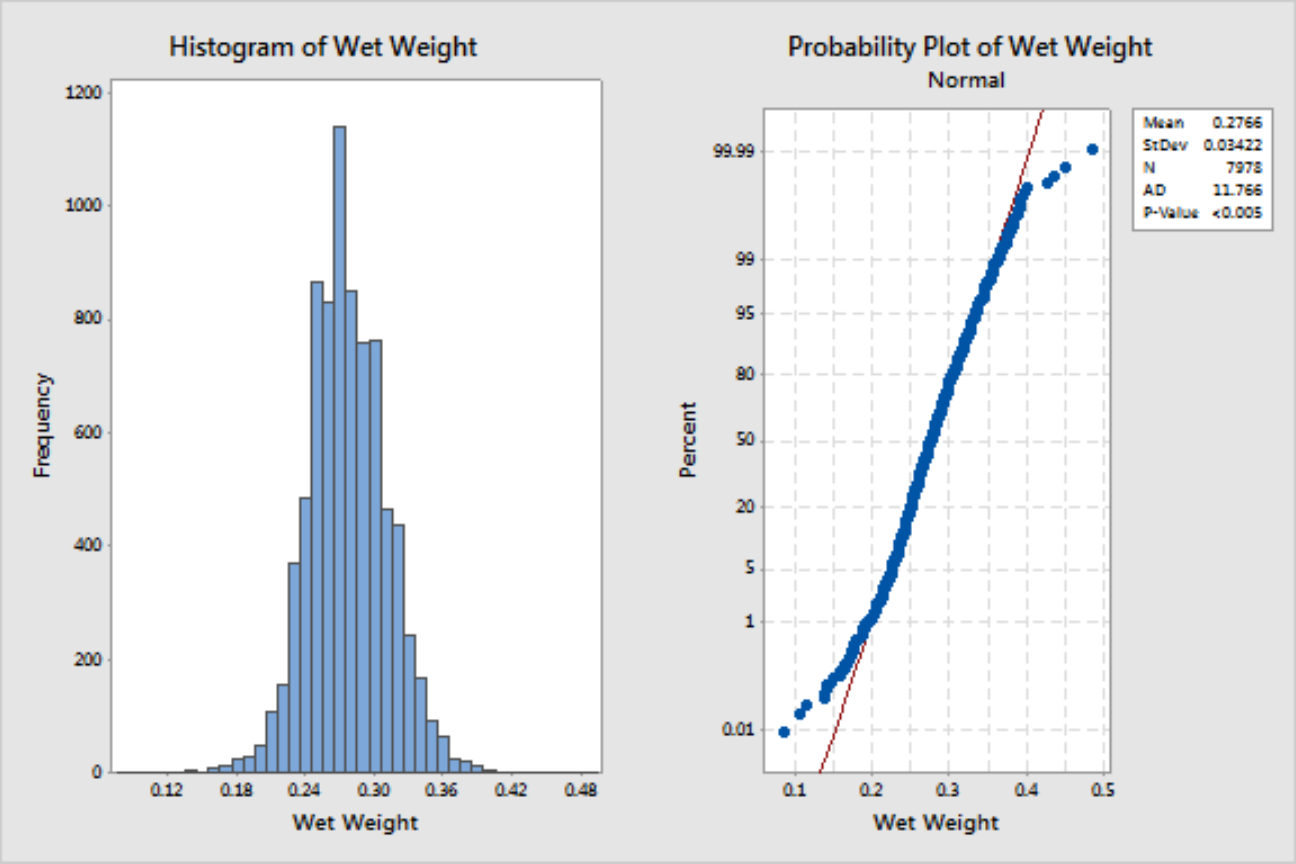
- What to do in a situation like this?
- Better control of the process. Implement corrective and preventive actions.



ASQ

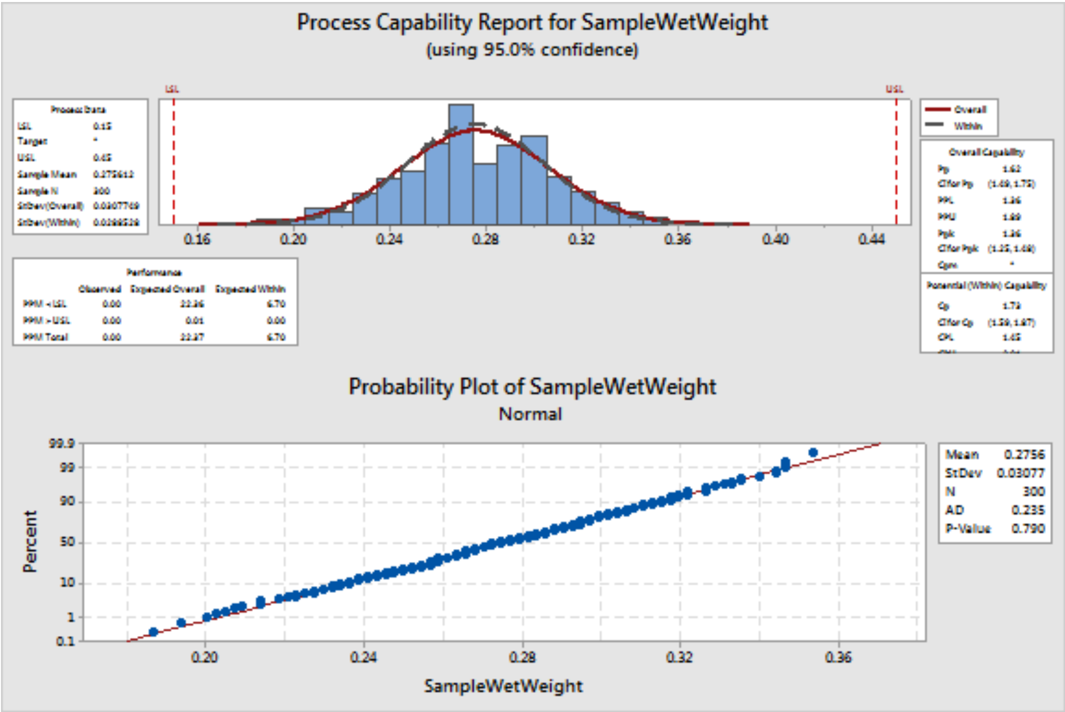
Normality Test Failed – Case III

- Revisit the “Wet Weight” example with over 7,000 observations. Data looks normal.



Normality Test Failed – Case III

- What is the issue?
- To quote G.E.P. Box in a slight different context: “All models are wrong, but some of them are useful.”



Easy solution: Use a nonparametric approach or get a random sample from your data.

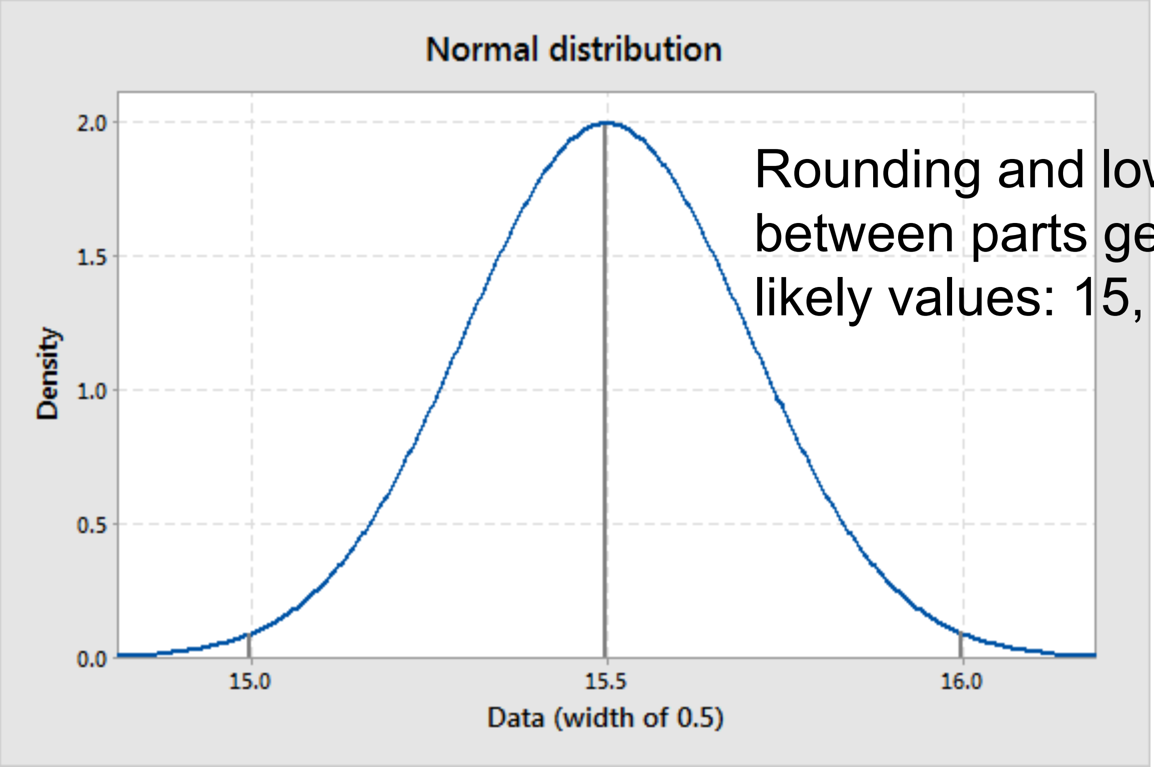
Normality Test Failed – Case III

- Comparing the capability estimates for the overall defect level for all analyses done on this dataset.

