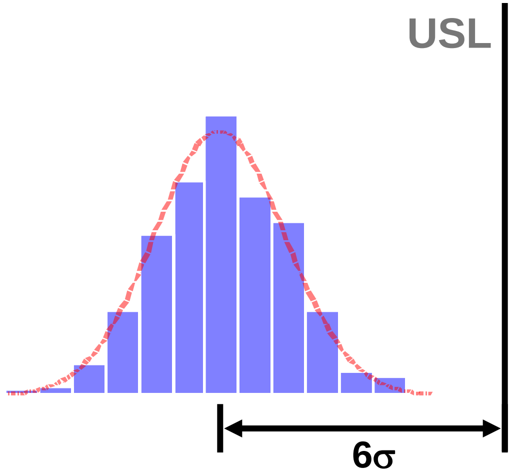
Basic Statistics
Basic Graphs



Objectives

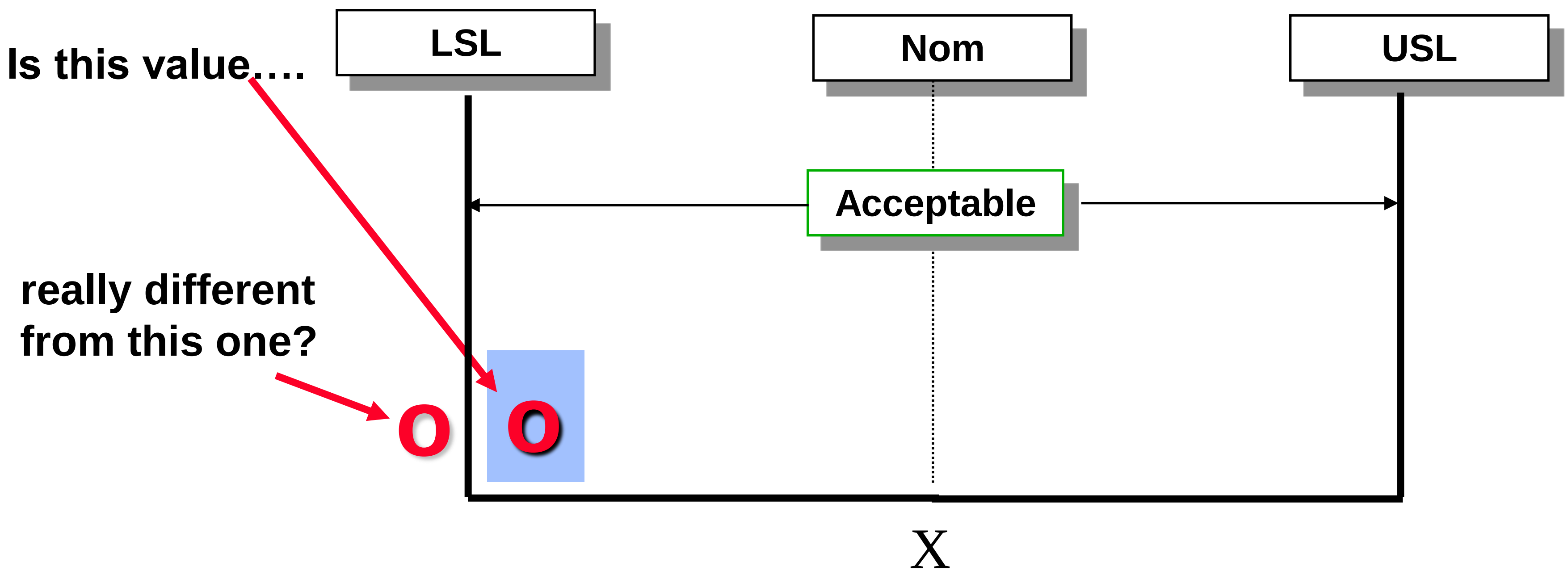
At the end of this module, the participant will be able to :

- Use graphs to visually represent variation
- Assess normality using the probability plot
- Describe a variable's distribution using graphs as well as shape, center, and spread metrics
- Use the normal distribution to calculate the probability of an outcome

Traditional View of Performance

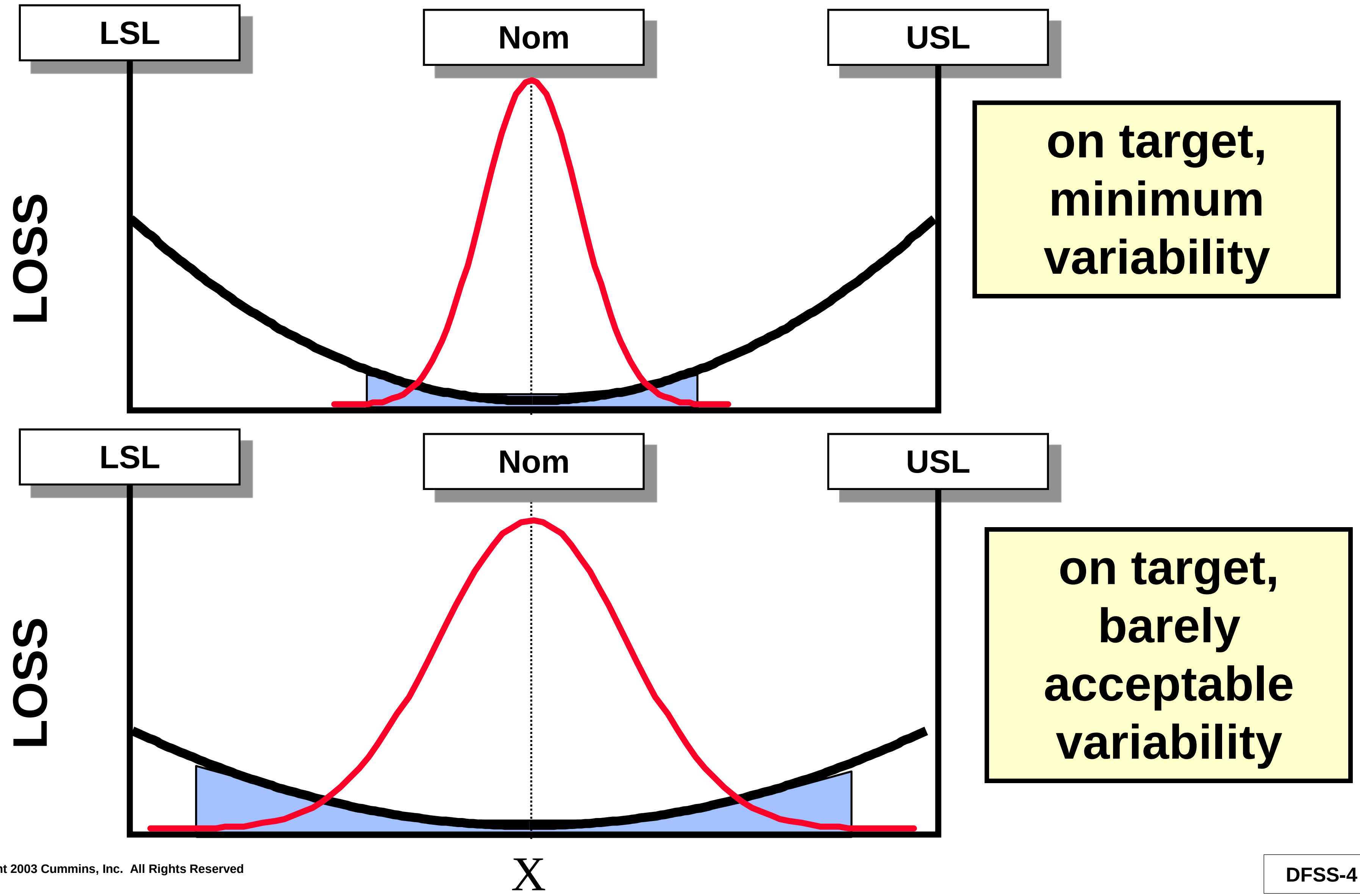
Our primary concern <u>WILL NO LONGER BE</u>
"Are we in spec?"

