

Statistical Analysis Prior to, During and After Process Validation

Overview

- Design of Experiments
- Process Capability
- Confidence Intervals
- Normal Distribution
- ANOVA

Analysis Progression

- Success / failure
 - One run, no factors varied, one outcome, yes/no
 - Lack of comparison, inefficient
- One-Factor-at-a-Time (OFAT)
 - Several runs, one factor varied, two outcomes
 - Can't find interactions and is inefficient

Analysis Progression

- Multivariate
 - Design Of Experiments (DOE) is used for multivariate data analysis
 - Is much more efficient
 - Comparison of different outcomes to evaluate the inputs

Multivariate Initiative

- Gather background information
- Define response to be measured
- List the potentially important CPPs
- Can I reduce the list to the up most critical
- Can I define a high and low level for each of the factors

Confounding Rules

- We need at least one more data value than the number of factors or effects we want to estimate
 - Estimate 7 effects in 8 runs
 - Estimate 15 effects in 16 runs
 - Estimate 99 effects in 100 runs

Main Effects Coding Table

Run #	Factors			Coded Factors			Cure Time
	Temp.	Filler	Resin	T	F	R	
1	90	A	1	+	-	-	57
2	65	B	1	-	+	-	75
3	65	A	2	-	-	+	42
4	90	B	2	+	+	+	28

Effects Coding Table

- Coded columns are the design table
- If we look at the plus and minus signs row wise, they tell us how to design and conduct the experiment
- If we look at the signs column wise, they tell us how to calculate the effects

Variability

- A process is considered well-understood when
 - All critical sources of variability are identified and explained (DOE)
 - Variability is managed by the process
 - Process and endpoint monitoring and control tools

Process Understanding

- Quality attributes can be accurately and reliably predicted
 - Ultimately, the ability to accurately and reliably predict product CQA (via mathematical models) reflects a high degree of process understanding
 - This level of process understanding is inversely proportional to the risk of producing a poor quality product

Sources of Variability

- Optimize CPPs
- Outputs have random variation
 - Conditions must be identical
- To isolate & identify particular causes of variability requires special experimental design and analysis

Sources of Variability

- Limiting variation will tighten U/LSL's
- Variability reduction adds value
 - increases process capability
 - $X \pm 3SD$ (normal distribution curve)
 - CpK (centering)
- Limiting variation minimizes the risk of deviations & OOS

Analysis of Variation

- Analysis of Variance Tools
 - Process capability
 - Confidence intervals
 - Normal distribution curves
 - ANOVA

General Terms

- Precision
 - Does not take into consideration target value (SD, RSD)
- Standard Deviation
 - Spread of group of individual observations