Method	Cpk	Defect rate	Yield
One proportion	N/A	0.14%	99.86%
Nonparametric percentile method	1.29*	0.14%	99.86%
Normal method	1.45	0.0022%	99.9978%

Normality Test Failed – Case IV

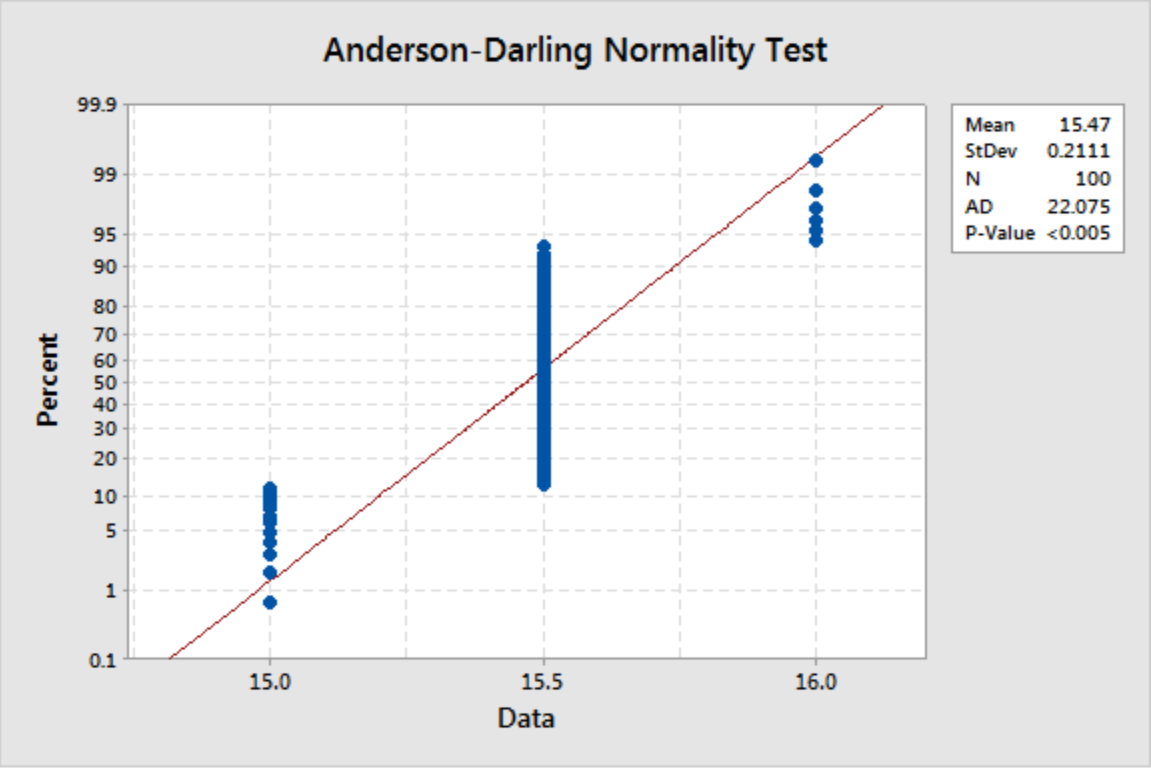
- Low discrimination of the measurement system.



Rounding and low discrimination between parts generates only three likely values: 15, 15.5, 16.

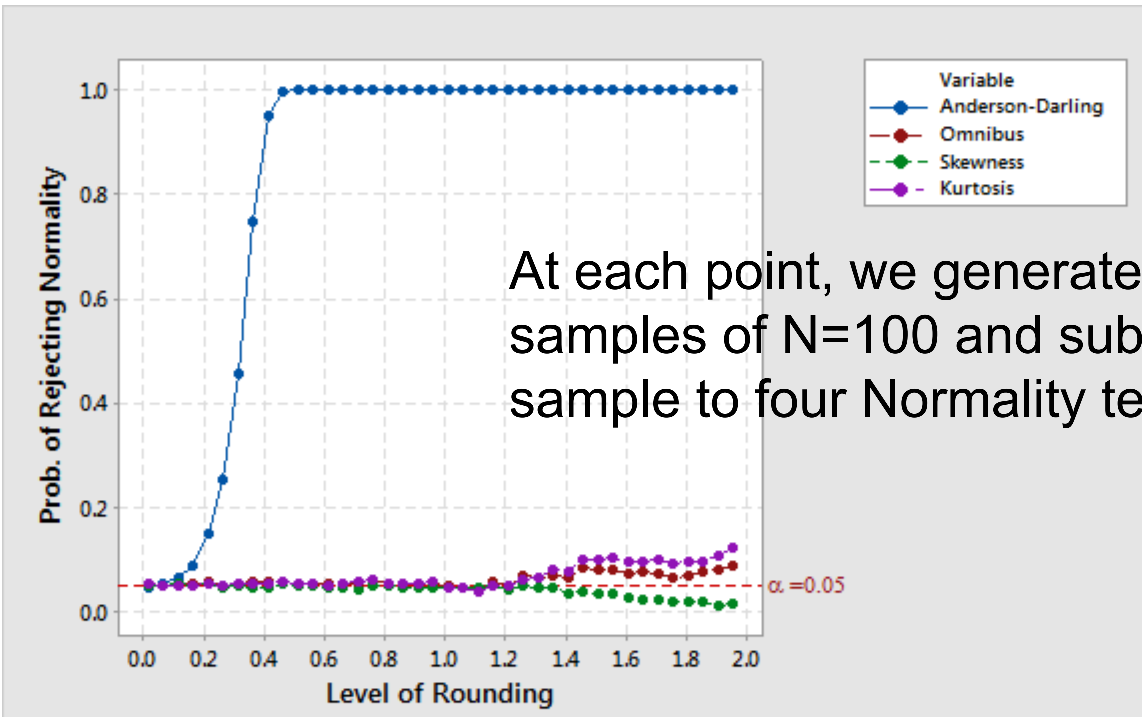
Normality Test Failed – Case IV

- Ties lead to the Anderson-Darling (AD) test failing normality.



Normality Test Failed – Case IV

- For heavily rounded datasets the Skewness (SK), Kurtosis (KT), Omnibus (OB) tests may be preferred.

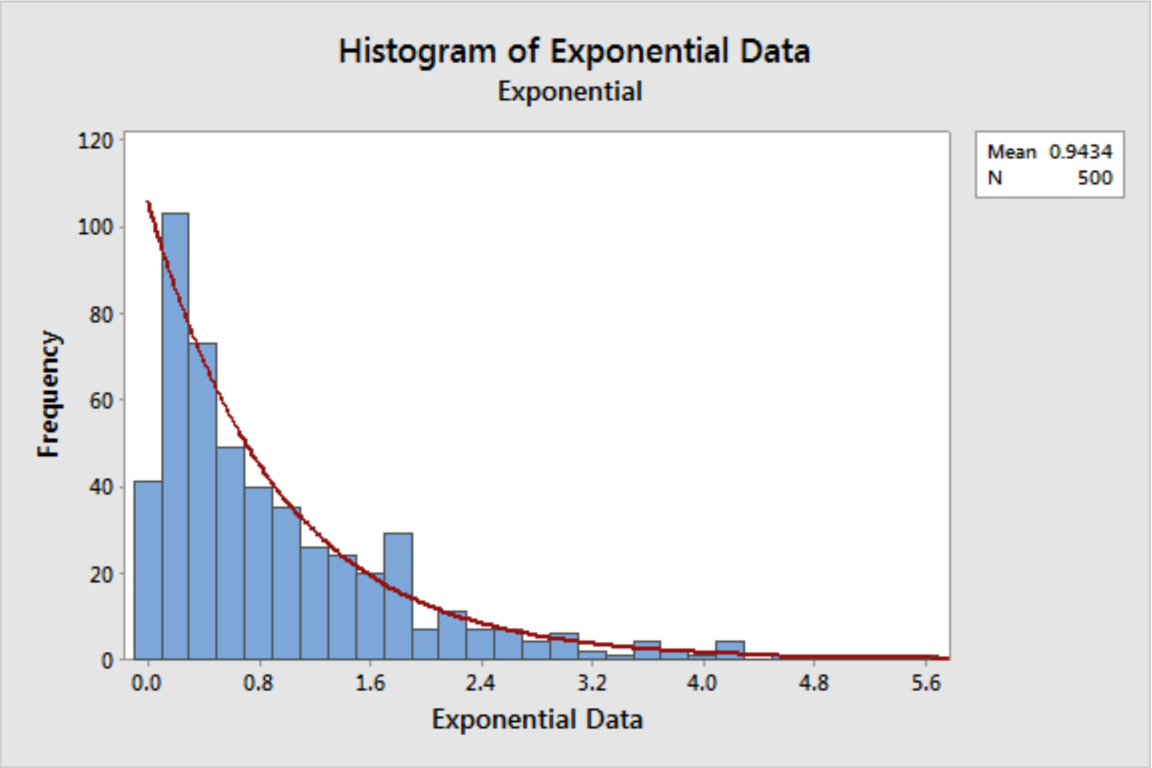


Note: Each simulation was made with a sample size of 100 normally distributed observations.

At each point, we generated 5000 normal samples of N=100 and subject each sample to four Normality tests.

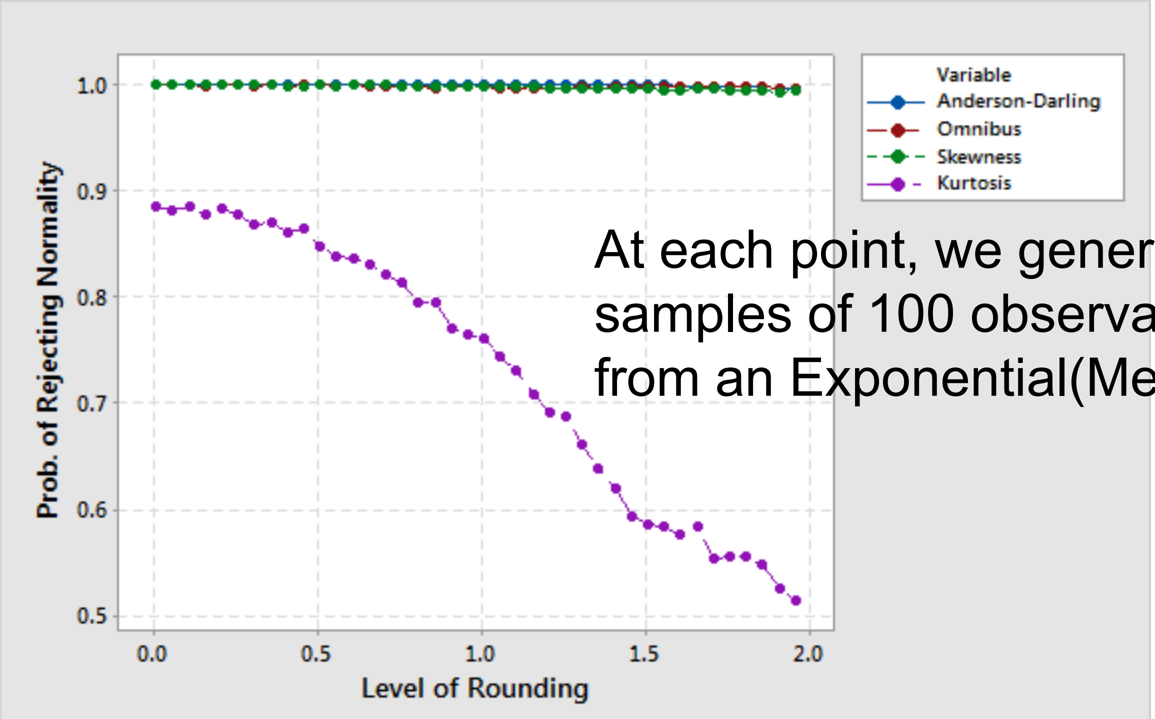
Normality Test Failed – Case IV

- Let's consider now a highly skewed distribution, namely the exponential distribution.