Traditional View

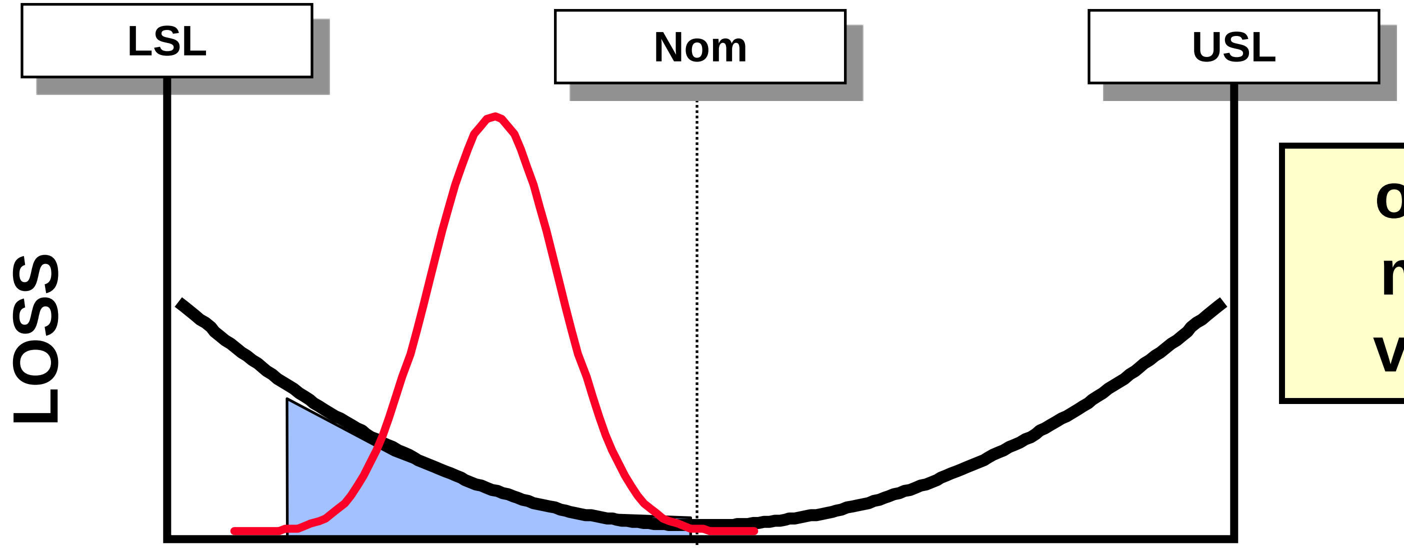


Yet we are ready to use one and throw the other out!

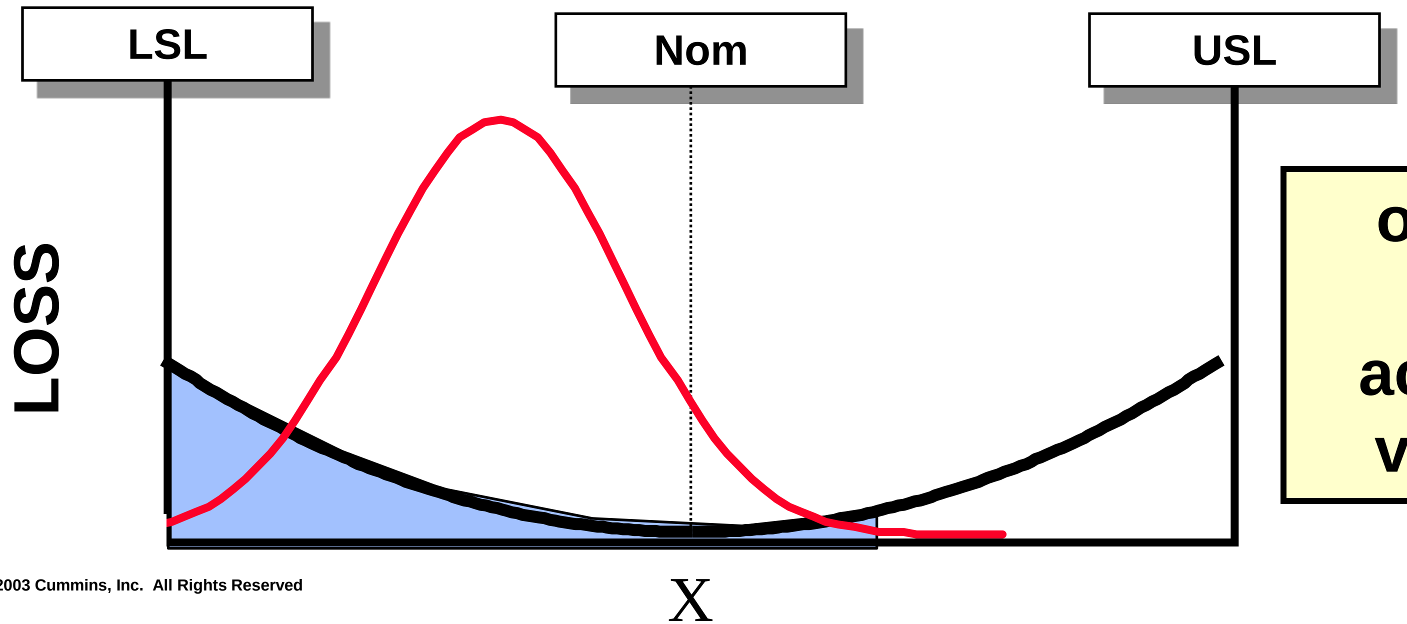
Cost of Variation when On Target



Cost of Variation when Off Target



off target, minimum variability



off target, barely acceptable variability

How to Assess Performance

- From now on, we will be concerned with the center and spread of the critical parameters … not only whether it's in spec or out of spec!

- We will assess product performance by
  - studying how the mean and variance (voice of the design) compare to the specification limits (voice of the customer).
    - use the mean of the distribution to determine if the response is on target & has sensitivity to noise
    - use the standard deviation of the distribution to determine variability & sensitivity to noise

