

# Process Capability Analysis and Process Analytical Technology

Presented by:  
Steven Walfish

President, Statistical Outsourcing Services

[steven@statisticaloutsourcingservices.com](mailto:steven@statisticaloutsourcingservices.com)

<http://www.statisticaloutsourcingservices.com>



# Agenda

- Introduction to Capability
  - What is capability
  - Histograms
  - Normal Distribution
- Capability Indices
  - Cp
  - Cpk
  - Pp
  - Ppk
- Calculating Sigma
  - Relating capability to percent nonconforming
- Capability with attribute data
  - Defects per Million Opportunities (DPMO)
- Process Analytical Technology (PAT)

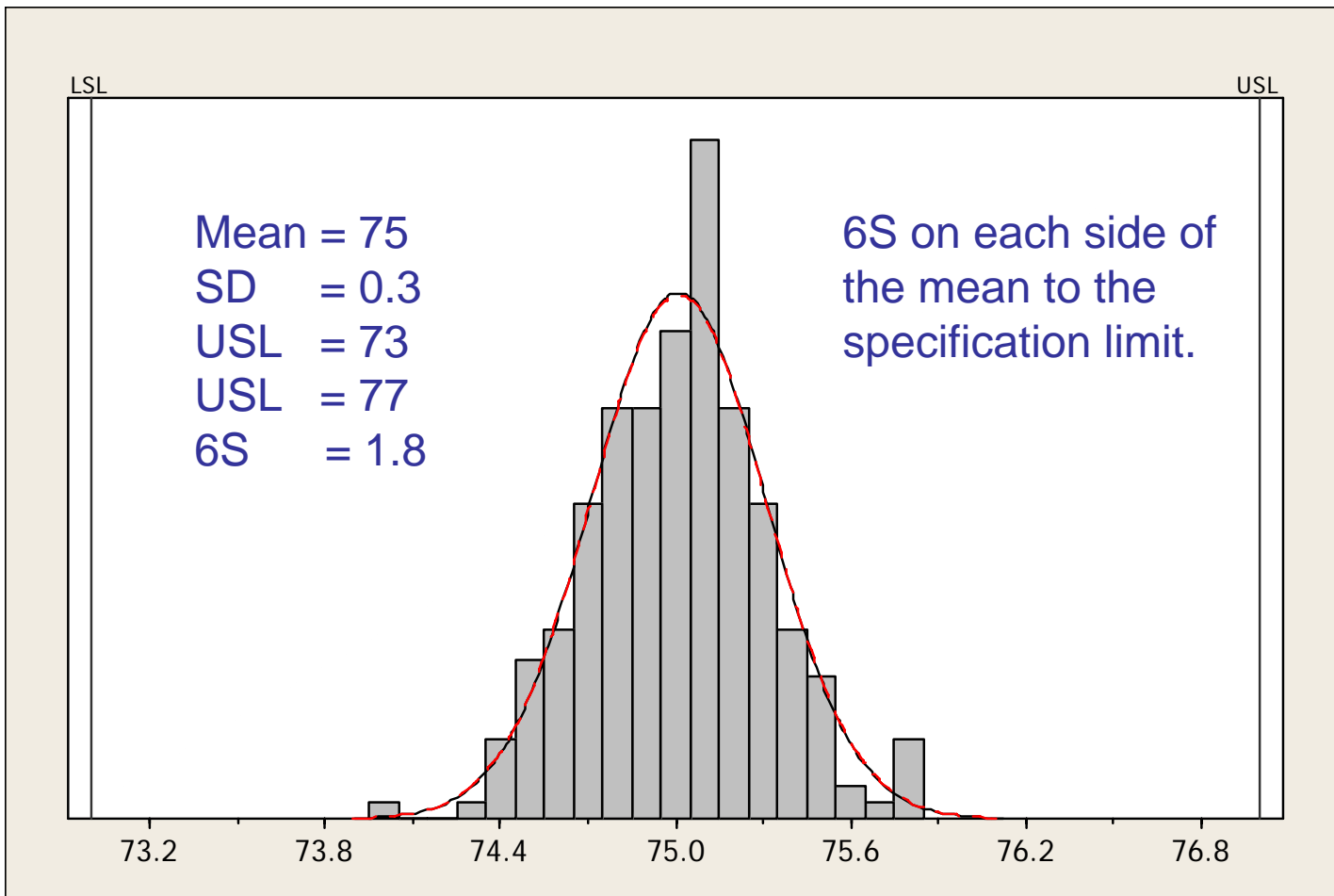


# What is Capability?

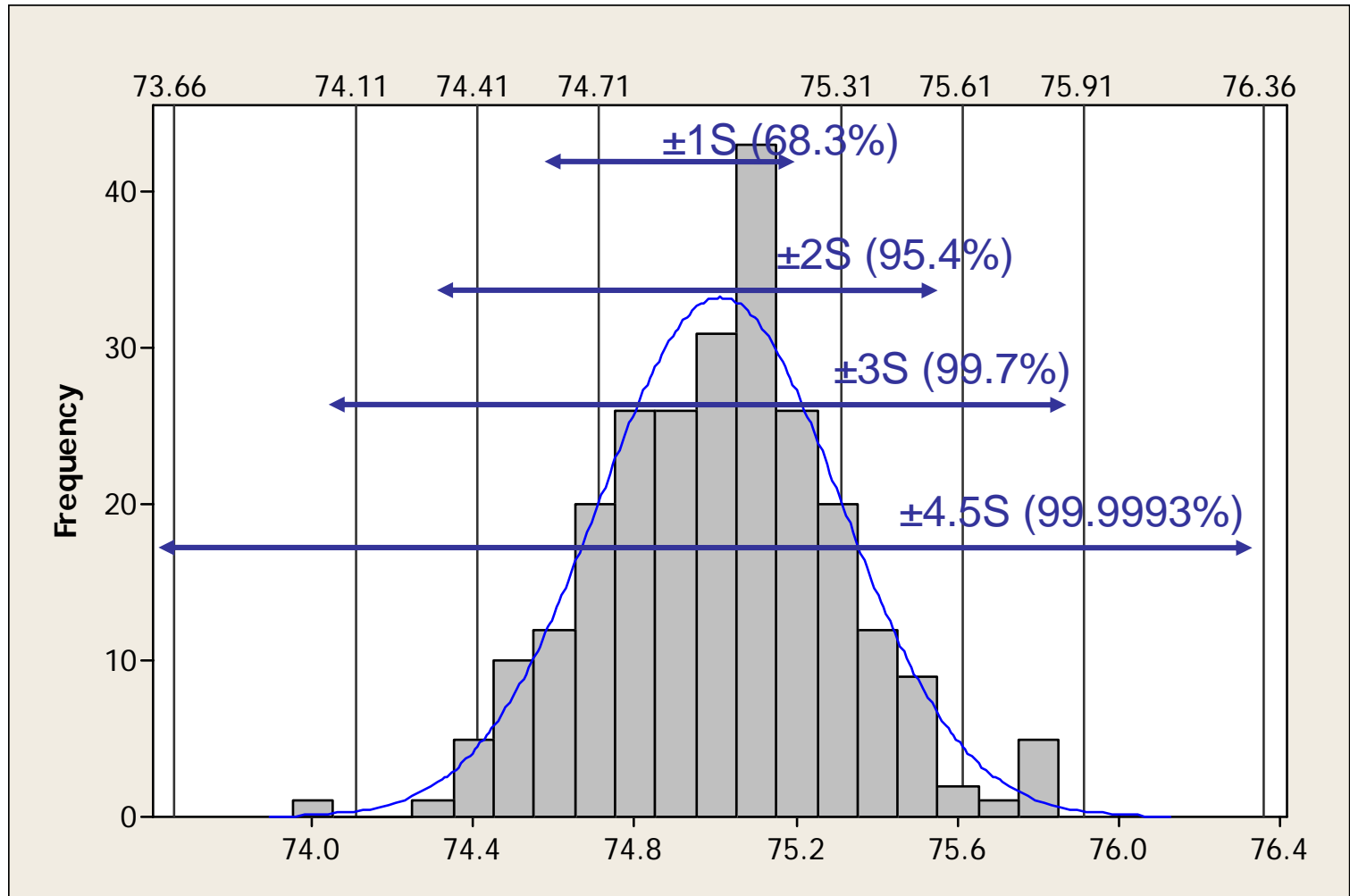
- Process capability compares the output of an *in-control* process to the specification limits.
- The comparison is made by forming the ratio of the spread between the process specifications (the specification "width") to the spread of the process values, as measured by process standard deviation.
- A capable process is one where almost all the measurements fall inside the specification limits.