Normality Test Failed – Case IV

- The AD seems to perform as well as the OB and SK tests, but its detection of non-normal data comes as a consequence of rounding.



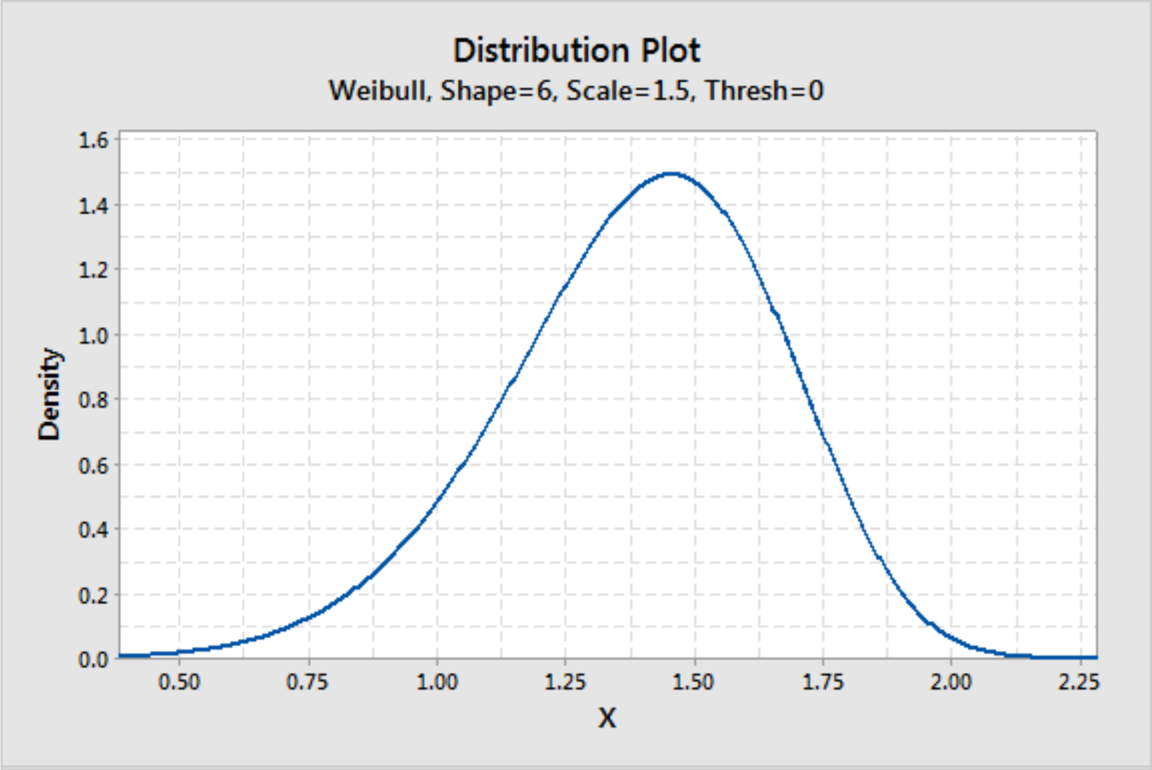
Note: Each simulation was made with 100 samples from an Exponential distribution.

At each point, we generated 5000 samples of 100 observations each from an Exponential(Mean=1).

ASQ

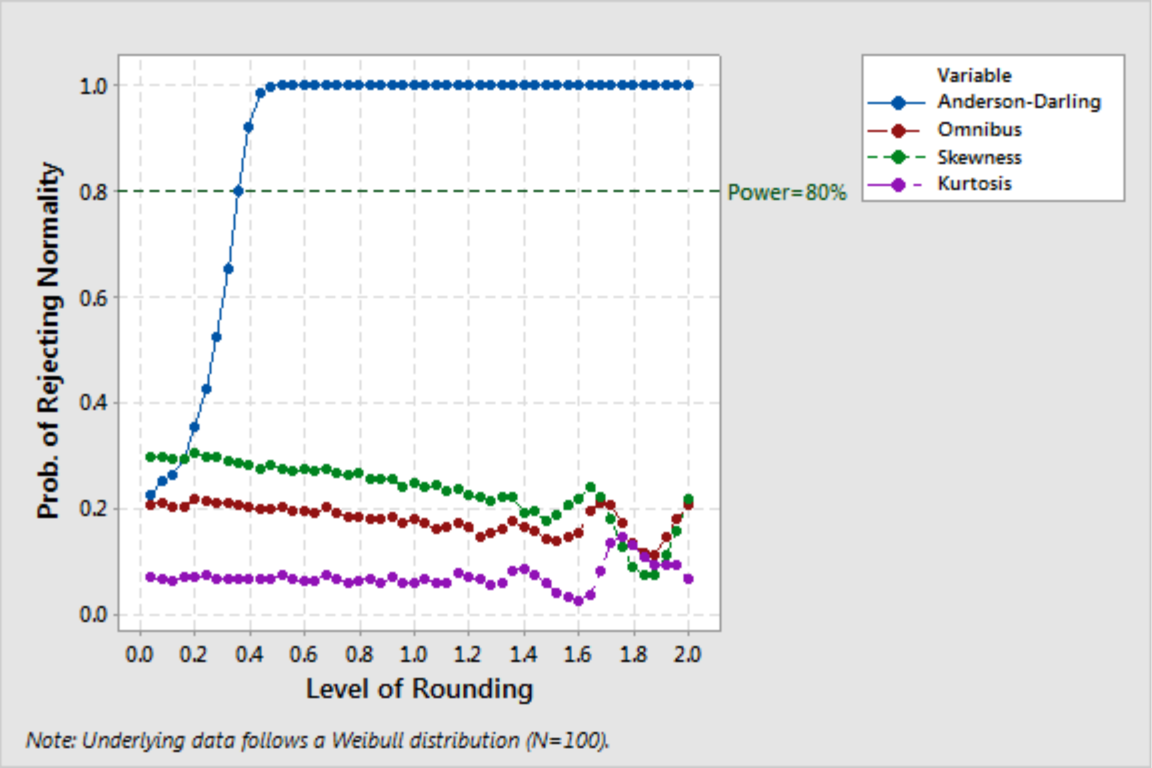
Normality Test Failed – Case IV

- What if the data is not severely skewed but still not normally distributed?



Normality Test Failed – Case IV

- The only test that seems to consistently reject normality (AD) does it as a consequence of ties, not effective detection.



What Normality Test to Use?

- Skewness fails to reject normality at the expected α level when the data are normal.
- Skewness has good power and is less sensitive to the degree of rounding.
- The Omnibus and Skewness tests have similar behavior.

What have we learned? (Obj. 3, 4)

- The nature of the data can make the use of the normal distribution inappropriate.
- Lack of controls in a process can produce samples that mix data from different distributions.
- Large sample sizes can make normality tests too sensitive.
- As the level of rounding increases, classic normality tests become less effective.

Capability with Rounded Data

- There a few approaches to estimate the capability of a process:

1. Classic approach
2. Adjust the estimates considering the bias induced by the measurement system
3. Handle the data as being interval-censored

Classic Capability Estimates

- The rounded data, denoted *Y**, is assumed to be normally distributed:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i^{\,*} \qquad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i^{\,*} - \bar{Y})^2}{n-1}}$$

- Proceed to estimate Cpk as usual.

Adj. Estimates – Sheppard's Correction

- Sheppard [5] describes the bias in the estimation of the standard deviation when the data is rounded.

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i{}^{*} - \bar{Y})^2}{n-1} - \frac{w^2}{12}}$$

where *w* is the width (incremental unit) of the measurement system.

Estimates from Interval-Censored Data

- Suppose we don't know the exact value that corresponds to the true measurement but we know the interval in which it must be.

Observed Values
5.1
5.1
5.0
5.2
5.1
5.0

Converted to

Start	End	Frequency
4.95	5.05	2
5.05	5.15	3
5.15	5.25	1

Estimates from Interval-Censored Data

- We get the following estimates:

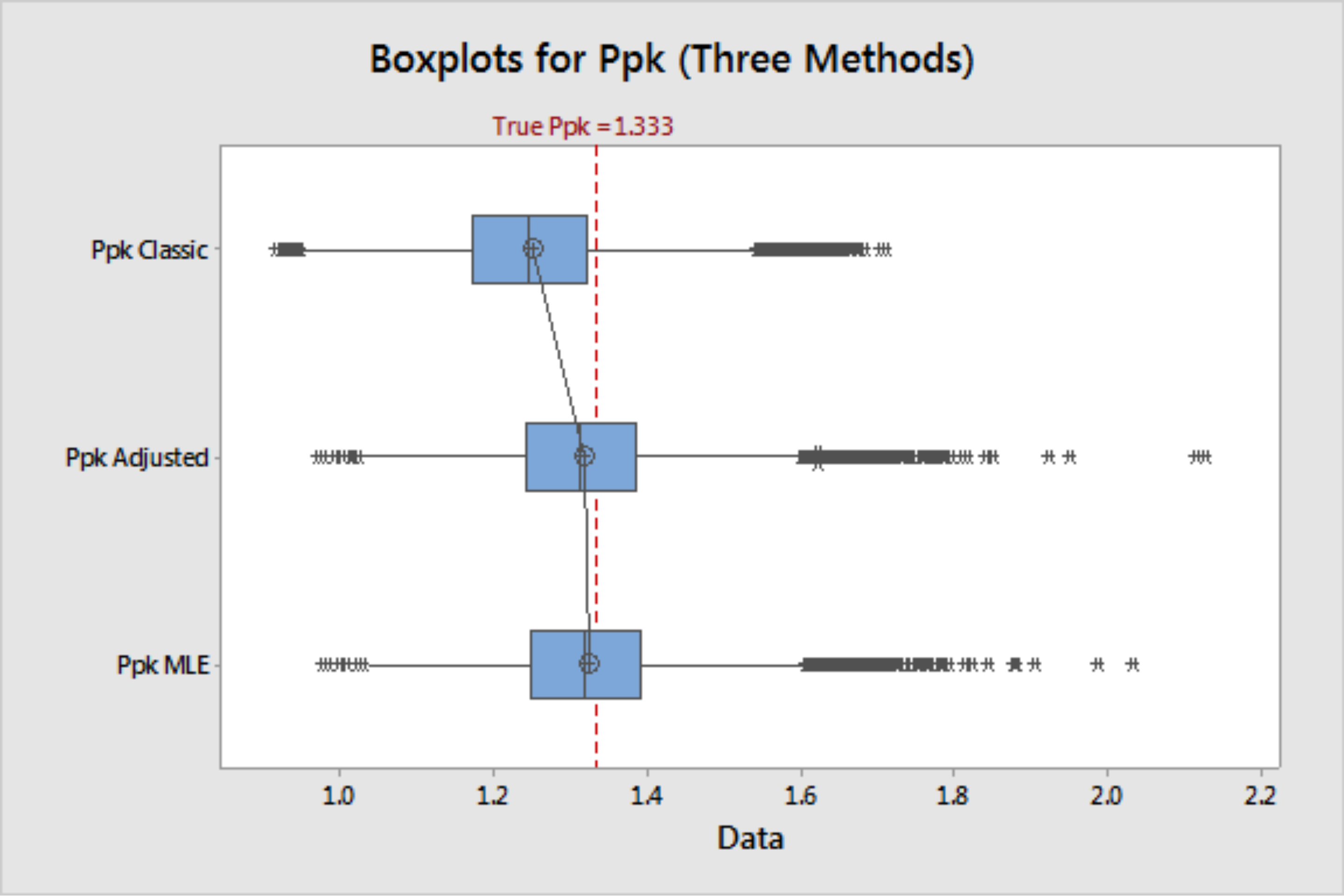
```
Variable Start: Start  End: End
Frequency: Frequency

Censoring Information     Count
Interval censored value       6
```

Estimation Method: Maximum Likelihood (MLE)

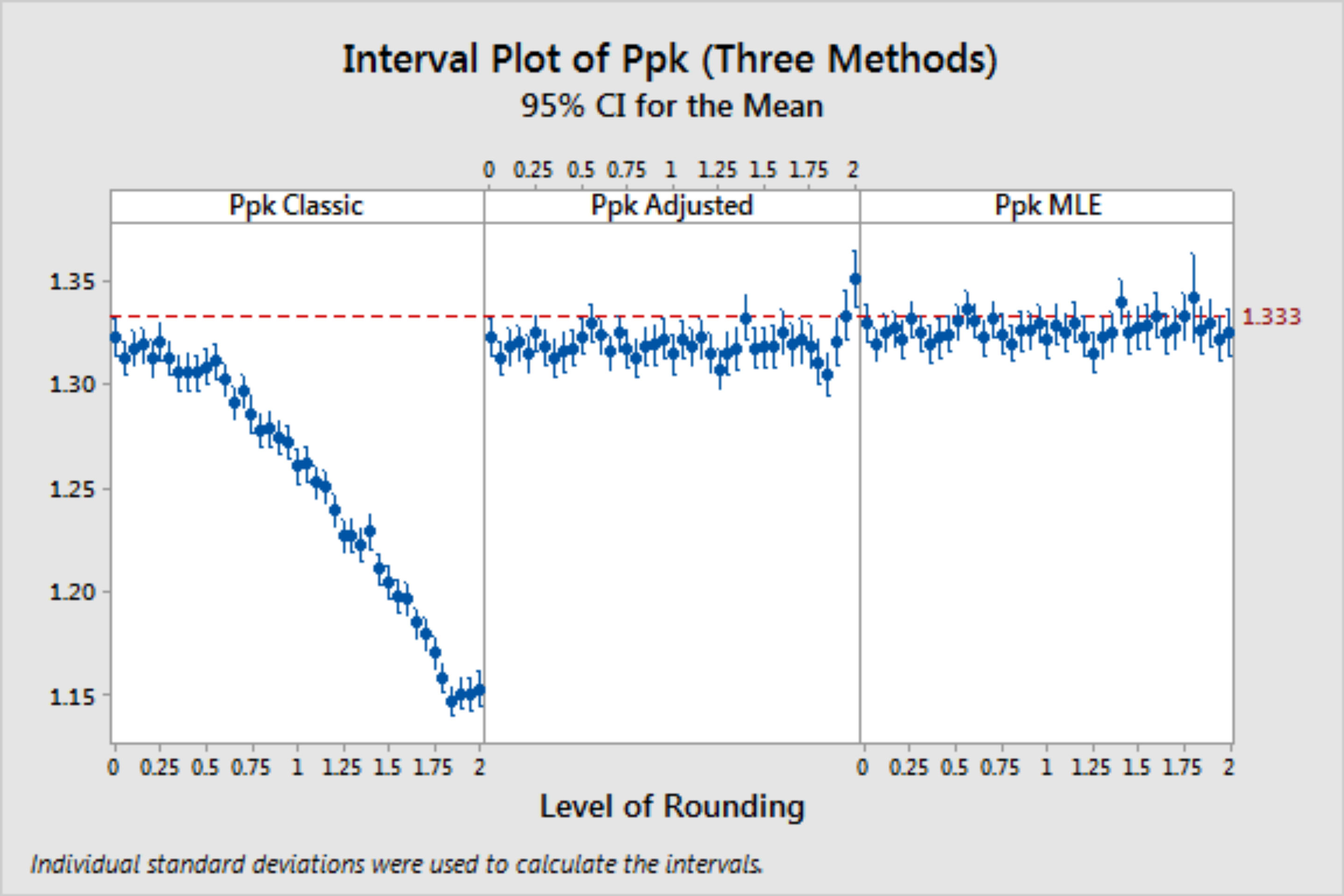
```
Distribution:   Normal

Parameter Estimates
```

Parameter	Estimate	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
Mean	**5.08345**	0.0278668	5.02883	5.13807
StDev	**0.0619204**	0.0218490	0.0310086	0.123648

Capability with Rounded Data



Capability with Rounded Data



What was the last lesson? (Obj. 5)

- Rounded data makes classic normality tests fail. [*They reject all the time no matter what type of data you have*]
- Rounded normal data should be symmetric, thus making skewness and omnibus tests extremely useful.
- Other issues in the data may not be captured by these tests.

What was the last lesson? (Obj. 5)

- Using interval-censored data with Maximum Likelihood Estimates (MLEs) seem to produce better estimates across the board.
- MLEs are asymptotically unbiased. [*As the sample size grows larger, the bias of the estimates becomes negligible*]
- The simplicity of Sheppard's adjustment to estimate σ makes it compelling too.

Conclusions

- The ultimate goal of a capability analysis is to estimate the defective level of a process.
- When the interest of an analysis is on the estimation of defects, the distribution assumption will be an important one.
- Another important assumption is ensuring the process is stable and in control.

Conclusions

- Of the two assumptions, normality is the one typically violated in practice.
- Non-normal (NN) capability analysis requires:
  - Using a transformation
  - Finding an alternative distribution that fits the data
  - Using a nonparametric approach which requires a large sample size

Conclusions

- Classic normality tests (AD, KS, SW) typically reject normality when the data is heavily-rounded regardless of the underlying distribution.
- When using a gauge with low discrimination, use different tests to check for normality, e.g. Skewness or the Omnibus test.

Conclusions

- If no evidence exists of the rounded data not being normal, assume normality.
- Utilize interval-censoring (MLE) to estimate the mean and standard deviation.

References

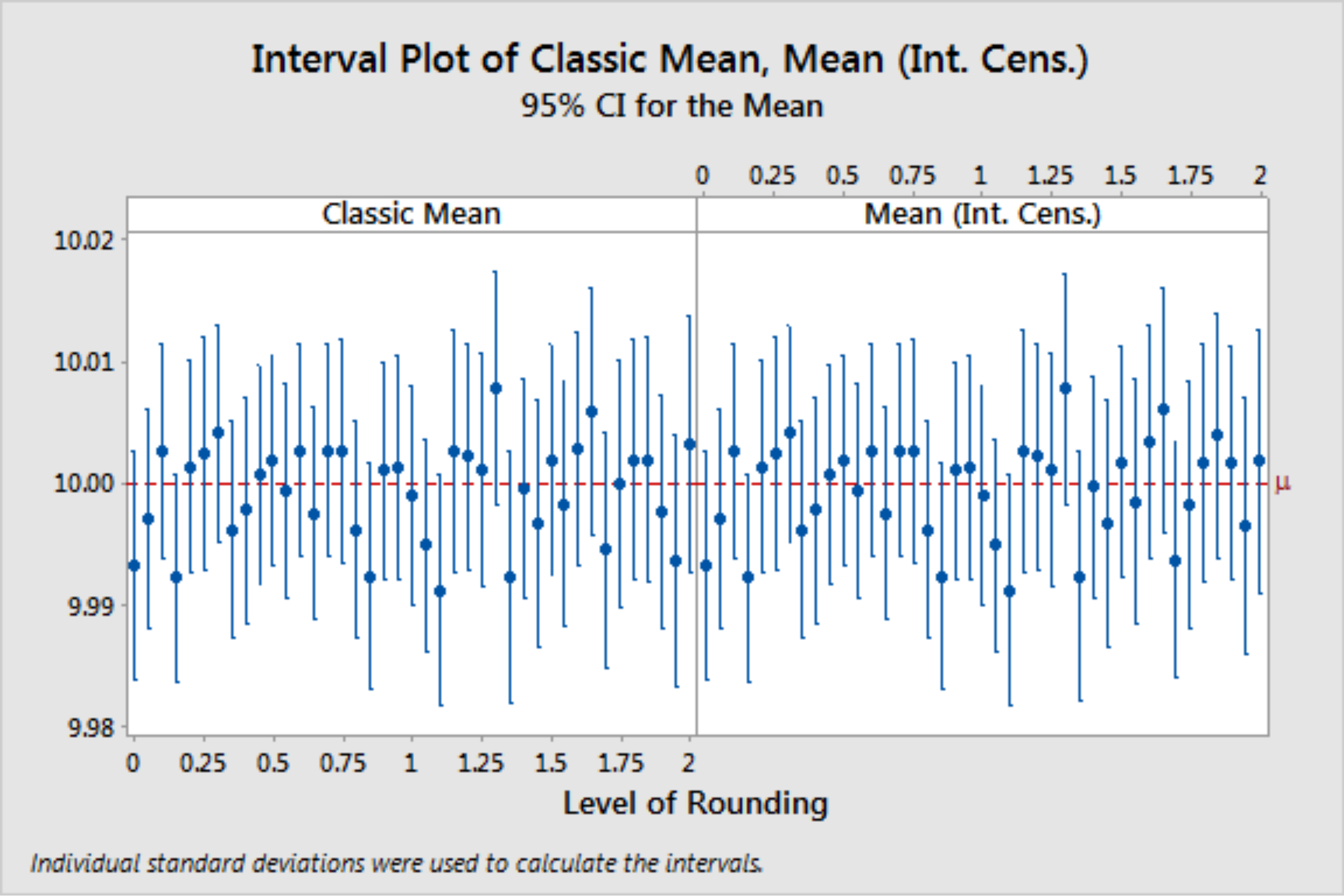
1. Juran, J.M., Godfrey, A.M. "Juran's Quality Handbook". 5th edition, McGraw-Hill. New York, 1999.
2. Kane, V.E. (1986) "Process Capability Indices". *Journal of Quality Technology*, 18, 41-52.
3. McComack, D.W., Harris, I.R., Hurwitz, A.M., and Spagon, P.D. (2000) "Capability Indices for Non-normal data", *Quality Engineering*. 12(4), 489-495.
4. Schneeweiss, H., Komlos, J., and Ahmad, A.S. (2006) "Symmetric and Assymetric Rounding." Working paper.
5. Sheppard, W.F. (1898). "On the calculation of the most probable values of frequency constants for data arranged according to equidistant division of a scale." *Proceedings of the London Mathematical Society.* 29, 231-258.
6. Tricker, A.R. (1984) "Effects of Rounding on the Moments of a Probability Distribution." *Journal of the Royal Statistical Society. Series D (The Statistician).* 33(4), 381-390.