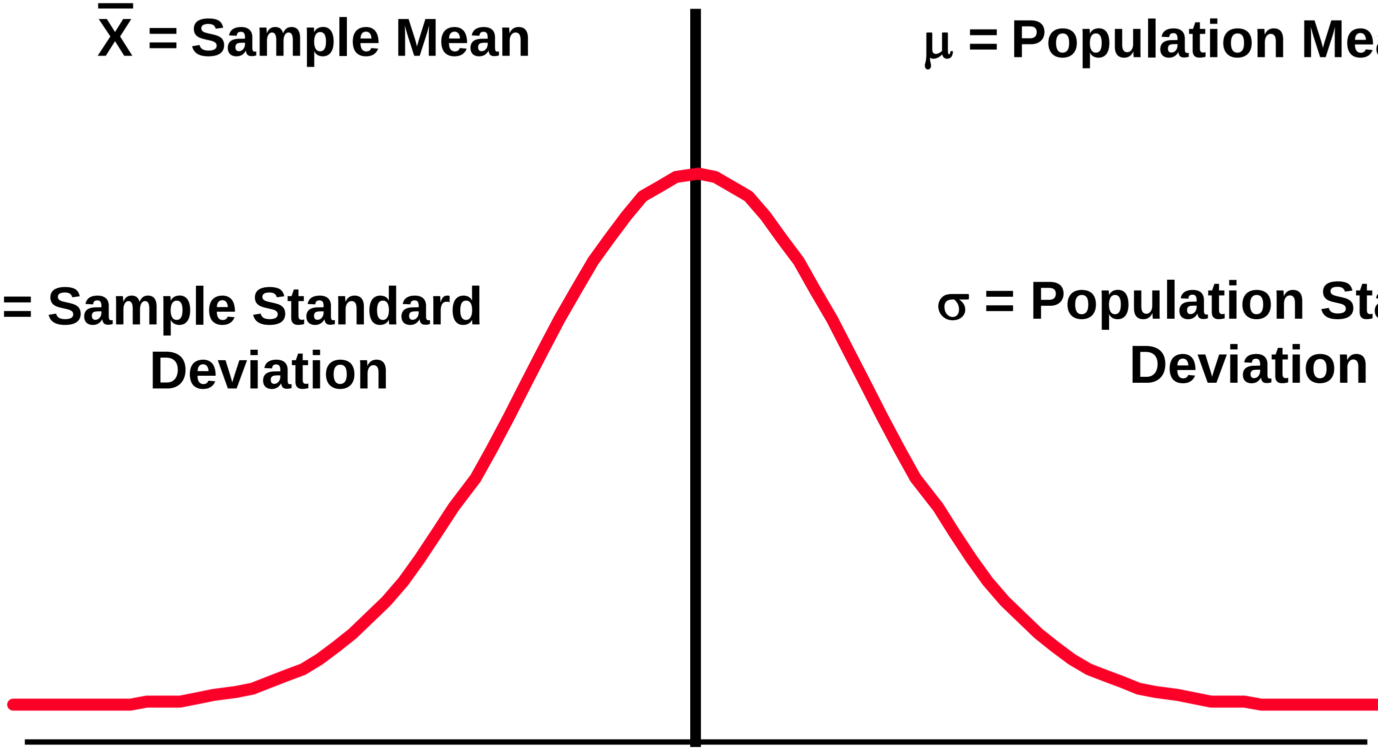
Sample Statistics vs Population Parameters

$\bar{X}$ = Sample Mean

μ = Population Mean

s = Sample Standard Deviation

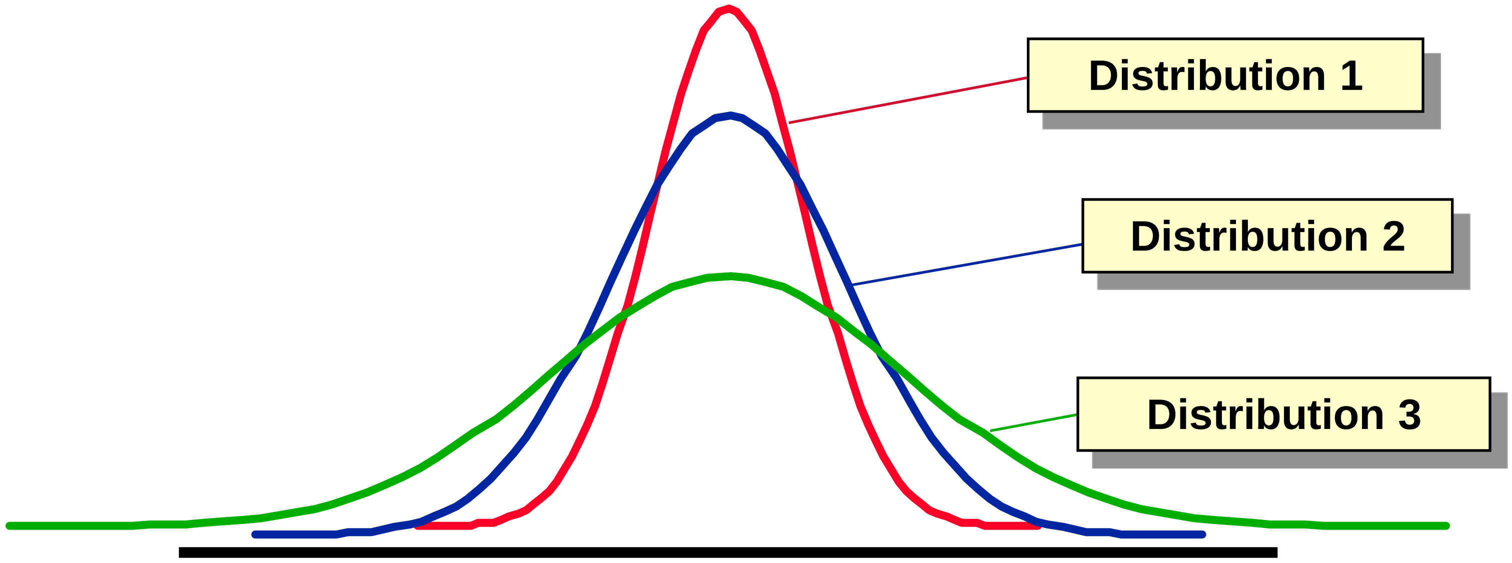
σ = Population Standard Deviation



Statistics → **estimate** → **Parameters**

The Normal Distribution

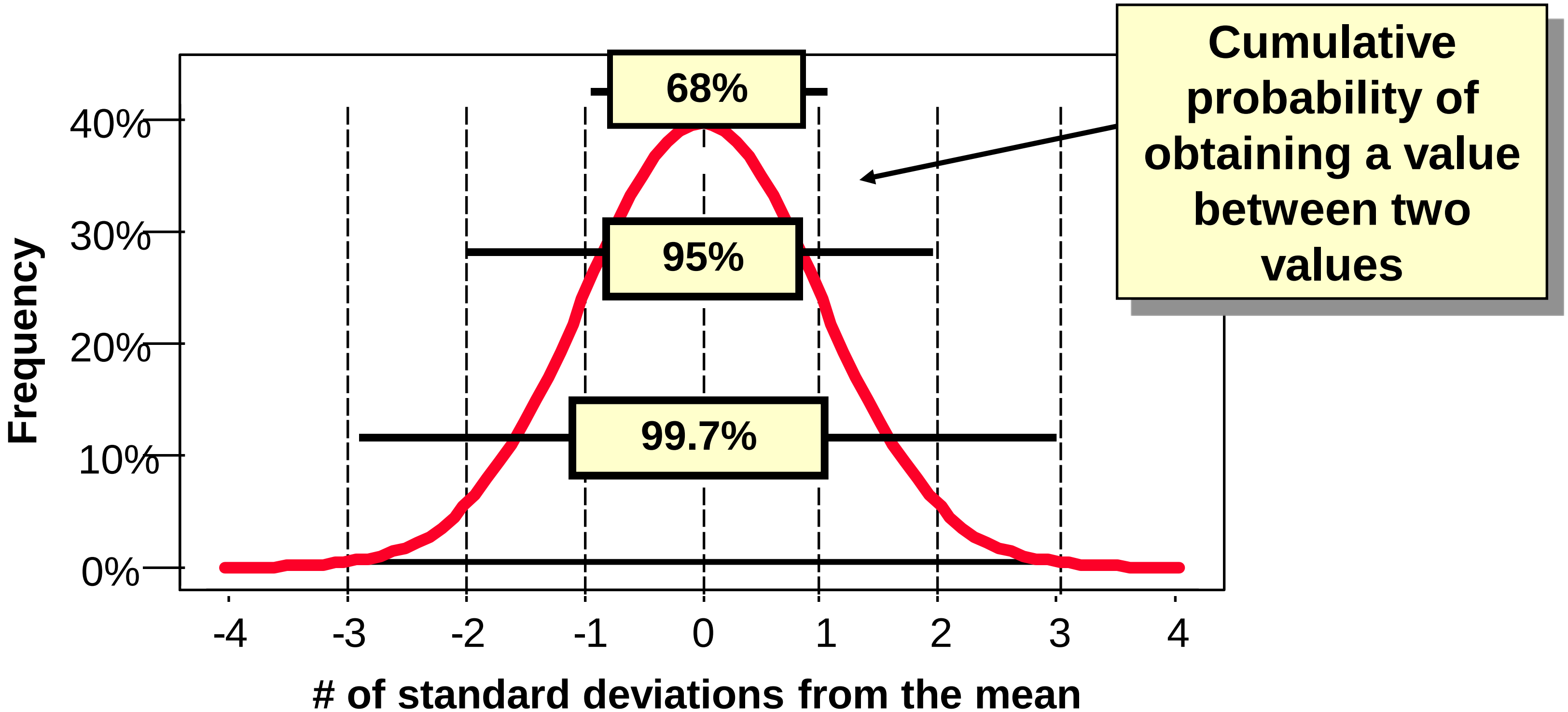
- Property 1: A normal distribution can be described completely by knowing only the :
  – mean and standard deviation