# Graphical Representation



# Histogram



# Examples of Capability

- Some examples of where capability analysis can be used:
  - Process that is not centered
  - Process with large variability
  - One-sided specifications
  - Setting/confirming customer specifications



# Percent Out of Specifications

- Based on the normal distribution, the percent of product that would fall out of specification can be calculated.
- This is best explained using an example.
- Assume we have a process with mean = 50, standard deviation = 4, USL = 58 and LSL = 46.
- We divide this problem into two parts. First the percent out of specification on the high end (greater than the USL) and then the percent out on the low end (less than the LSL).



# The Normal Distribution

- The normal distribution is:

$$Z = \frac{USL - \bar{X}}{S}; \frac{\bar{X} - LSL}{S}$$

- Z is the number of standard deviations that the specification is from the mean.
- Normal probability tables give you the percent of the distribution that would exceed the specification limit for a given z value
- Remember that 68.3% of the data is within  $\pm 1S$  (therefore 31.7% is outside of  $\pm 1S$ ).



# Out of Specification Calculations

$$Z = \frac{USL - \bar{X}}{S}; \frac{\bar{X} - LSL}{S}$$

$$Z = \frac{58 - 50}{4}; \frac{50 - 46}{4}$$

$Z = 2$  for the upper specification and 1 for the lower specification

- A  $z = 2$  is 2.28% out of specification;  $z=1$  is 15.9%.
- The total percent expected to be out of specification would be 18.1%
- We will discuss summary statistics for this later.



# Estimating Sigma

- There are several methods for estimating sigma ( $S$ ) used in capability analysis.
  - Control charts
    - $R\bar{c}$
    - $S\bar{c}$
    - Moving Range
    - MSSD
  - Pooled standard deviation
  - Total standard deviation (Long-Term)



# Short-Term

- Statistical Process Control methods such as control charting provide estimates for short term variability.
- Short term variability is defined as the average within subgroup variability.
- For subgroup sizes greater than 1, based on the distribution of the range, the average range within a subgroup can be divided by a constant  $d_2$ . The constant is based on the sample size of the subgroup.
- Similarly, the average standard deviation can be divided by  $c_4$ . The constant  $c_4$  is also based on the sample size of the subgroup.



# Short Term for n=1

- When the subgroup size is a single observation, there are two methods to estimate the sigma; moving range or mean squared successive differences.
- The moving range, like the average range for subgroup size greater than 1 uses the constant  $d_2$ . Here the value of  $d_2$  equals 1.128 and the average range is the average of the range of successive points.
- A variation of the moving range is the mean squared successive difference (MSSD).

$$\frac{\sqrt{\frac{1}{2} * \frac{(\sum d_i^2)}{(n-1)}}}{c_4}$$

# Long Term

- Total variability can be estimated from the entire data set.
- Total variability is estimated by treating the data as one big sample using only the overall mean and looking at how the data points vary around this one overall mean.
- Total variability estimates the long term state of the process variability.
- If a process is stable, then the variability seen short term is consistent with what you expect to see long term.



# Assumptions

- There are two critical assumptions to consider when performing process capability analyses with continuous data, namely:
  - The process is in statistical control.
  - The distribution of the process considered is Normal.
- If these assumptions are not met, the resulting statistics may be highly unreliable.
- In a later modules we will discuss capability analysis for non-normal data.



# Capability Indices

- There are several statistics that can be used to measure the capability of a process:  $C_p$ ,  $C_{pk}$ ,  $P_p$  and  $P_{pk}$ .
- The statistics assume that the population of data values is normally distributed.
- Variability can be stated as either short-term or long-term.
- $C_p$  and  $C_{pk}$  are based on short term variability
- $P_p$  and  $P_{pk}$  are based on total variability



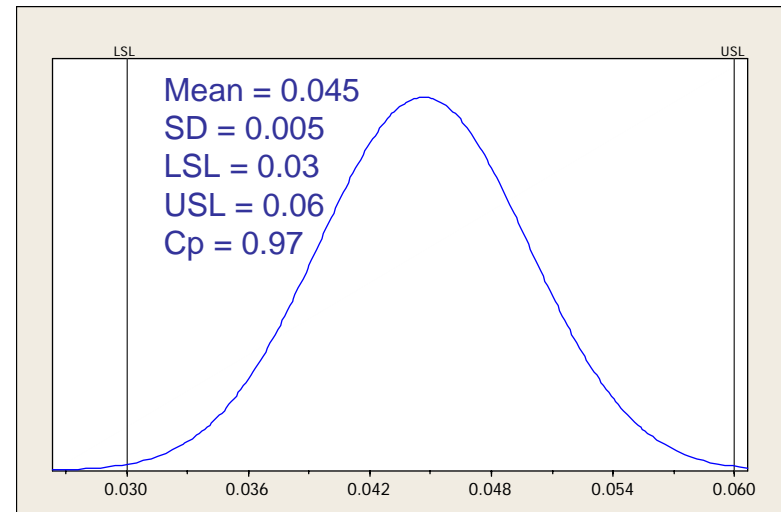
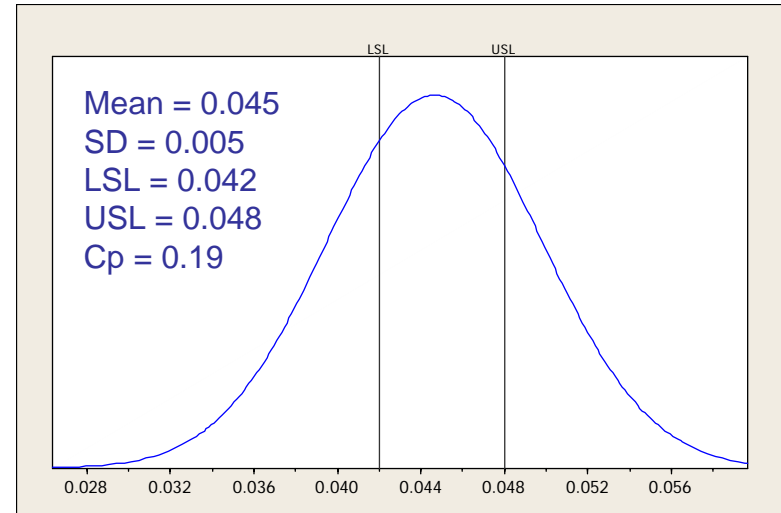
# C<sub>p</sub>

- Approximately 99.7% of the data from a normal distribution is contained between  $\pm 3\sigma$ .
- If the process is in control and the distribution is well within the specification limits then the difference between the Upper specification (U) and Lower specification (L) should be larger than  $6\sigma$ .
- If the specifications are larger than  $6\sigma$ , the ratio will be less than 1.
- If  $C_p$  is greater than 1 then the process has the potential to meet specifications as long as the mean is centered.