ASQ

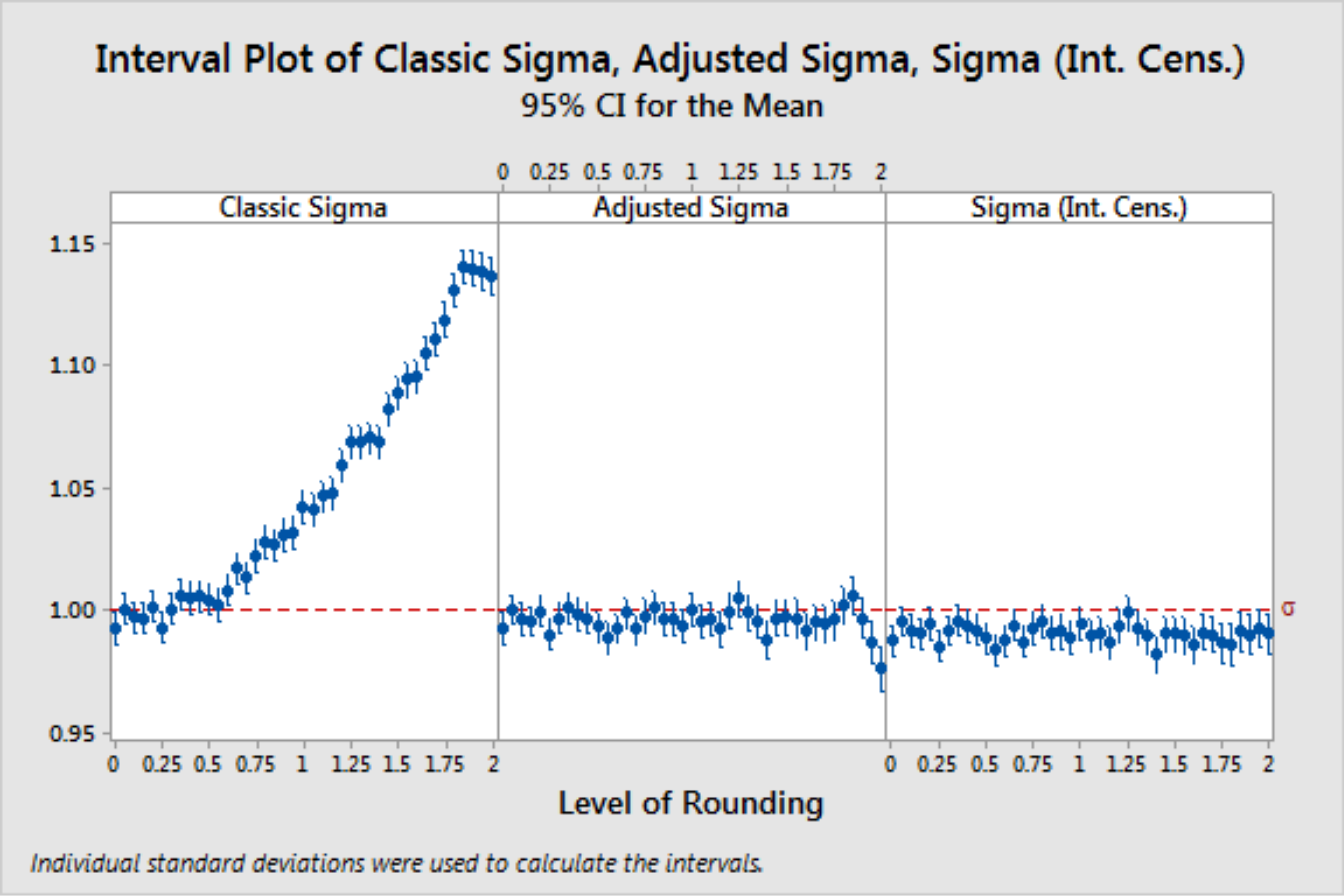
ASQ

ADDITIONAL SLIDES

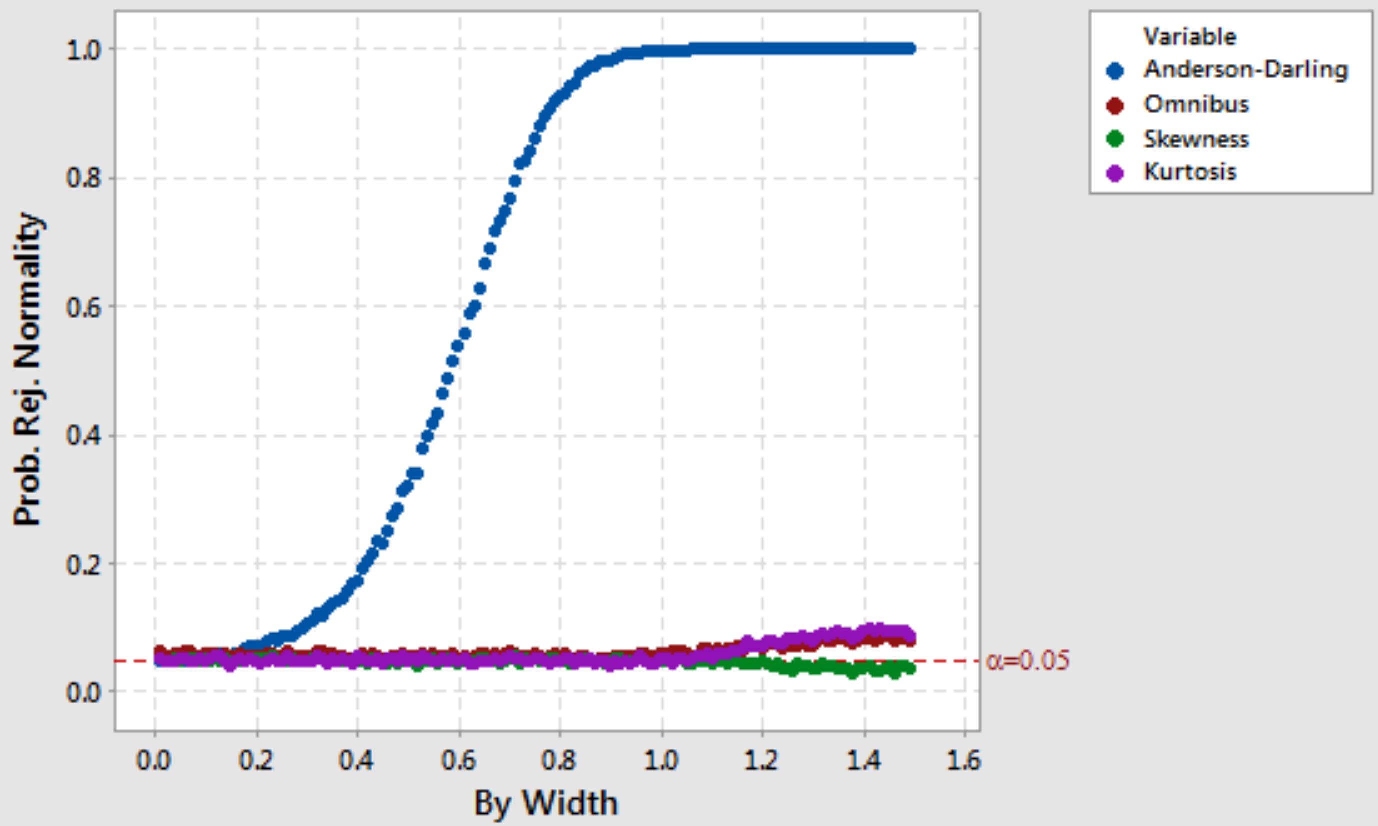
Capability with Rounded Data (N=100)



Capability with Rounded Data (N=100)