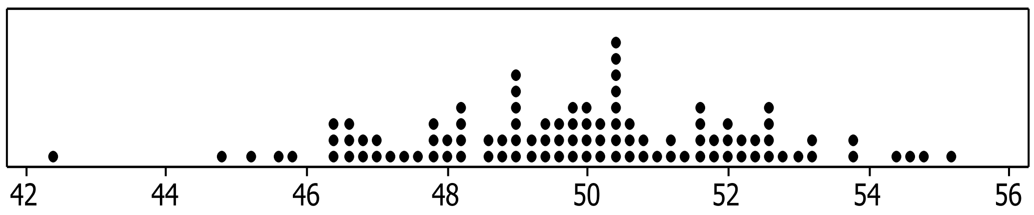
What is the difference between these normal distributions?

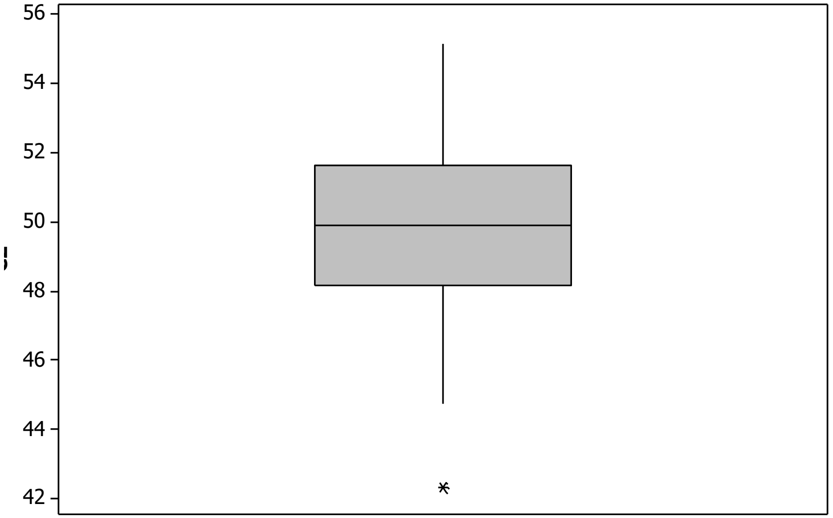
The Normal Curve and Probability Areas Associated with the Standard Deviation

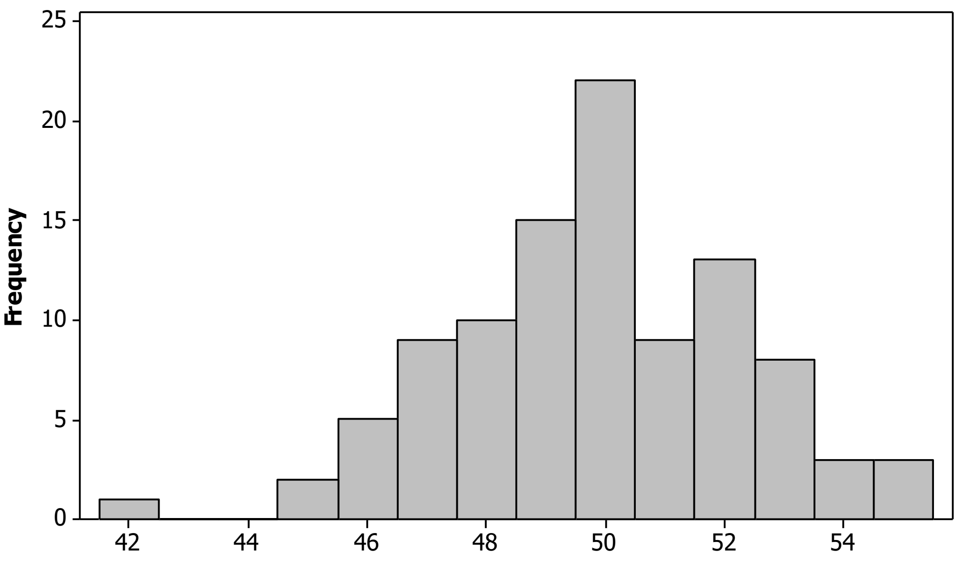
- Property 2: The area under the curve can be used to estimate the probability of certain “events” occurring