# Cp

$$C_p = \frac{\text{UpperSpec} - \text{LowerSpec}}{6 \cdot S}$$



# Cpk

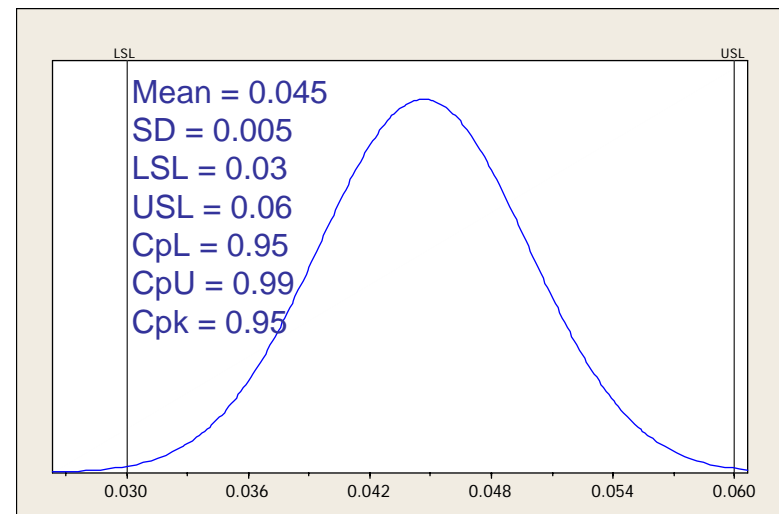
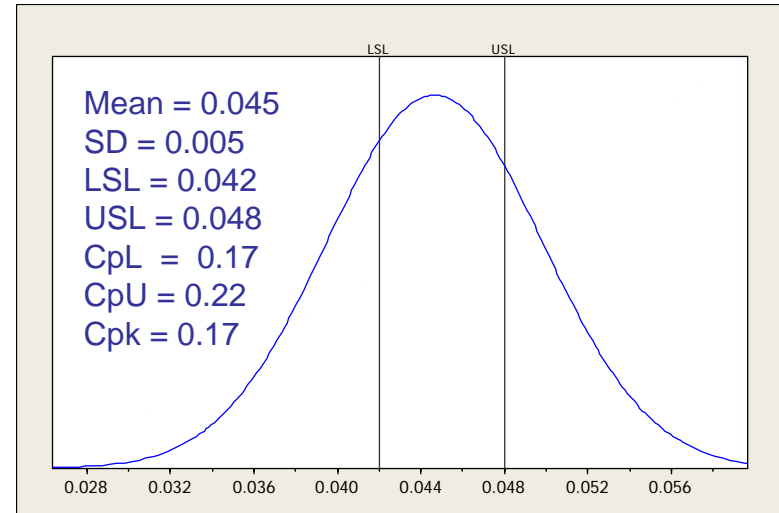
- Cpk is an process capability index that assesses how close the process mean is from the specification limit.
- If the process is in control and the distribution is well within the specification limits then the difference between the Upper specification (U) and then mean or the difference between the Lower specification (L) and the mean should be larger than  $3\sigma$ .
- If  $C_{pk}$  is greater than 1 then the process mean is sufficiently far from the specification limit.

# Cpk

$$C_{pU} = \frac{UpperSpec - \bar{\bar{X}}}{3 \cdot S}$$

$$C_{pL} = \frac{\bar{\bar{X}} - LowerSpec}{3 \cdot S}$$

$$C_{pk} = \min(C_{pL}, C_{pU})$$



# Cpk

- $C_{pk}$  greater than 1 shows the process is probably centered and usually able to meet specifications
- $C_{pk}$  less than 1 indicates either the mean is not centered between the specifications or there is problem with variability
- $C_{pk}$  is meant to be used with processes that are in control – gives us a measure of whether the in-control process is capable of meeting specifications
- $C_{pk}$  is not an appropriate measure if there are trends, runs, out-of-control observations or if the process is too variable

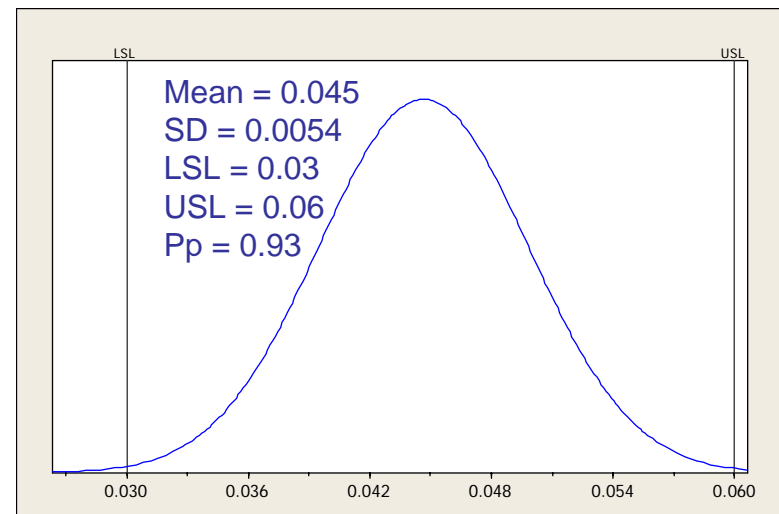
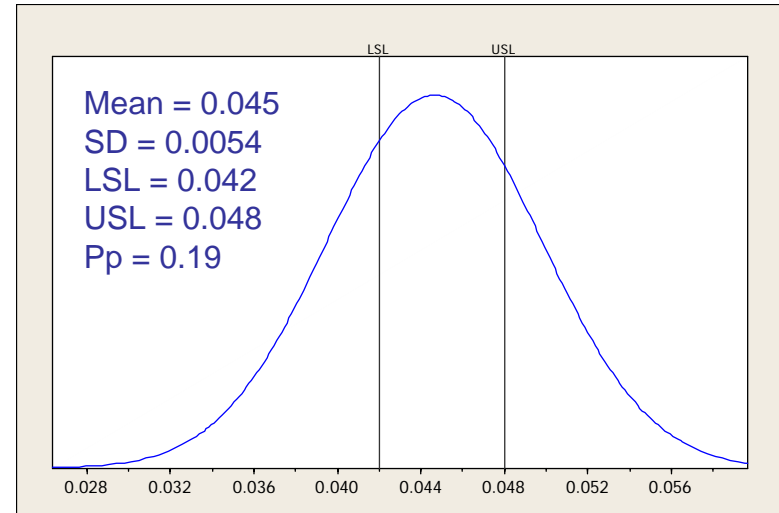


# Pp

- Pp is an overall capability similar to Cp.
- Total variability is used in the denominator instead of the short term.
- If the process is stable and in control the estimate of Pp is similar to the estimate of Cp.
- If  $P_p$  is greater than 1 then the process is meeting the specifications as long as the mean is centered.