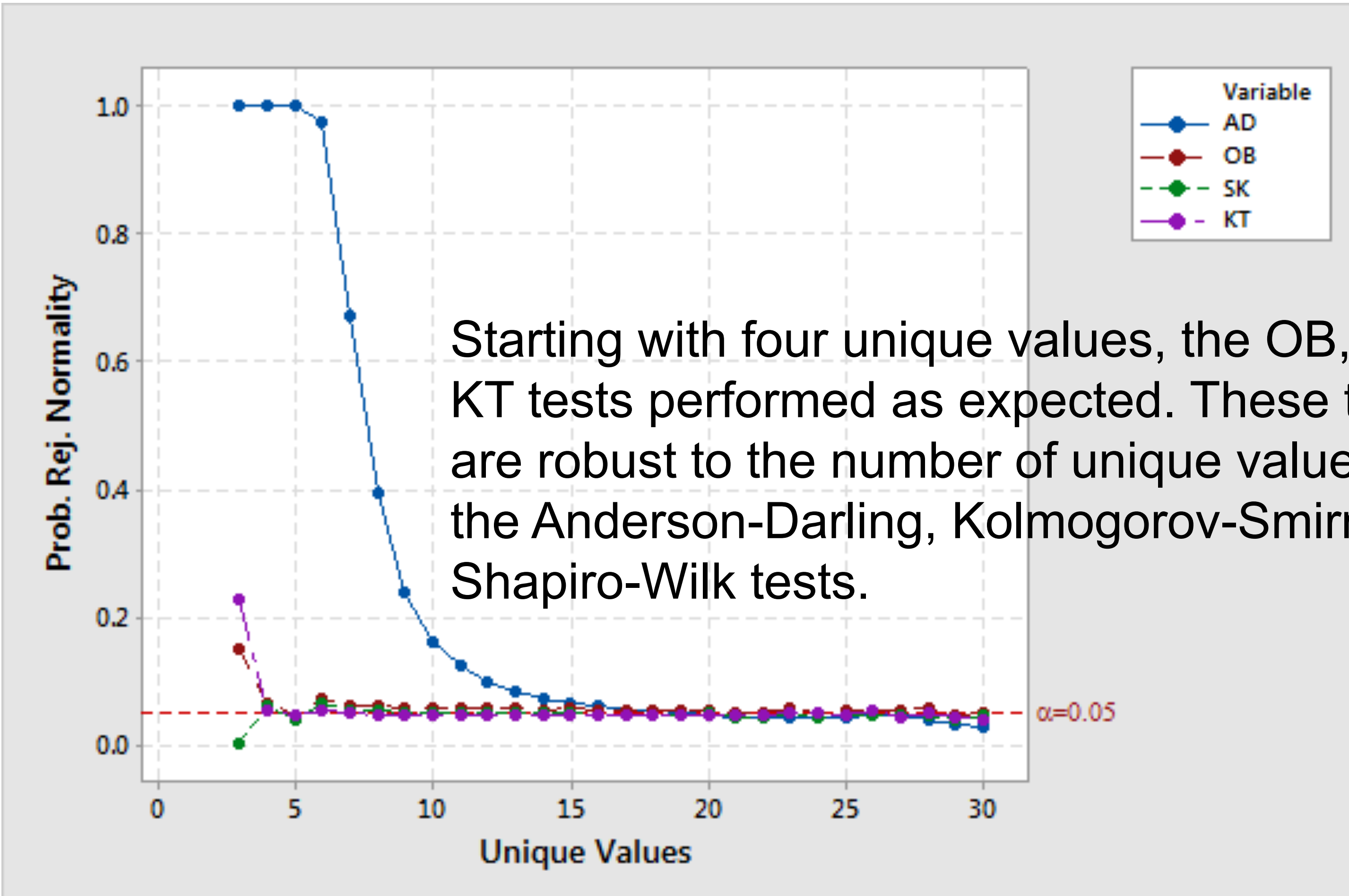


Normality tests for Normal data (N=30)



Note: Each simulation was made with a sample size of 30 observations.

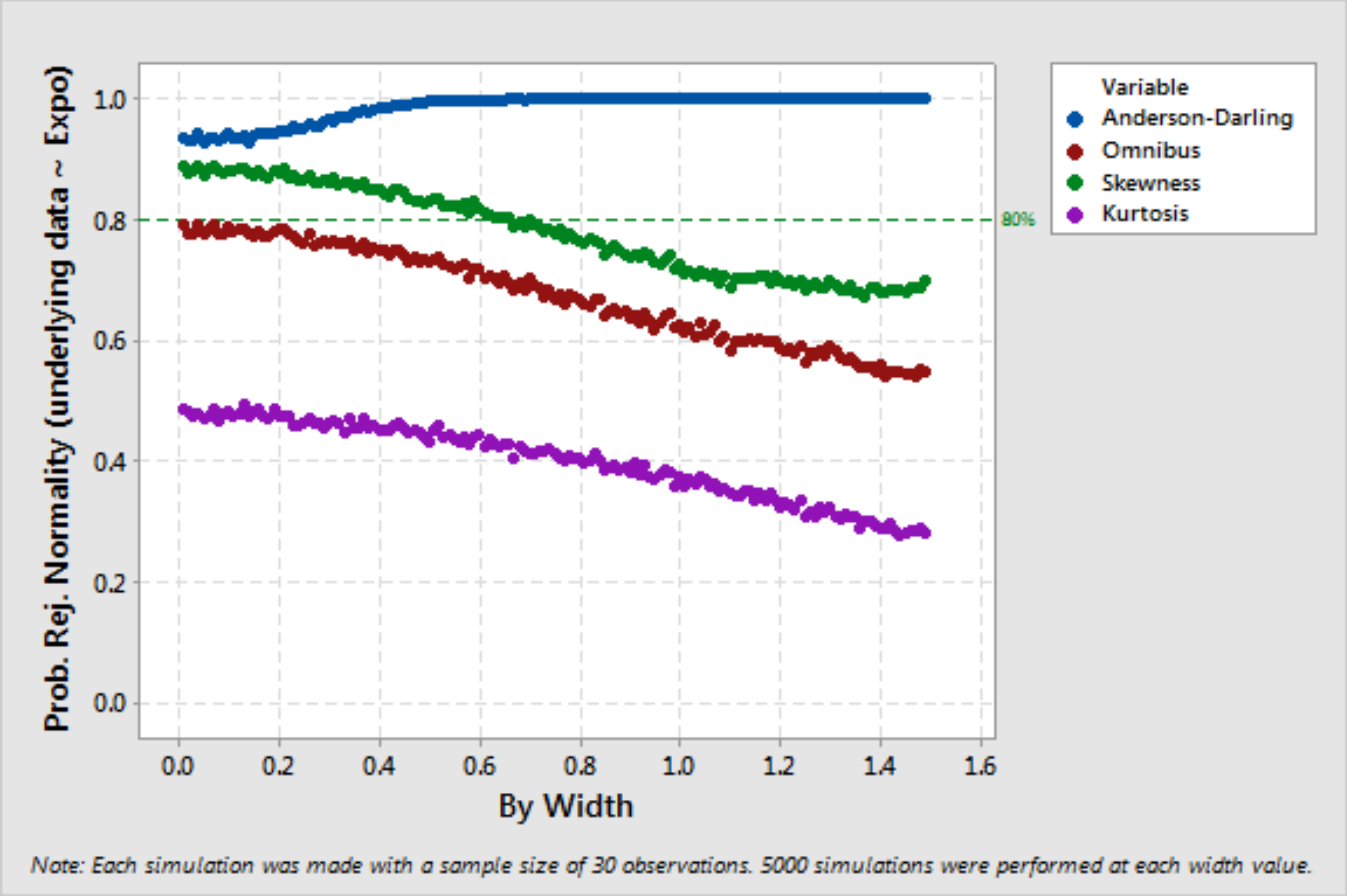
Normality tests performance (Ties)



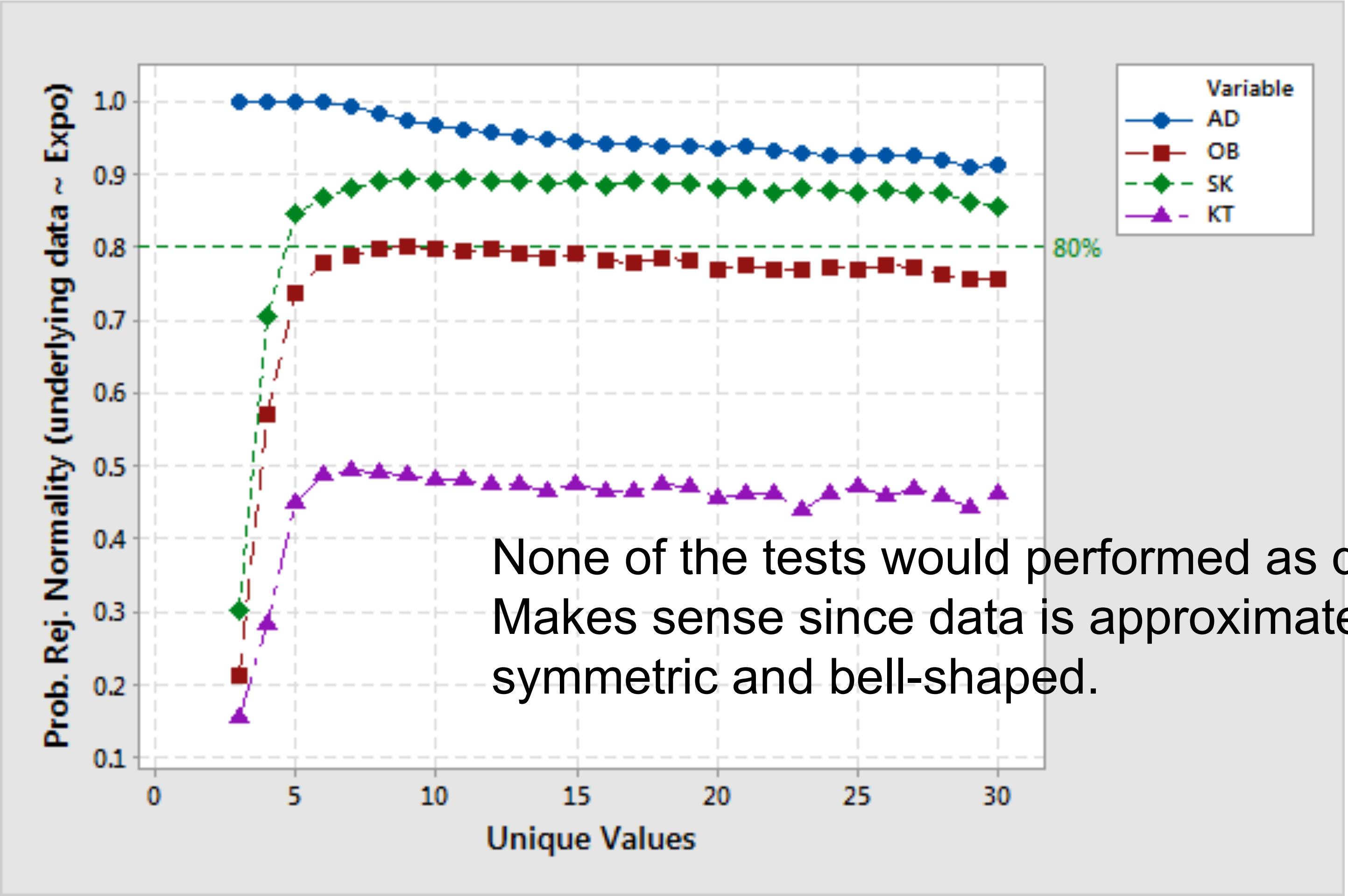
Starting with four unique values, the OB, SK, and KT tests performed as expected. These three tests are robust to the number of unique values, unlike the Anderson-Darling, Kolmogorov-Smirnov, or Shapiro-Wilk tests.