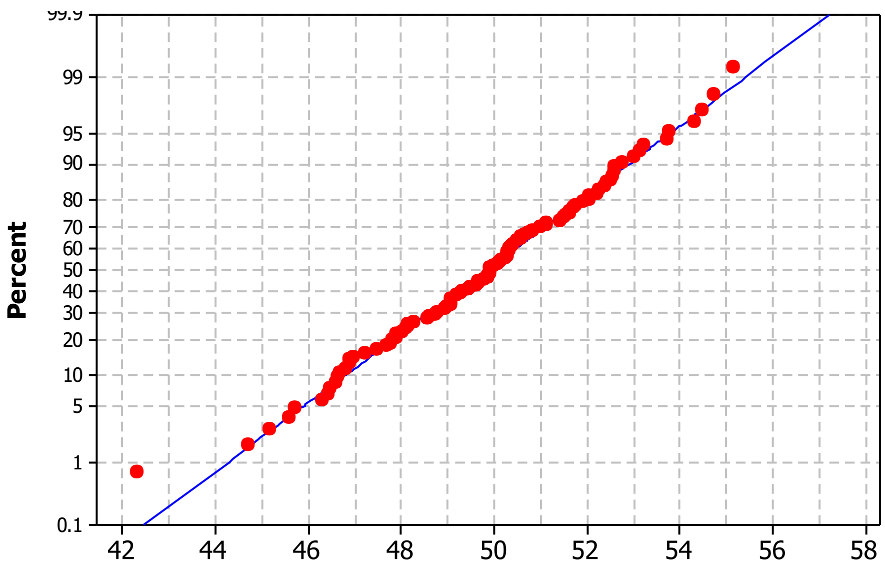


Study Shape: Normal Data

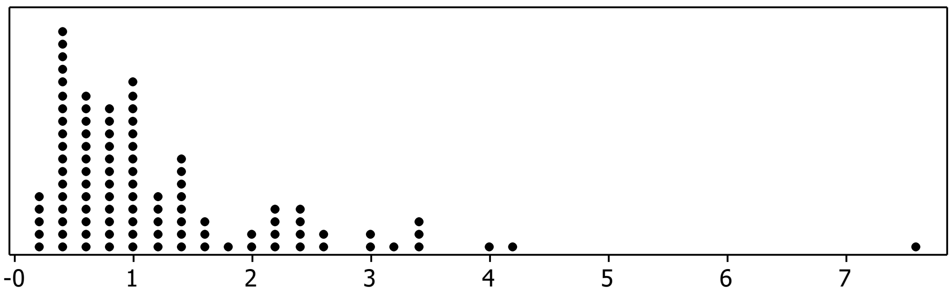


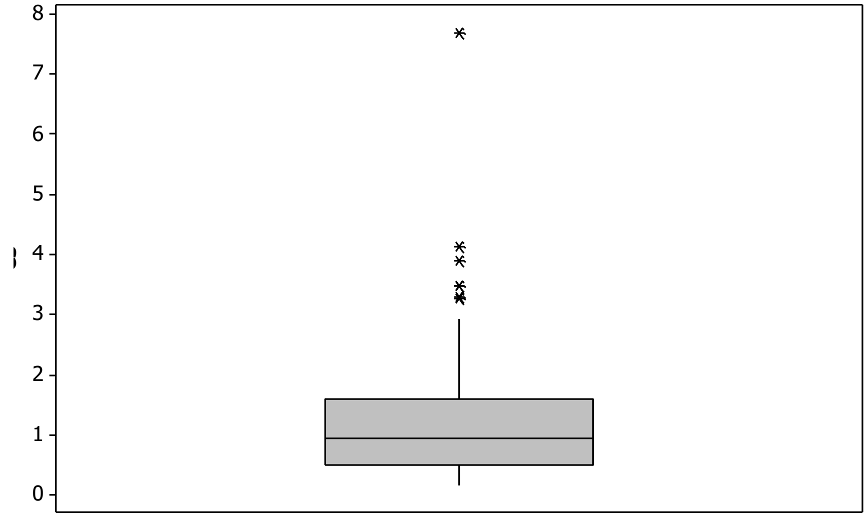


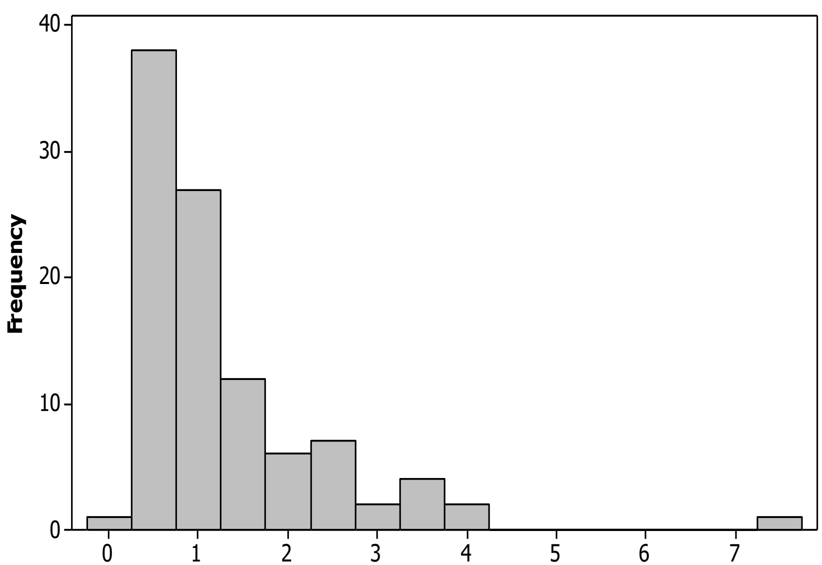


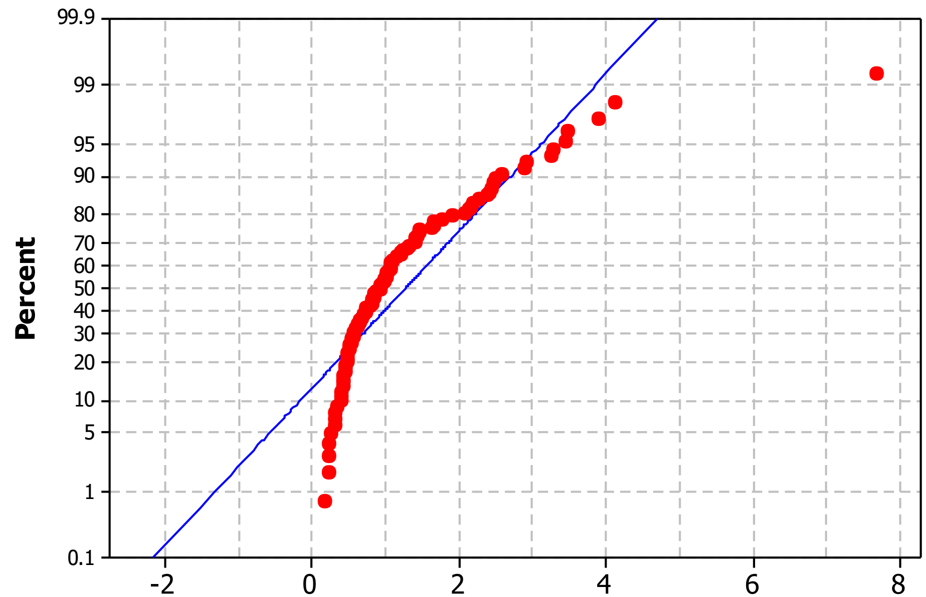


Study Shape: Non-Normal Data









Measures of Center

- Mean: arithmetic average of a set of values
  - reflects the influence of all values
  - strongly influenced by extreme values

Sample

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Population

$$\mu = \frac{\sum_{i=1}^{N} \mathrm{X_i}}{\mathrm{N}}$$

- Median: center number after a set of numbers has been sorted
  - is the 50% rank
  - is “robust” to extreme scores

When to use which measure?

Measures of Spread

- Range: distance between the extreme values of a data set (highest - lowest). The range is sensitive to extreme values.
- Variance: average squared deviation of each data point from the mean, Variance = s^2 or σ^2
- Standard Deviation: square root of the variance

Sample

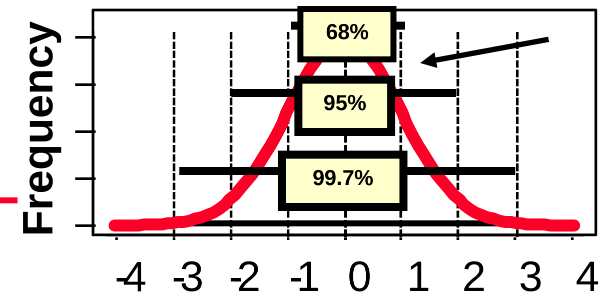
$$\hat{\sigma} = s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$$

Population

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

When to use which measure?

A Simple Example



- Calculate the standard deviation for the data : 2 1 3 5 4

	X	$X - \bar{X}$	$(X - \bar{X})^2$
1	2	-1	1
2	1	-2	4
3	3	0	0
4	5	2	4
5	4	1	1
Σ	15		10
Mean	3		
s-square			2.5
s			1.58

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

$$\sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

Exercise: Turbocharger Data

- CFR : Variable Geometry Turbine Efficiency (%)
- Sample : 100 turbos measured in the Charleston test cell
- Objective : Study shape, center, spread, outliers
- Question : What is the % of turbos with efficiency 62% or higher?
- Project File : Basic Statistics.mpj
- Worksheet : data for exercises, Column 1

Exploratory Data Analysis (EDA)

- In general, data analysis should follow some basic steps:
  - Take a practical look at the raw data to identify any abnormalities (errors, unexpected values)
  - Explore the data graphically to get a visual sense of the data
  - Analyze the data statistically to get a numerical sense of the data (we'll learn how to do this in future modules)

P - Practical

G - Graphical

A - Analytical

Basic Analysis Steps … We’ll do !