# Pp

$$P_p = \frac{\text{UpperSpec} - \text{LowerSpec}}{6 \cdot S}$$



# Ppk

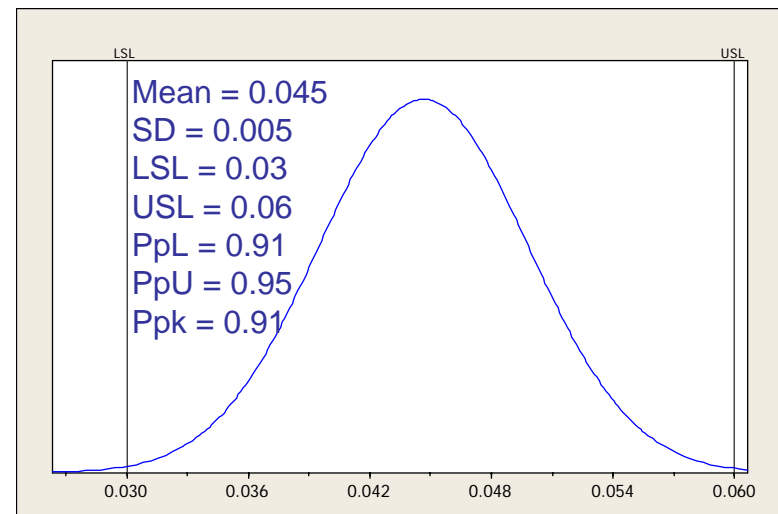
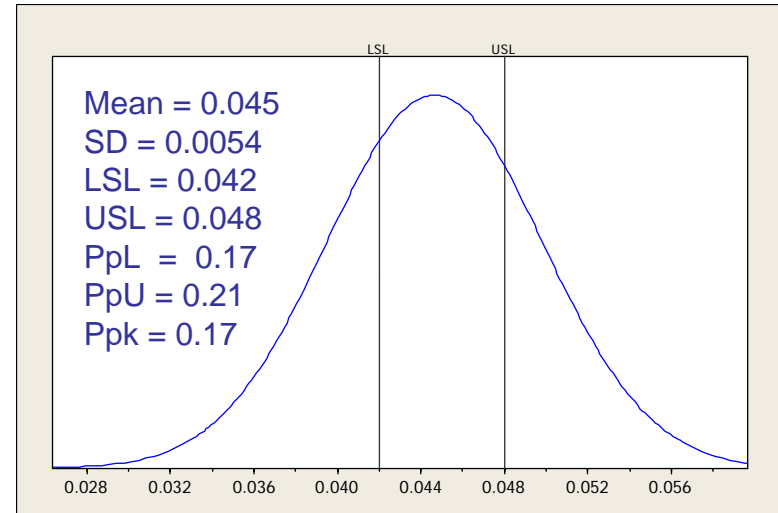
- Ppk is an process capability index that assesses how close the process mean is from the specification limit.
- Total variability is used in the denominator instead of the short term.
- If the process is in control and the distribution is well within the specification limits then the difference between the Upper specification (U) and then mean or the difference between the Lower specification (L) and the mean should be larger than  $3\sigma$ .
- If  $P_{pk}$  is greater than 1 then the process mean is sufficiently far from the specification limit.

# Ppk

$$P_{pL} = \frac{\bar{\bar{X}} - \text{LowerSpec}}{3 \cdot S}$$

$$P_{pU} = \frac{\text{UpperSpec} - \bar{\bar{X}}}{3 \cdot S}$$

$$P_{pk} = \min(P_{pL}, P_{pU})$$





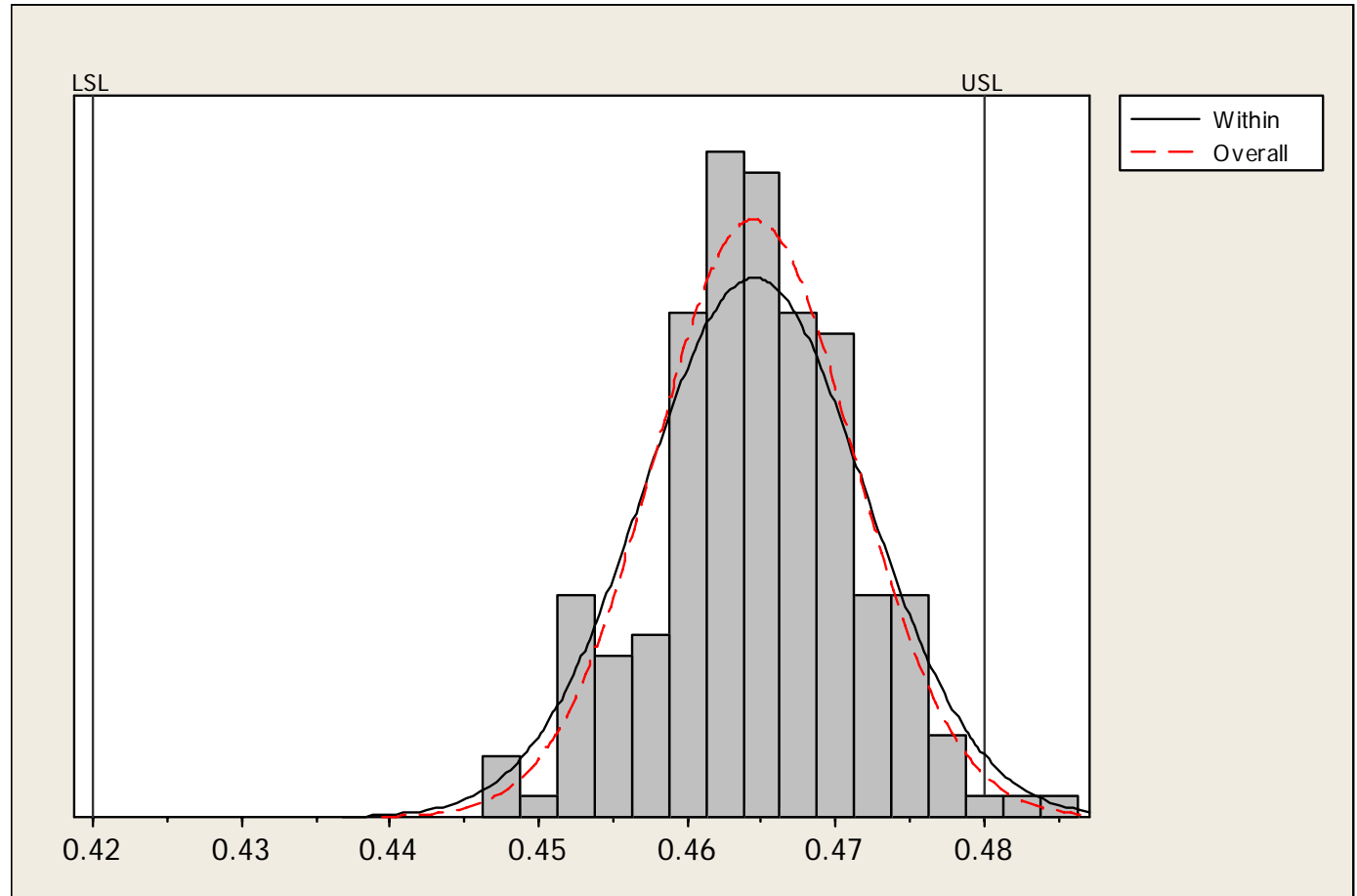
# Example

- The concentration after fermentation is critical to downstream processing.
- Tolerances are  $0.45 \mu\text{g}/\text{mL} \pm 0.03$
- Low concentration would lead to insufficient product and too high a concentration would lead to loading issues.
- A capability study was performed, with the following results.
  - Mean = 0.465
  - Standard Deviation (Short Term) = 0.0075
  - Standard Deviation (Total) = 0.0067



# Example – Continued

- $C_p = 1.34$
- $C_{pL} = 1.99$
- $C_{pU} = 0.69$
- $C_{pk} = 0.69$
- $P_p = 1.49$
- $P_{pL} = 2.20$
- $P_{pU} = 0.77$
- $P_{pk} = 0.77$



# Attribute Data

- When examination of an item or event results in a **PASS** or **FAIL** rather than a measurement, the capability analysis must be based on a discrete distribution.
- When the measure is counts or proportions, the binomial is used to estimate capability.
- When the relevant measure of performance is a rate, then the capability analysis is based on the Poisson distribution.