ASQ

Normality tests for Expo data (N=30)

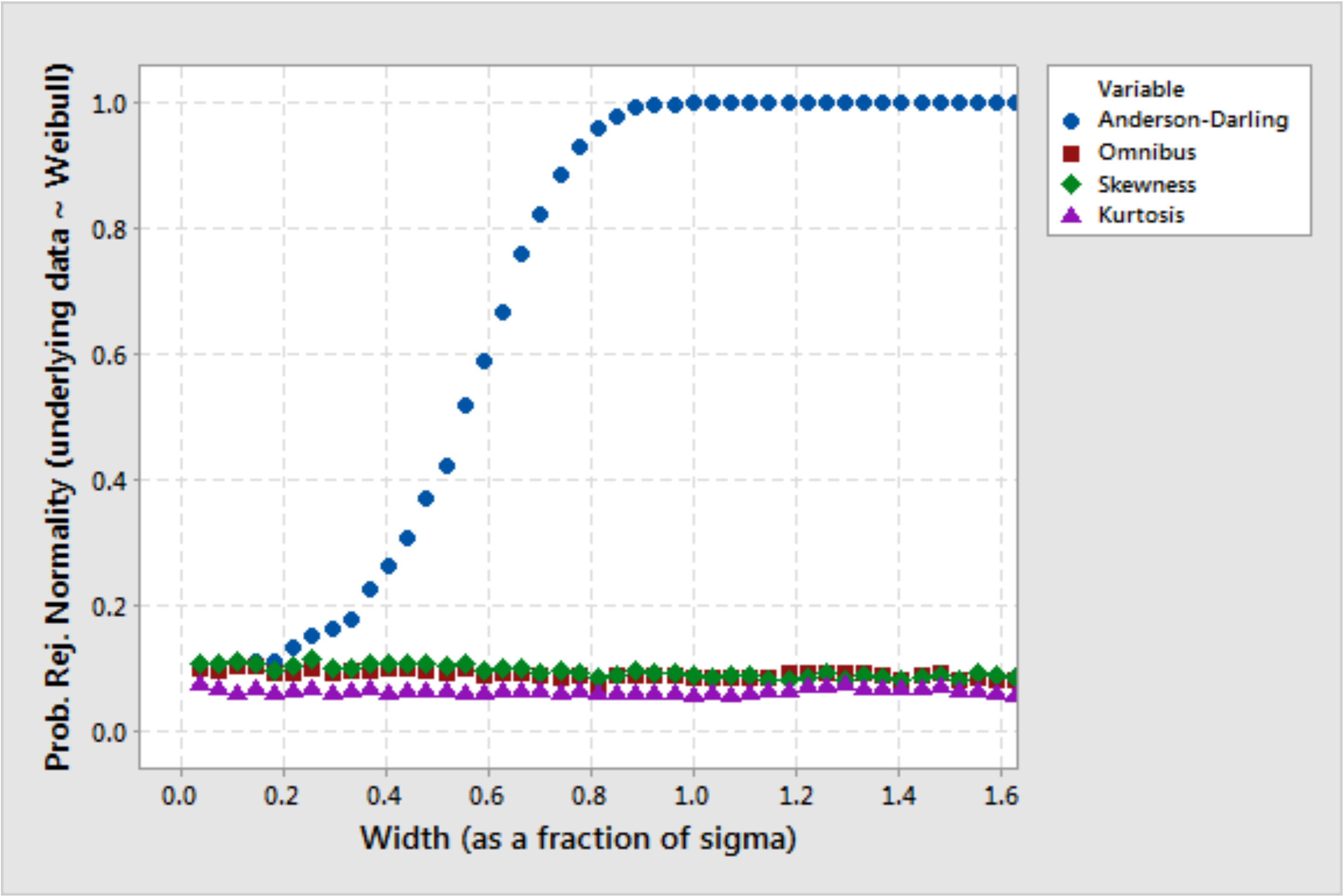


Normality tests performance (Expo)

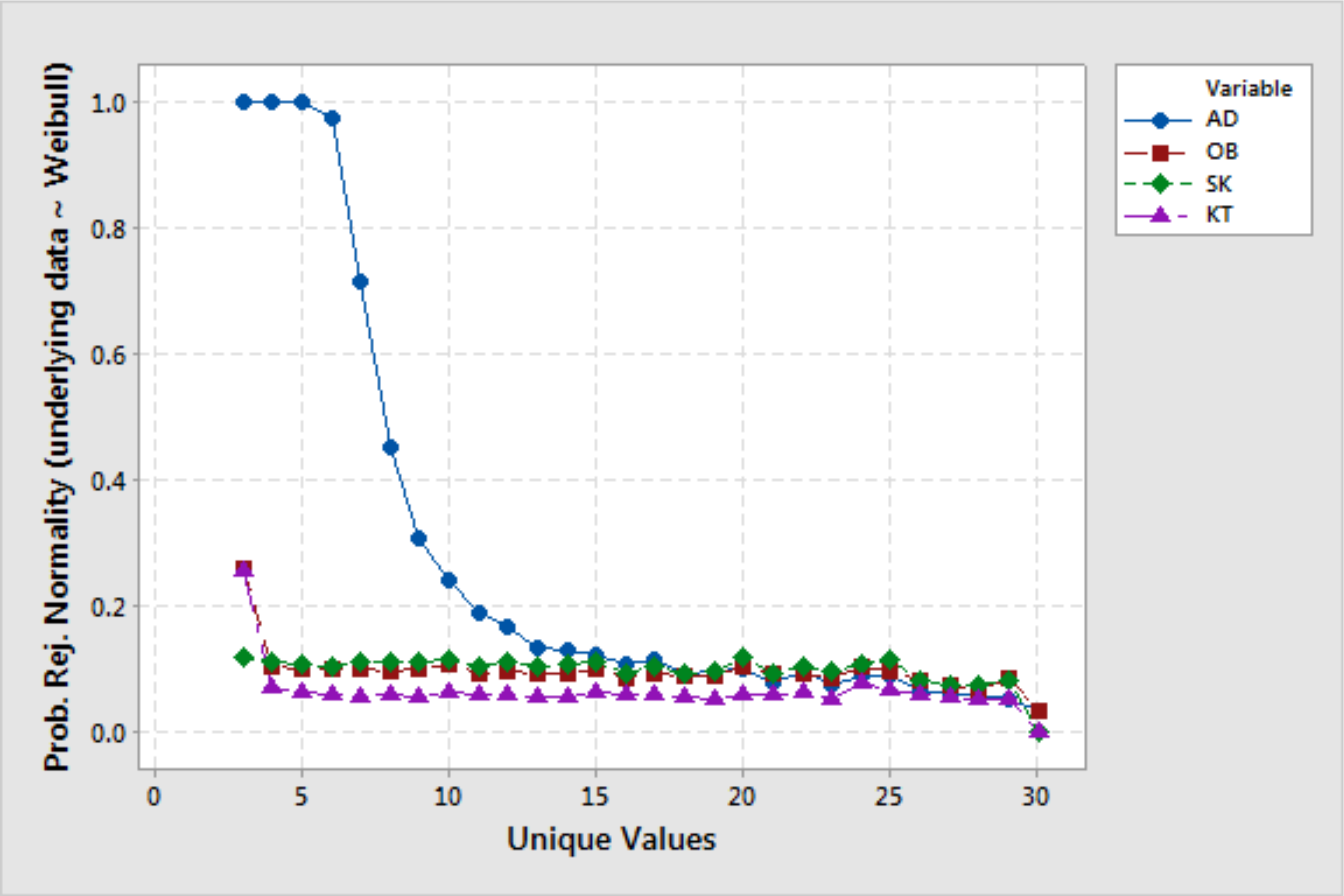


None of the tests would performed as desired. Makes sense since data is approximately symmetric and bell-shaped.