- **Practical** – take a visual look at data in worksheet
- **Graphical**

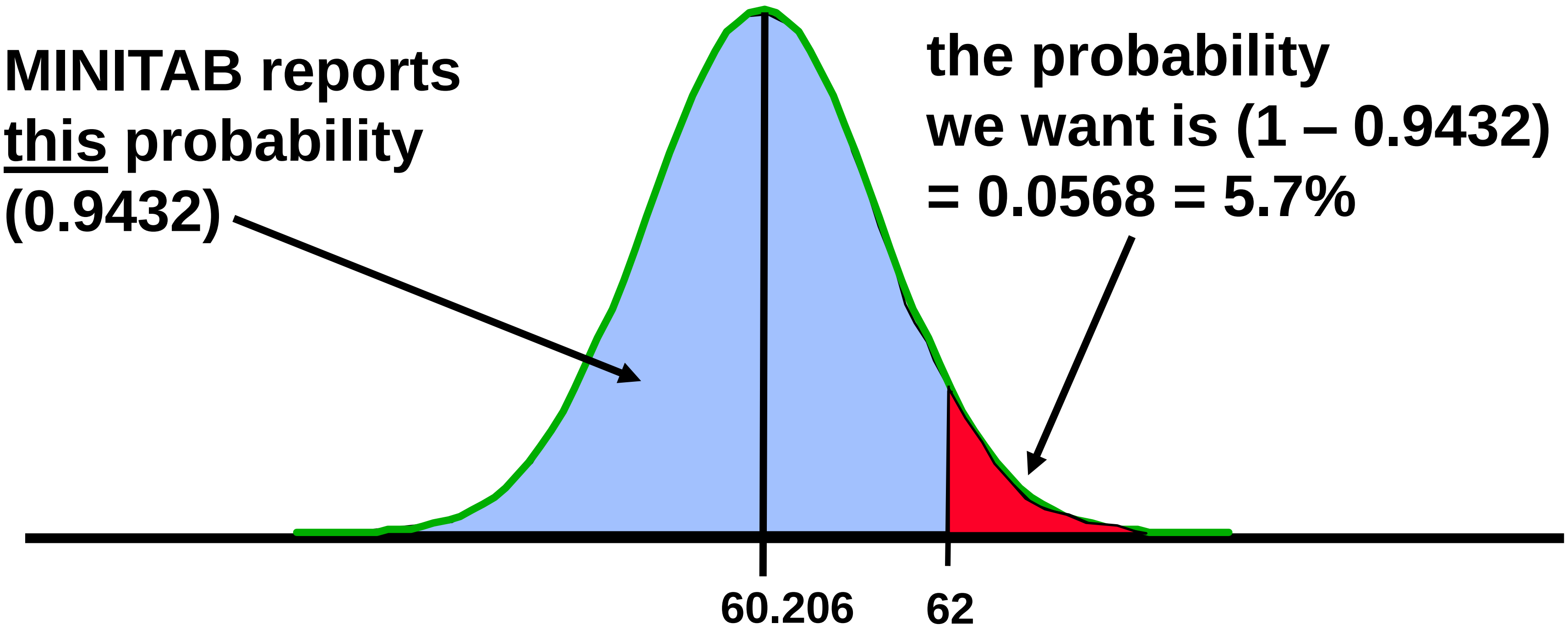
Graph > Histogram
Graph > Dotplot
Graph > Boxplot
Graph > Probability Plot
Stat > Basic Statistics > Normality Test

- **Analytical**

Stat > Basic Statistics > Display Descriptive Statistics
Stat > Basic Statistics > Graphical Summary
Calc > Probability Distributions > Normal (see next page)

How to Use the Probabilities: Example

Calc > Probability Distributions > Normal

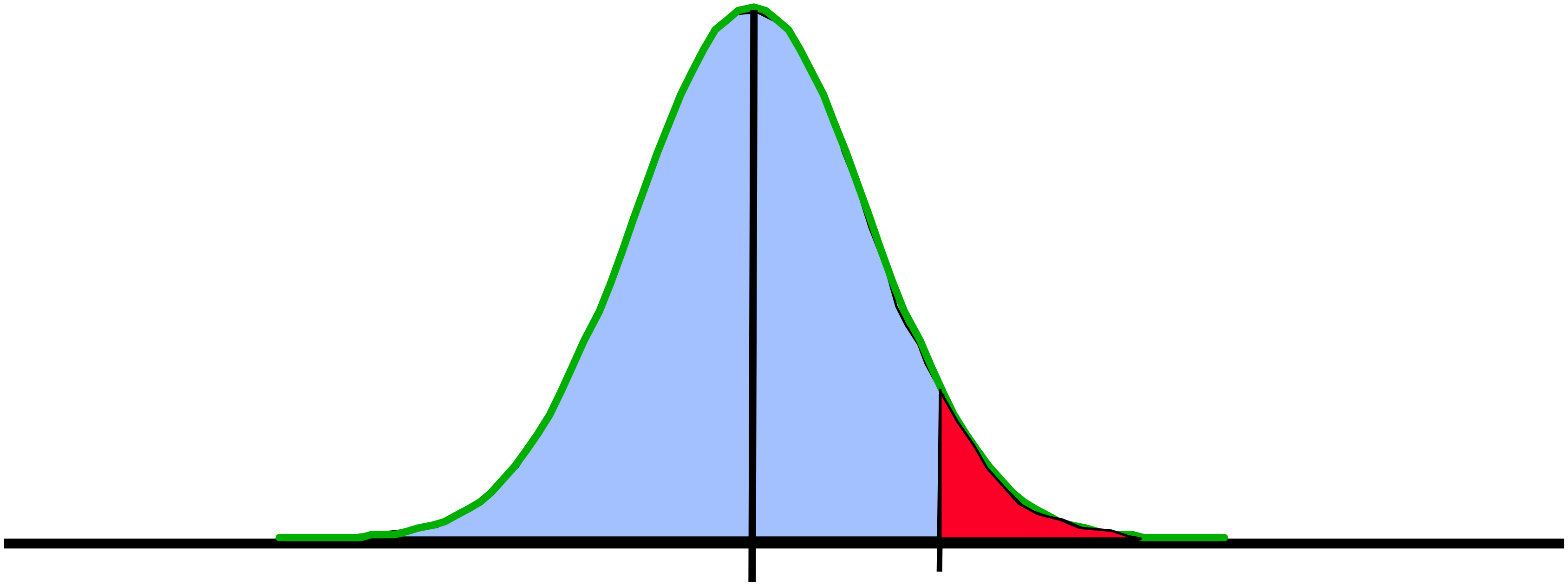



```
Cumulative Distribution Function

Normal with mean = 60.206 and standard deviation = 1.134

     x   P( X <= x )
    62      0.9432
```

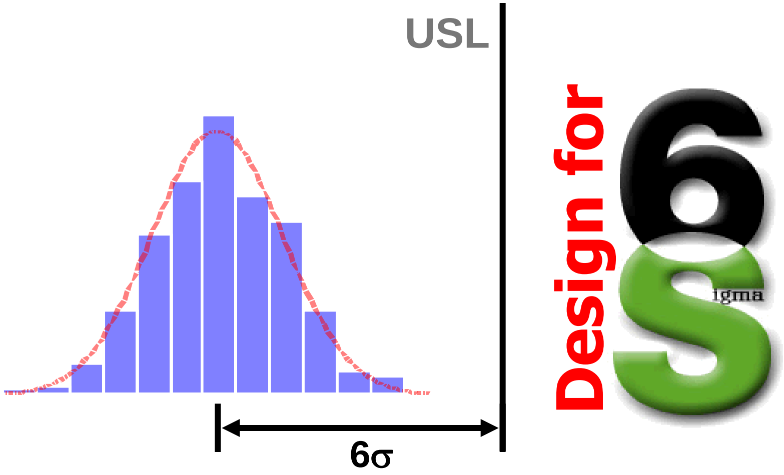
How to Use the Probabilities: Reality Check