# DPMO

- A capability index for an attribute process, especially those that have multiple inspections per part is to calculate Defects per Million Opportunities (DPMO).
- Defects per Million Opportunities (DPMO) =  
(Number of Defects Generated /  
Number of Opportunities for Defects in Sample) x  $10^6$
- If there are 2 defects generated in a sample of 100 with each unit having 5 areas inspected  
 $DPMO = (2 / (5 * 100)) * 10^6 = 4000$

# Binomial Data

- First plot your data on a p-chart. This allows you to assess if the process is in statistical control. Remove any points outside the limits and recalculate the average percent defective.
- Using the overall percent defective, calculate the upper and lower 95% confidence interval on percent defective.
- Multiply the percent defective by 1,000,000 to get the DPMO.
- Calculate Process Z by determining the Z-value that corresponds to the percent defective. A higher Process Z is desirable.

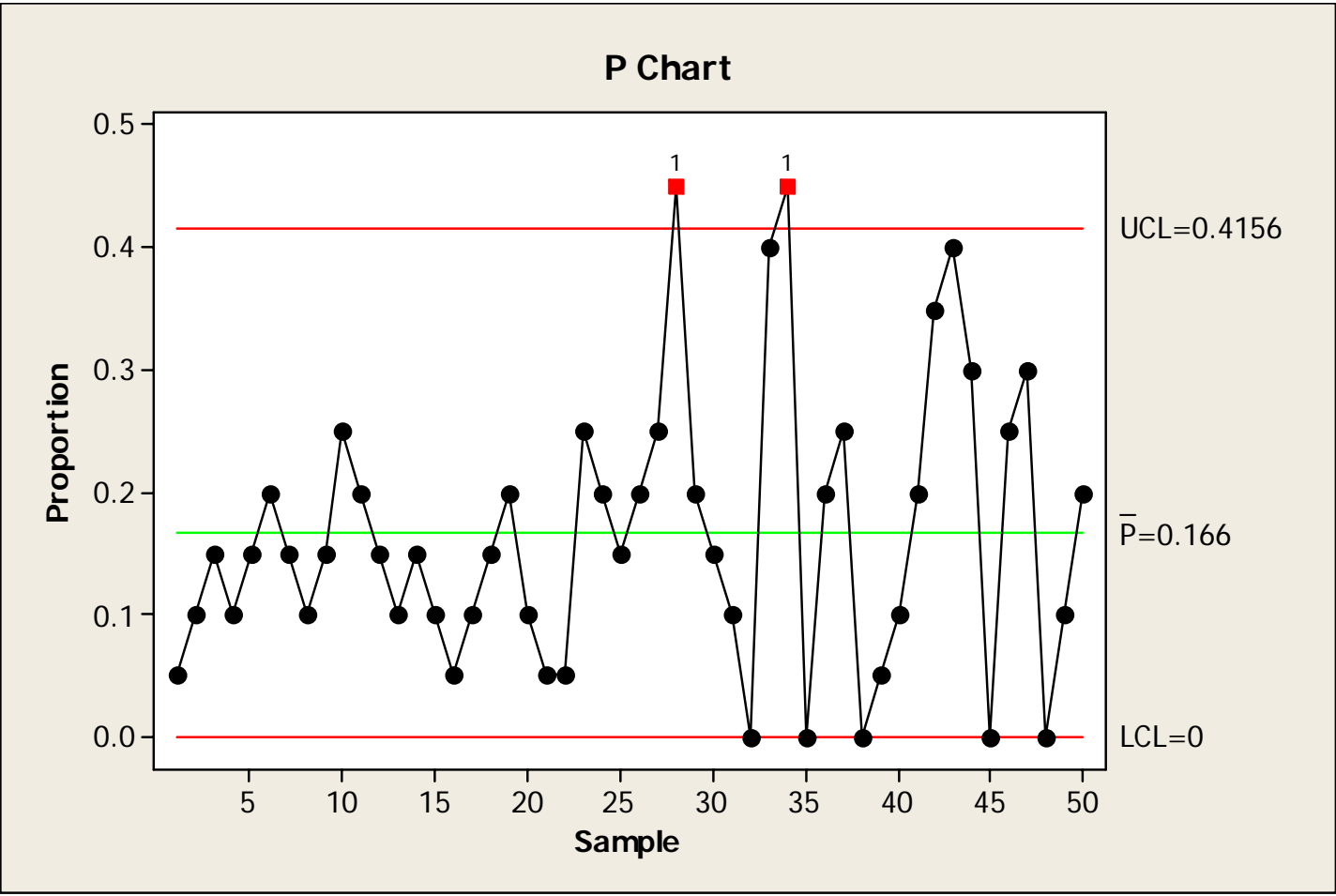


# Example

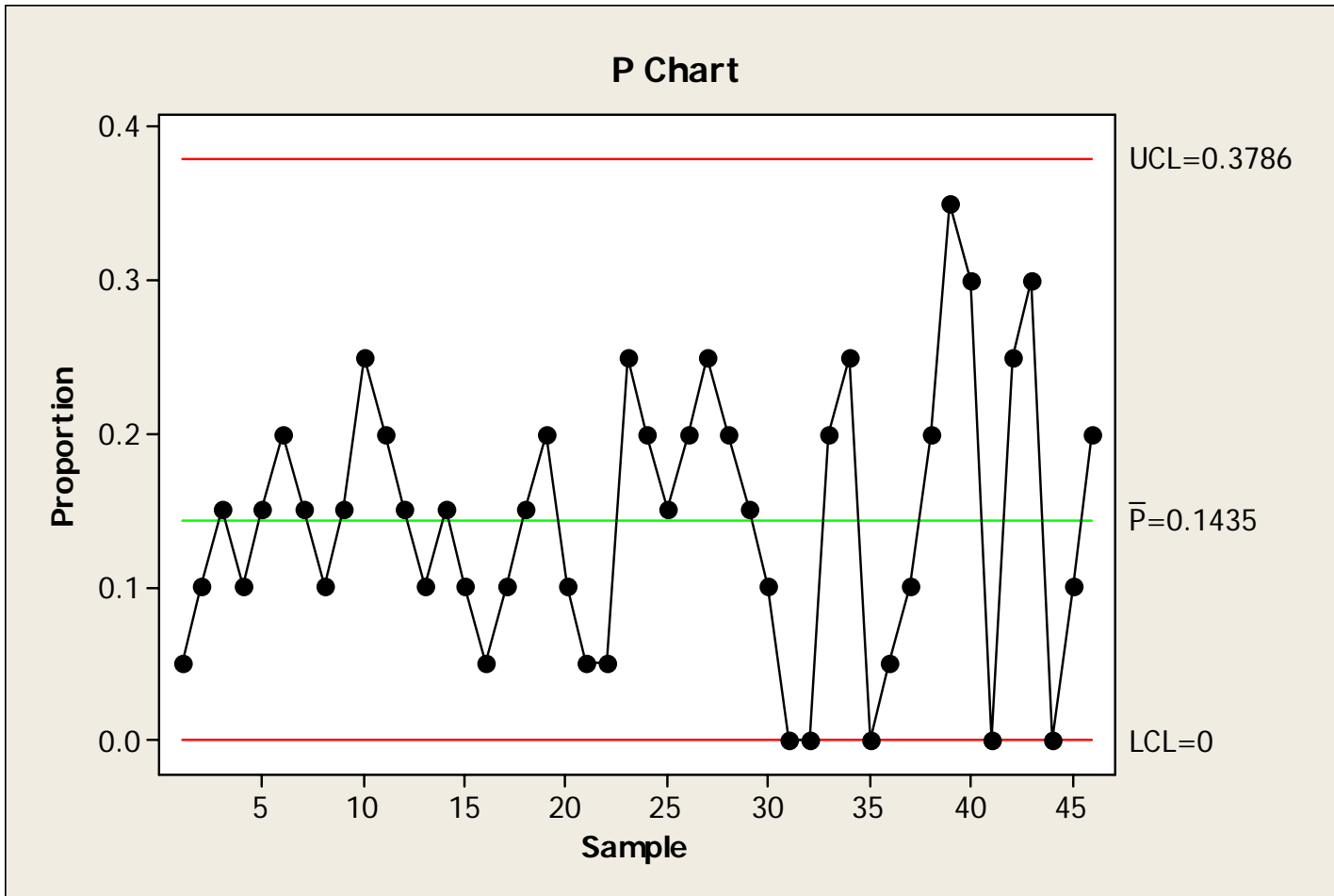