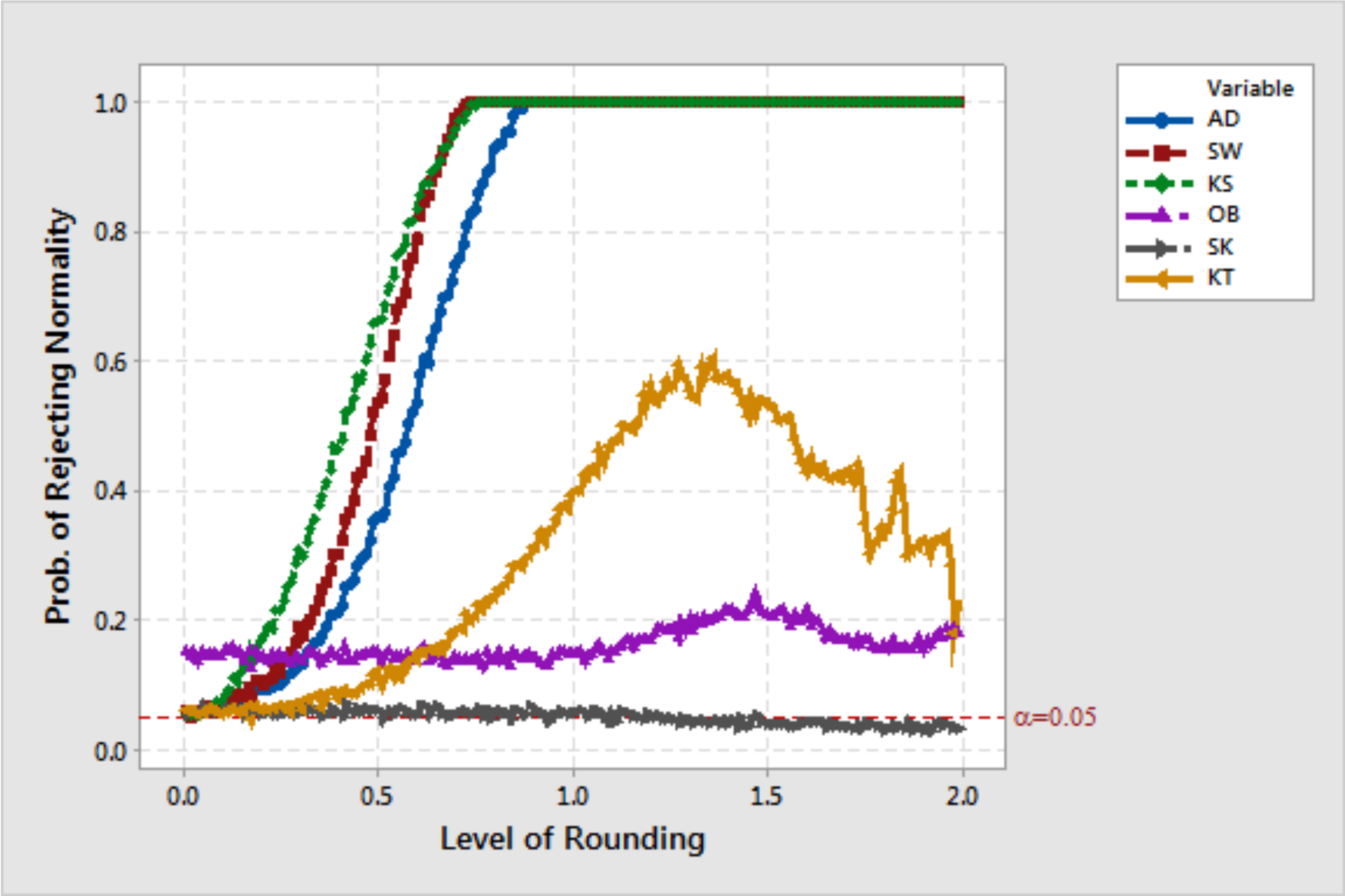
Normality tests for Weibull data (N=30)



Normality tests performance (Weibull)



Comparison of Classic Normality Tests



Case Study

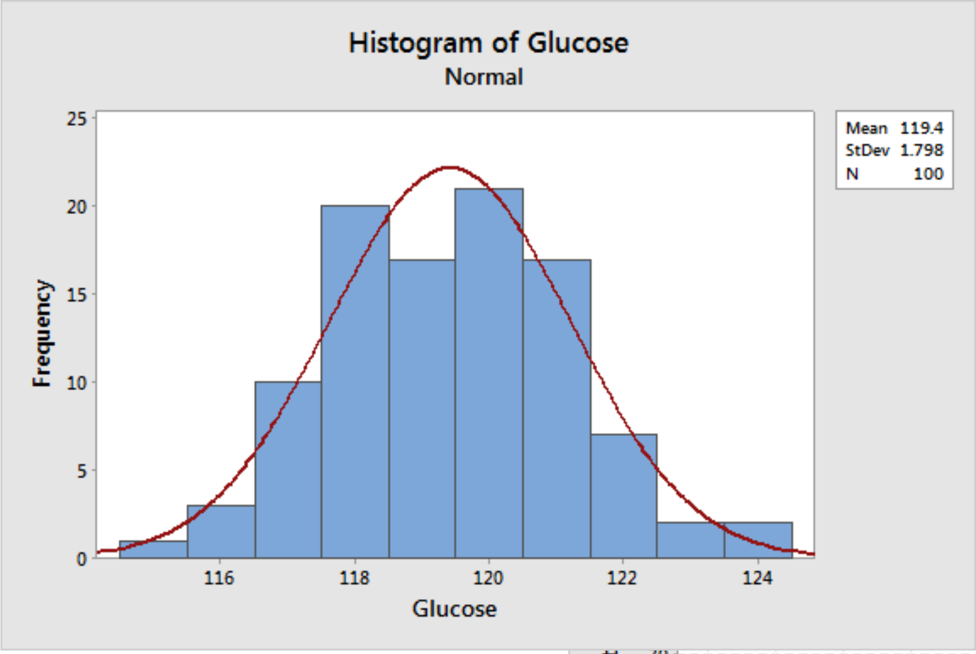
A medical device manufacturer builds a blood glucose measurement apparatus for diabetics to use at home. The reading has to be truncated so that it is easy for the customer to read and understand. They measure a standard solution on 100 devices to set a baseline. The specs are [99, 136].

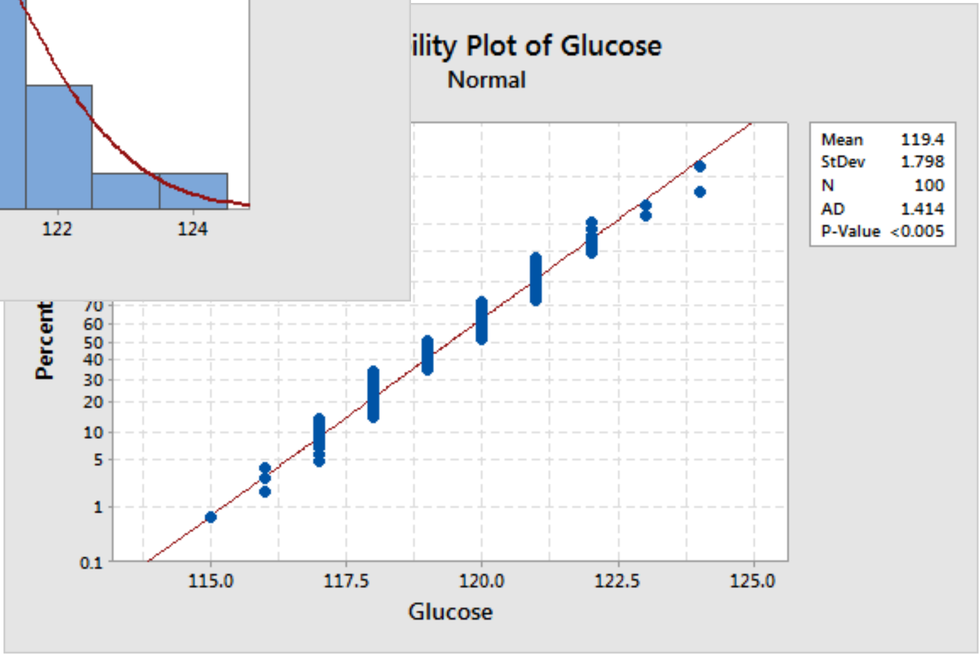
Data

122	121	119	119	123	116	119	120	119	121
118	120	118	120	117	116	120	118	121	120
117	118	119	120	118	118	120	119	120	123
120	117	119	121	120	121	118	117	119	118
120	120	120	122	118	120	117	119	121	117
121	118	117	118	122	119	120	120	120	118
122	119	121	118	118	119	118	121	119	120
116	122	120	117	124	117	120	121	120	115
124	121	118	119	118	121	119	118	122	121
117	118	122	121	121	121	121	119	118	119

Case Study

- Classic Normality tests fail.





Case Study

- Try an alternative normality test instead, such as the Skewness test.

```
Total number of observations in Glucose = 100
```

Data Display

```
Z         0.600407
P-value 0.548235
```

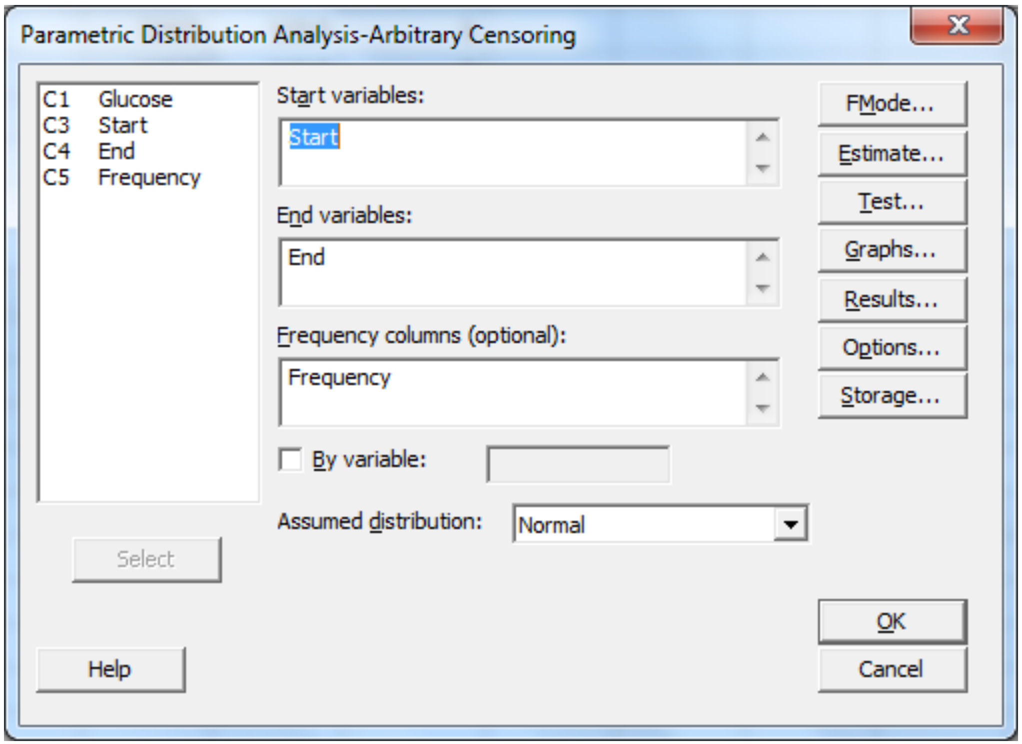
Case Study

- Convert the data to the following format.

Start	End	Frequency
114.5	115.5	1
115.5	116.5	3
116.5	117.5	10
117.5	118.5	20
118.5	119.5	17
119.5	120.5	21
120.5	121.5	17
121.5	122.5	7
122.5	123.5	2
123.5	124.5	2

Case Study

- Treat the data as interval-censored and analyze it with Parametric Distribution Analysis to get the estimates of μ and σ.



Case Study

- Finally, with the estimates of 119.41 for the mean and 1.766 for the standard deviation proceed to estimate Ppk as usual.

$$Ppk = min\left[\frac{USL - \hat{\mu}}{3\hat{\sigma}}, \frac{\hat{\mu} - LSL}{3\hat{\sigma}}\right] = 3.13$$