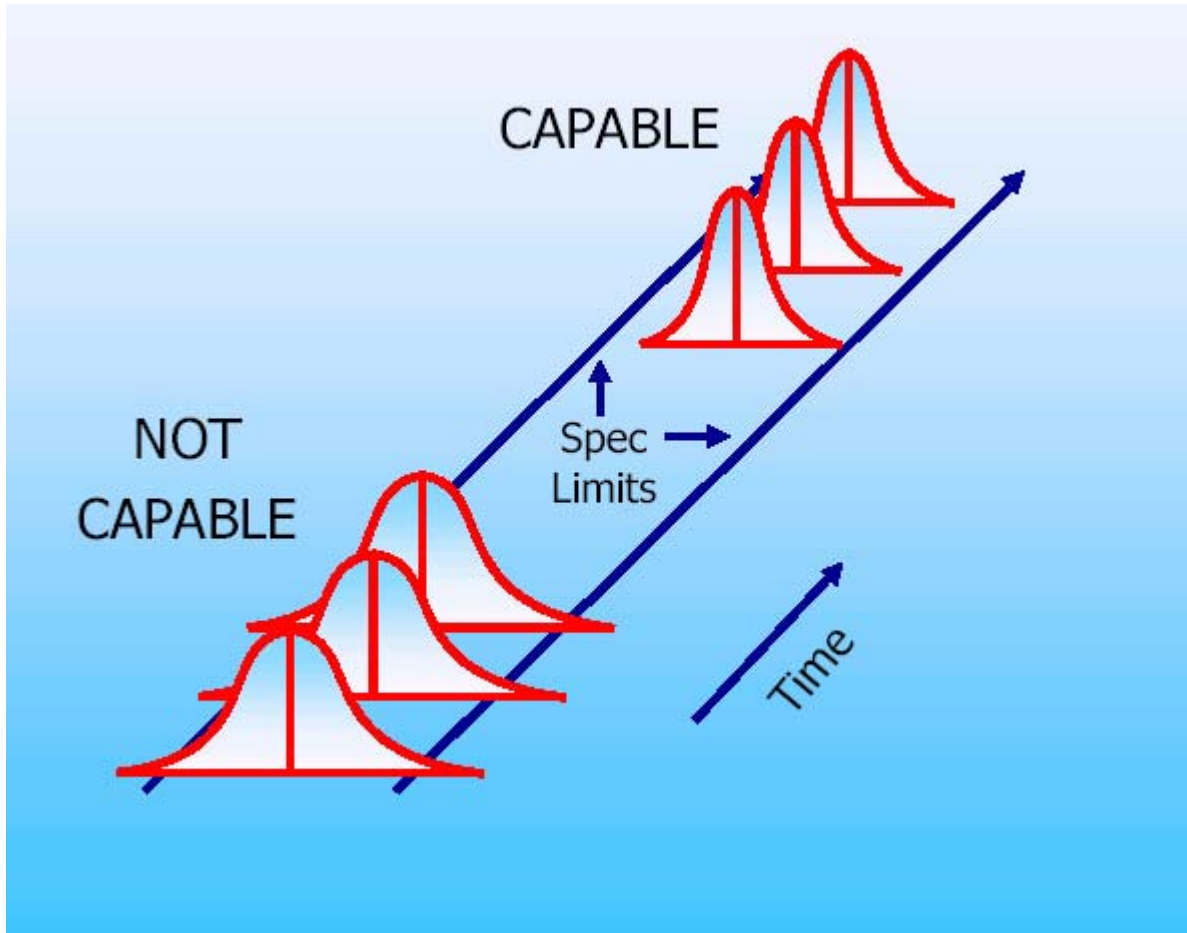
General Terms

- Relative Standard Deviation
 - The relative standard deviation is a measure of precision, calculated by dividing the standard deviation for a series of measurements by the average measurement

Process Capability

- Cpk
 - Capability index which accounts for process centering
 - $(USL - \text{Mean}) / 3 \sigma$ **or** $(\text{Mean} - LSL) / 3 \sigma$
- σ = process capability
- Value of 1.33 is the goal. 1.00 minimum standard

Process Capability



Confidence Intervals

- It is possible to use your sample to calculate a range within which the population value is likely to fall
 - "Likely" is usually taken to be "95% of the time," and the range is called the **95% confidence interval**
- The values at each end of the interval are called the **confidence limits**
 - All the values between the confidence limits make up the confidence interval
 - You can use *interval* and *limits* almost interchangeably

Confidence Intervals

- Assumes only one true value, and that the confidence interval defines the range where it's most likely to be
- The confidence interval is NOT the variability of the true value or of any other values between subjects. It is nothing like a standard deviation