- The following are the number of defectives found during the inspection of 20 units in each lot.

1	4	1	2	4
2	3	1	0	7
3	2	5	8	8
2	3	4	9	6
3	2	3	0	0
4	1	4	4	5
3	2	5	5	6
2	3	9	0	0
3	4	4	1	2
5	2	3	2	4

# P-Chart



# Updated P-Chart

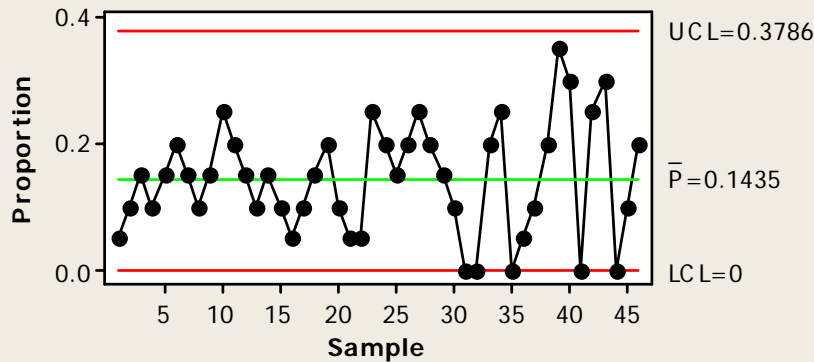




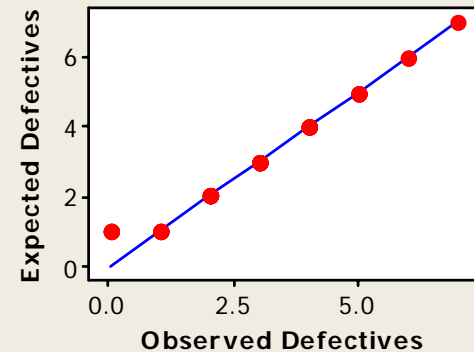
# Summary Statistics

## Binomial Process Capability Analysis

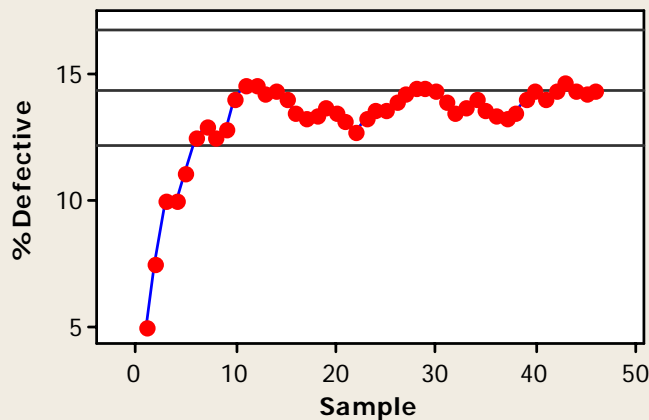
P Chart



Binomial Plot

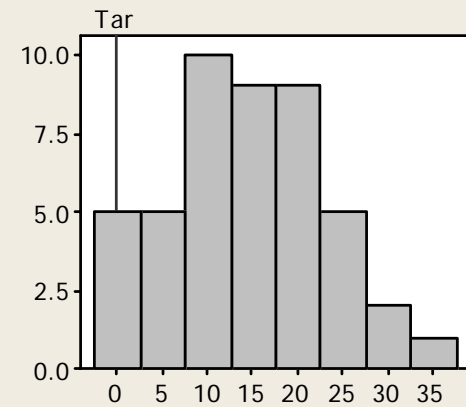


Cumulative % Defective



Summary Stats	
(using 95.0% confidence)	
% Defective:	14.35
Lower C I:	12.15
Upper C I:	16.78
Target:	0.00
PPM Def:	143478
Lower C I:	121452
Upper C I:	167814
Process Z:	1.0648
Lower C I:	0.9628
Upper C I:	1.1678

Dist of % Defective



# Poisson Data

- First plot your data on a u-chart. This allows you to assess if the process is in statistical control. Remove any points outside the limits and recalculate the average defects per unit (DFU).
- Using the overall defects per unit, calculate the upper and lower 95% confidence interval on percent defective.