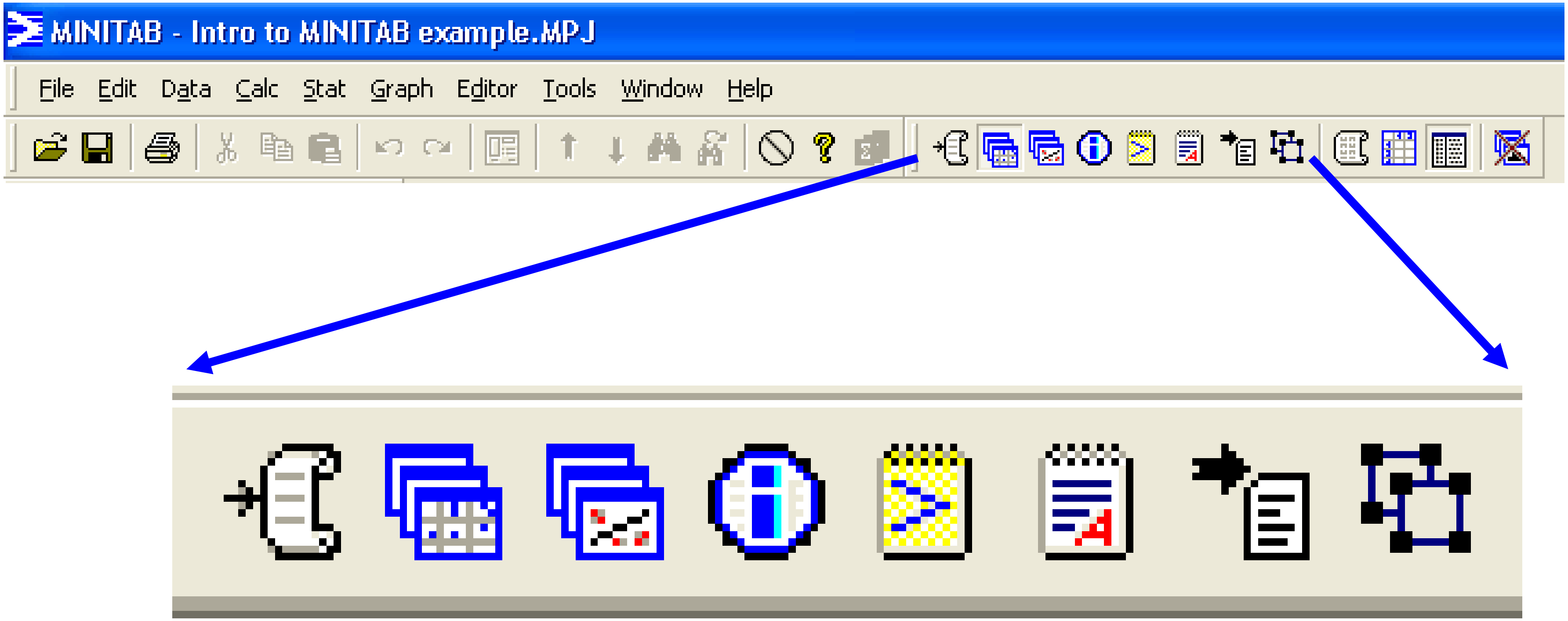
Does 5.7% make sense? How would you check it?

$\bar{x}$ + 1s = 60.206 + 1(1.134) = 61.34 => 16% above
$\bar{x}$ + 2s = 60.206 + 2(1.134) = 62.474 => 2.5% above
$\bar{x}$ + 3s = 60.206 + 3(1.134) = 63.608 => 0.15% above

MINITAB “Getting Around My File” Tips

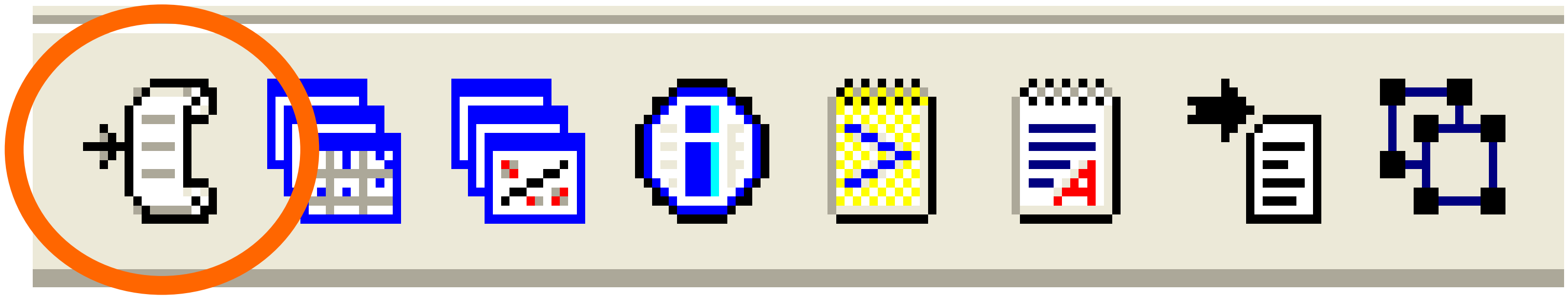


Project Manager Tool Bar



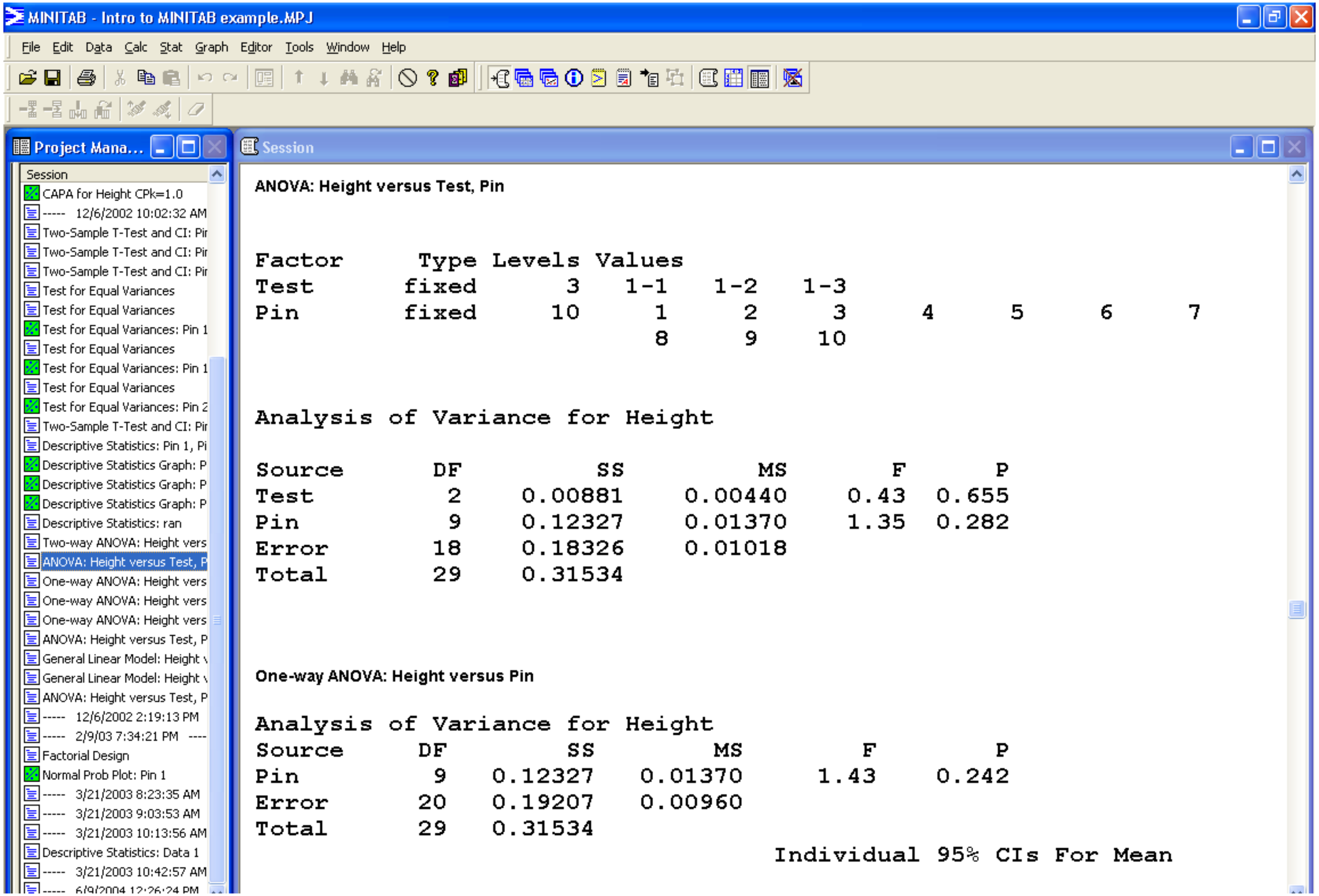
- Open MINITAB file: Basic Statistics.mpj
  - Worksheet: xxxxxxxx
- A Project Manager toolbar allows quick & easy access to everything in your MINITAB project