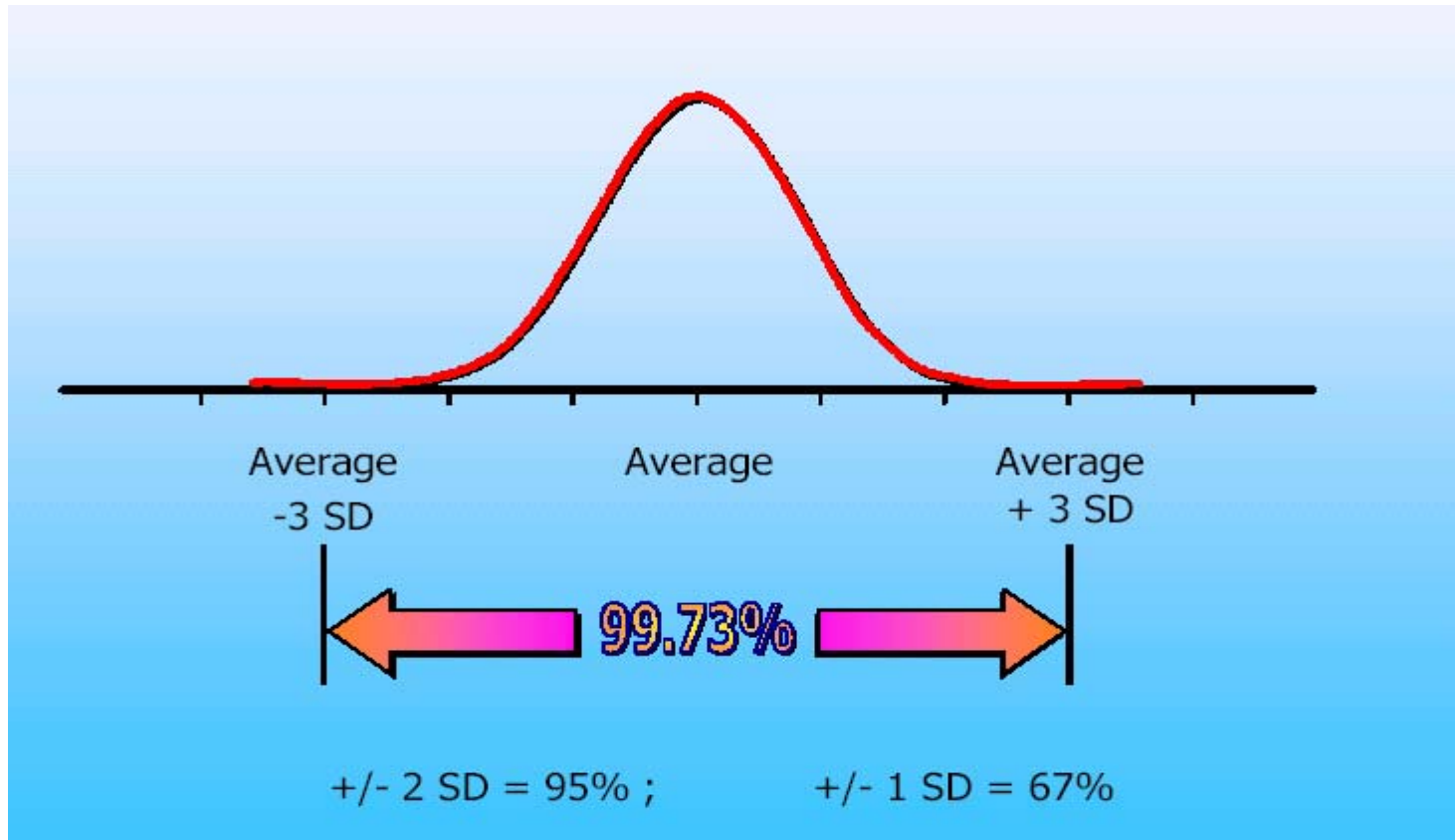
Confidence Intervals

- A confidence interval is an interval within which we believe the true mean lies
- When intervals are constructed, we are 95% confident the interval contains the true mean

Normal Distribution

- Model is a response
- The response should be the target value + random error (normally distributed)
- Errors must be normal with constant variance (trend analysis)
- Errors are independent