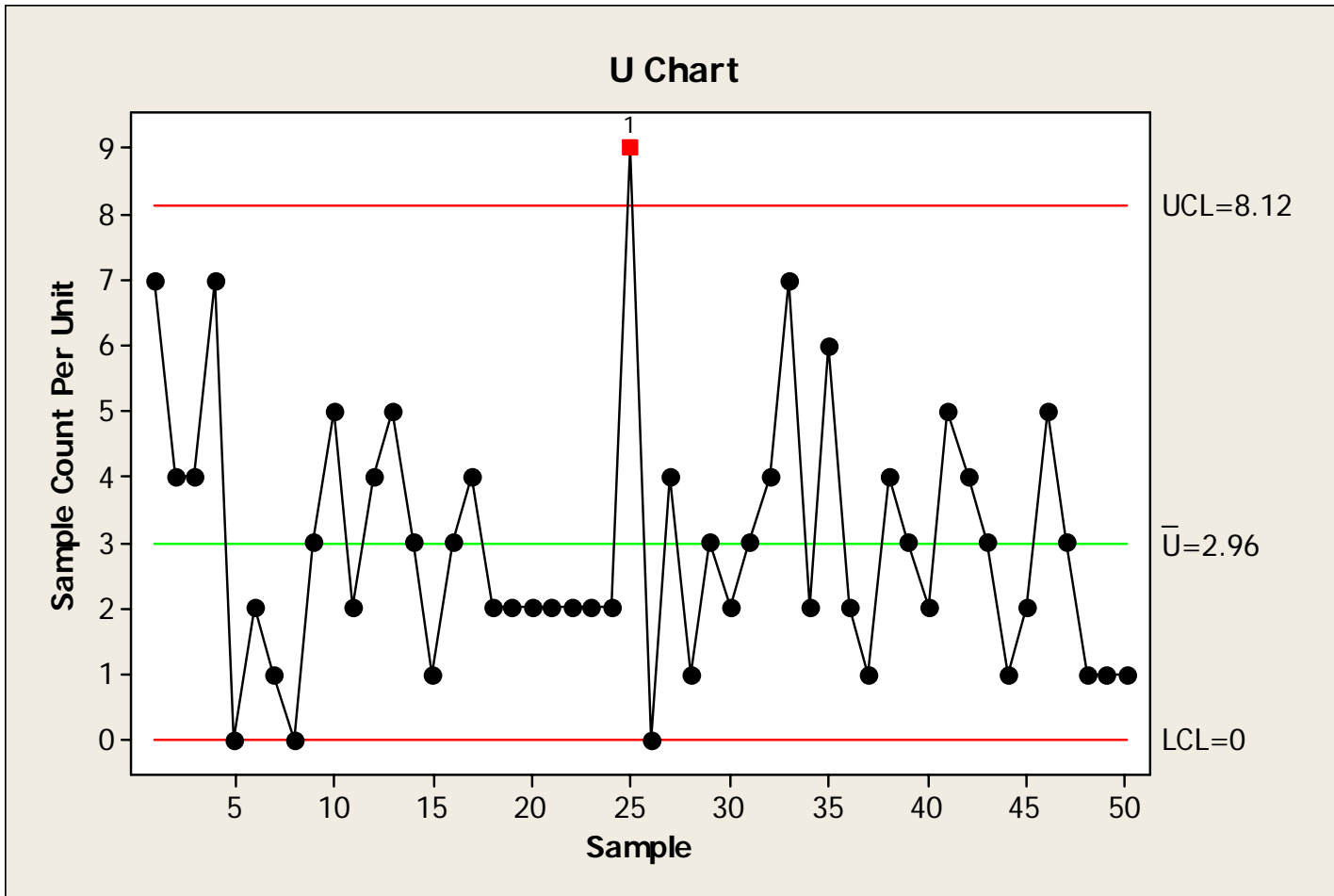


# Example

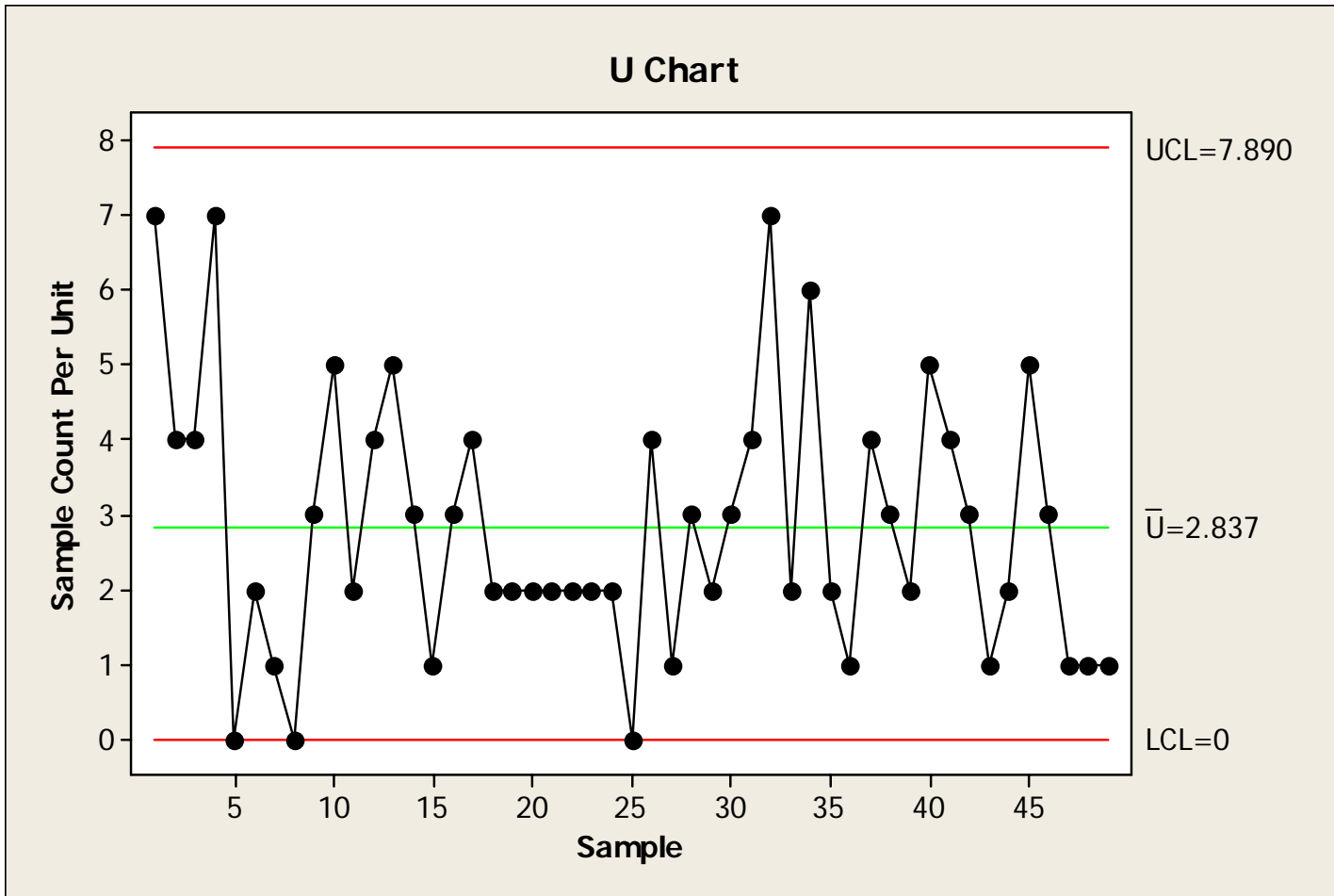
- The following are the number of defectives found per unit during the inspection of 50 units.

7	2	2	3	5
4	4	2	4	4
4	5	2	7	3
7	3	2	2	1
0	1	9	6	2
2	3	0	2	5
1	4	4	1	3
0	2	1	4	1
3	2	3	3	1
5	2	2	2	1

# U-Chart

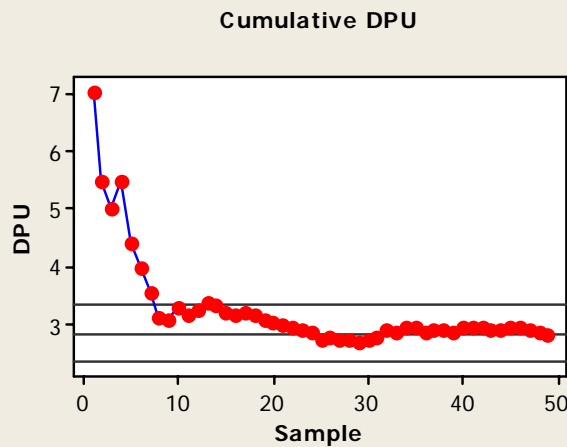
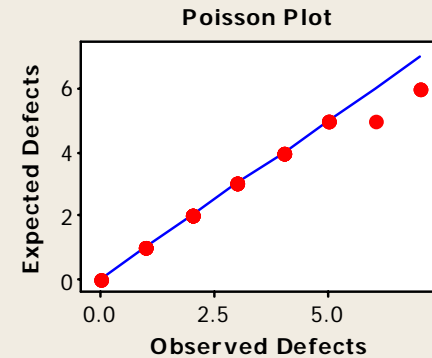
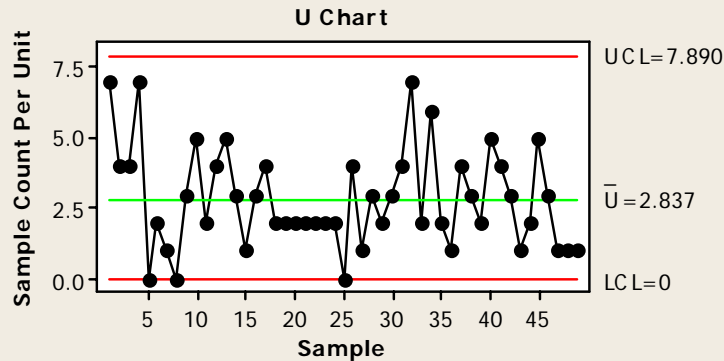


# Updated P-Chart

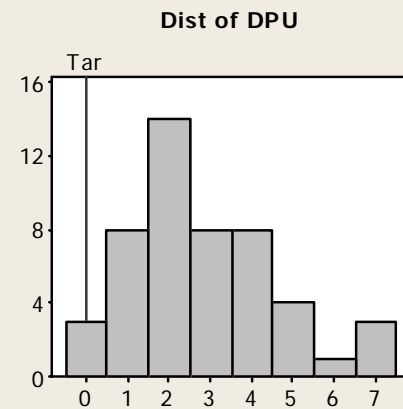


# Summary Statistics

## Poisson Capability Analysis



Summary Stats	
(using 95.0% confidence)	
Mean Def:	2.8367
Lower CI:	2.3848
Upper CI:	3.3494
Mean DPU:	2.8367
Lower CI:	2.3848
Upper CI:	3.3494
Min DPU:	0.0000
Max DPU:	7.0000
Targ DPU:	0.0000

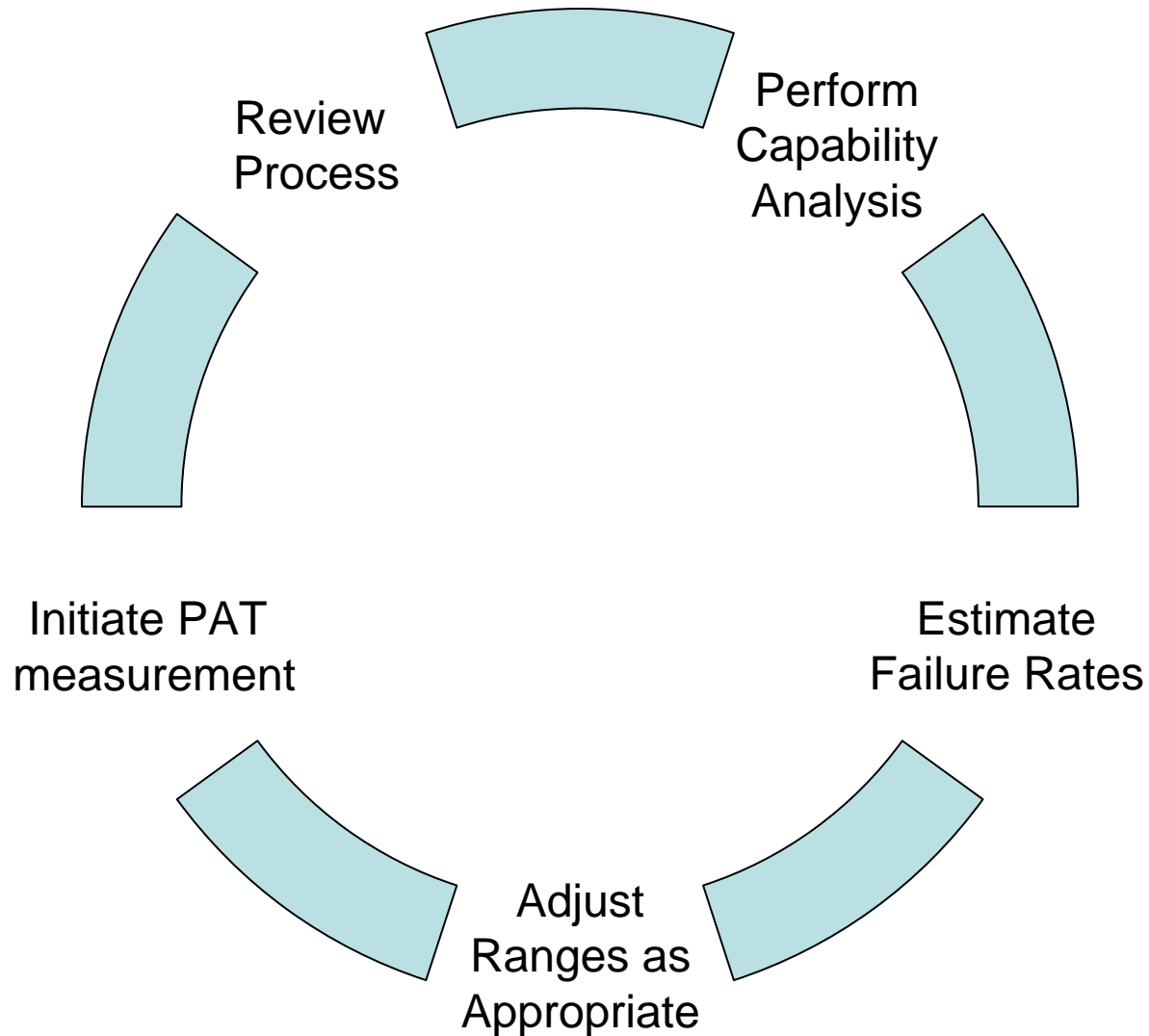


# Process Analytical Technology

- The goal of PAT is to achieve sufficient process understanding and control to enable quality assurance in “real-time”.
- When a process is capable and repeatable, this can be achieved.
- When there is too much variability, then further process understanding is required.



# PAT Lifecycle





# Conclusions

- Process capability analysis can be predictive of expected out of specification results.
- The process must be in statistical control prior to doing a capability analysis.
- Data transformation can be used for non-normal data.
- Attribute data can use confidence intervals to predict process performance.



# Questions?



# Statistical Outsourcing Services

- Statistical Outsourcing Services can provide:
  - Bioassay Analysis
  - Acceptance Sampling
  - Method Validation
  - Process Validation
  - Stability Testing
  - Comparability Studies
  - Quality Systems
  - Statistical Process Control (SPC)
  - Third Party Expert Review
  - Clinical Protocol Review
  - Business Intelligence (BI)
- The following is a sample list of available training programs:
  - Introduction to Statistics
  - Statistics for Non-Statisticians
  - Using Statistical Software
  - Statistical Process Control (SPC)
  - Regression Analysis
  - Design of Experiments
  - Method Validation