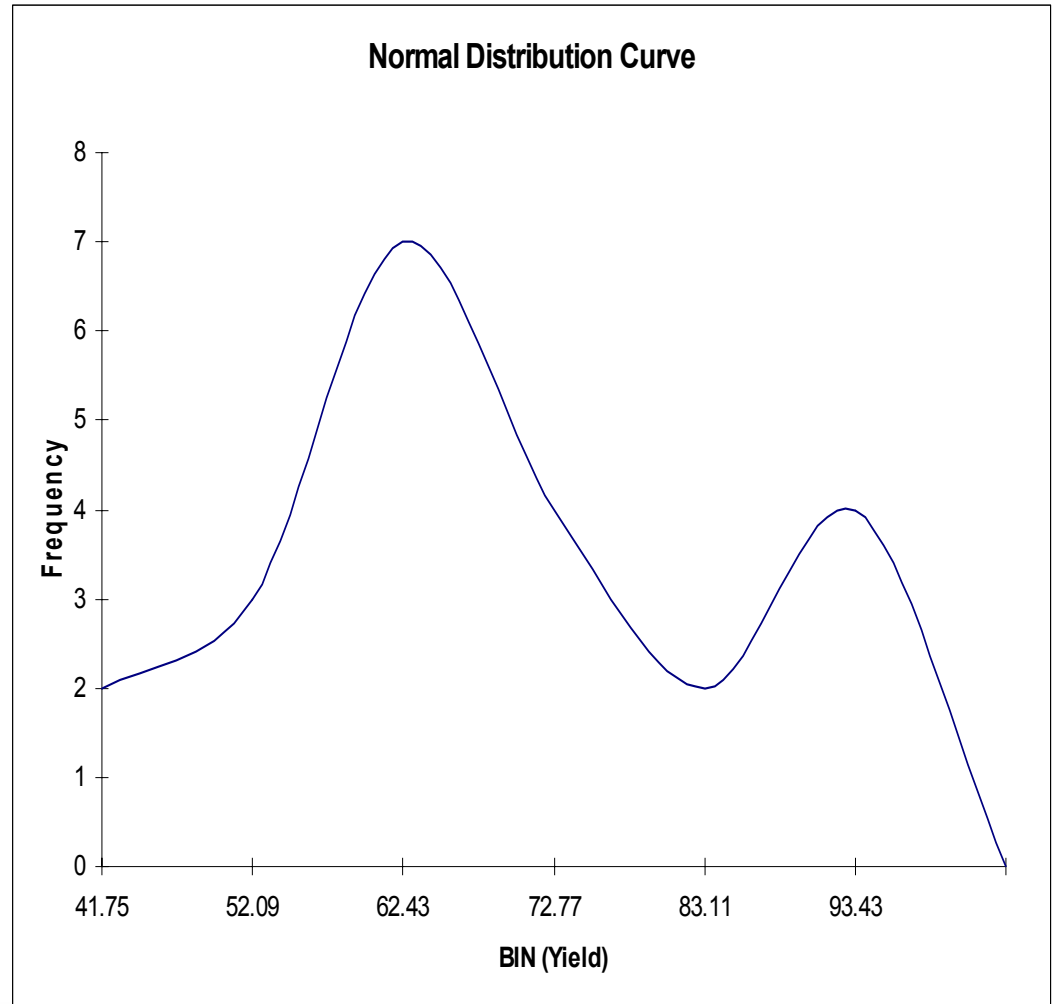
Normal Distribution Curve (Population)



Source: Process Validation Guidance, GHTF, 1999

Normal Distribution

- Evaluation: Two distinct populations (Bi-modal)
- Possible reasons:
 - Sampling Methods
 - Different Inputs
 - CPPs
 - Different equipment
- Possible Solutions:
 - Try to segregate 2 populations
 - 2 different shifts
 - 2 different equipment
 - 2 different operators



Analyze Mean Variance

- Analyze 2 means
 - Student T-Test
 - Z Test
- Analyze 3 or more means
 - ANOVA is a general method of analyzing data from designed experiments to test the hypothesis that means from two or more samples are equal (drawn from populations with the same mean).

One-Way ANOVA

- Used for a single response
- Used when we wish to test the equality of means in experiments where two or more means are randomly assigned to different, independently experimental units
 - Eg. Blend homogeneity at top, middle and bottom of blender
 - Eg. 3 batches from process validation

ANOVA

- ANOVA is concerned with differences between *means* of groups, not differences between *variances*

Thank-You