Session Folder

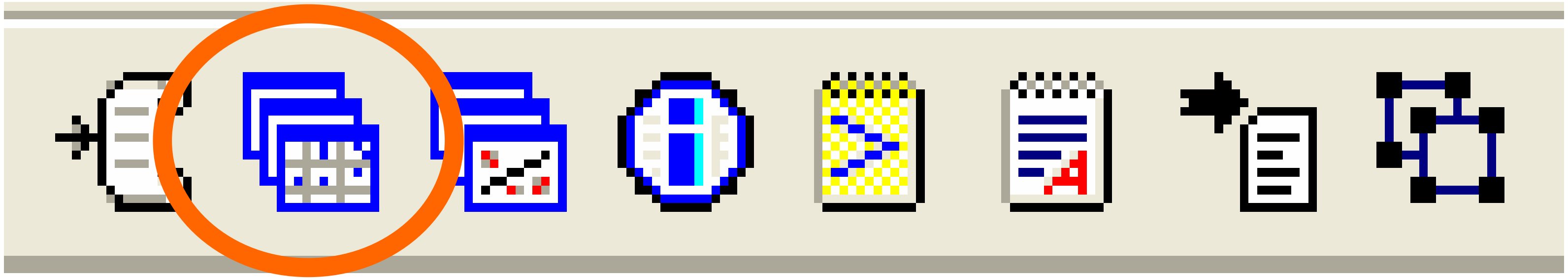


- Go to session folder by clicking this button once
- Double-clicking on a specific item takes you to that analysis in the session window
- right-mouse menu (edit title, delete, print, open StatGuide, etc.)

Session Folder Example

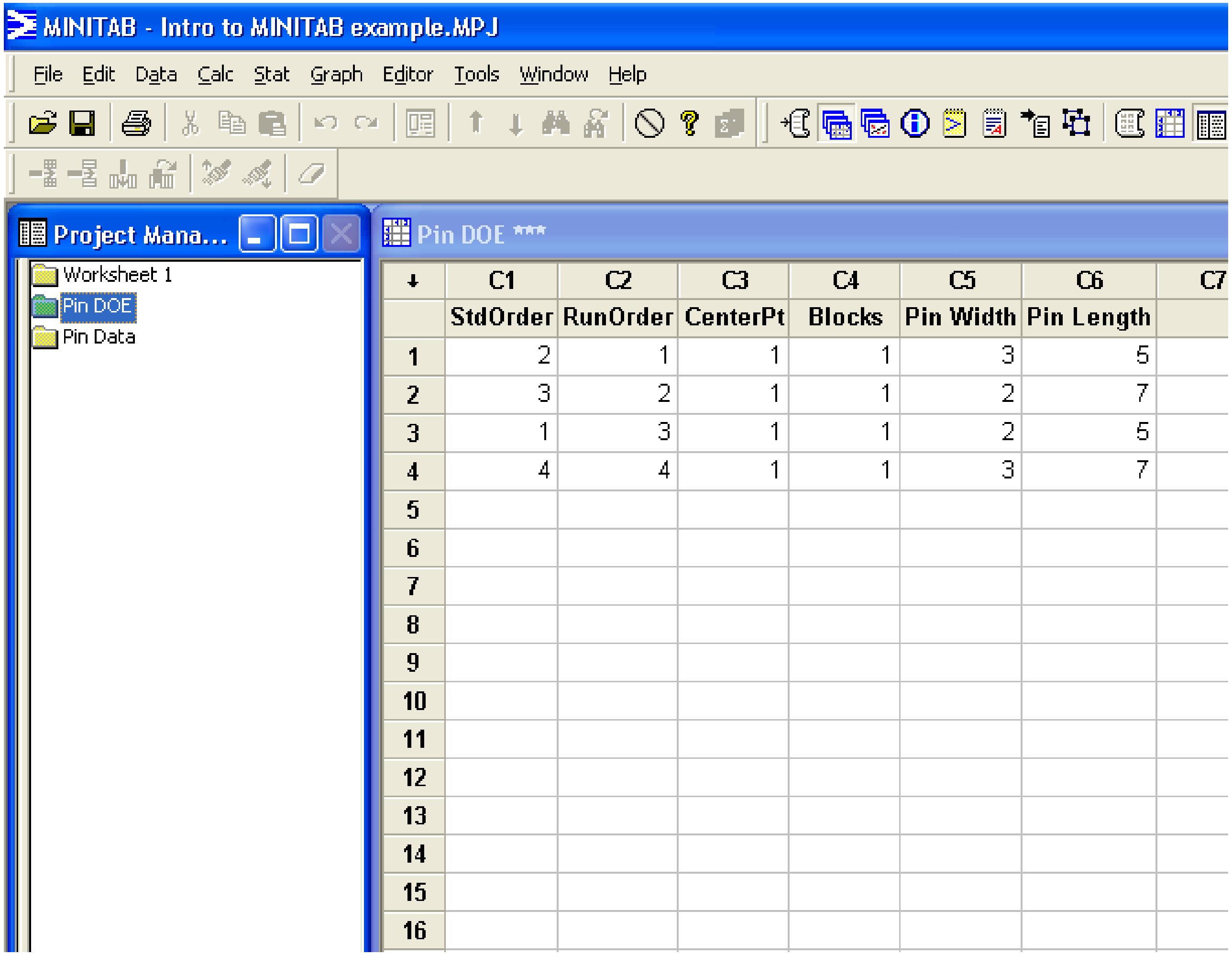


Worksheets Folder

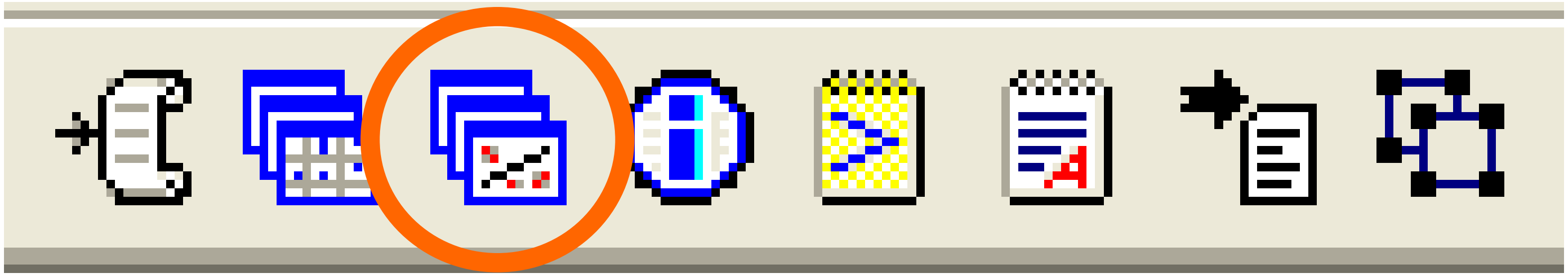


- Go to worksheet folder by clicking this button once
- Double-clicking on a specific worksheet brings it to the front, for analysis
- right-mouse menu (tile, print, save, close, rename, description, etc.)

Worksheets Folder Example

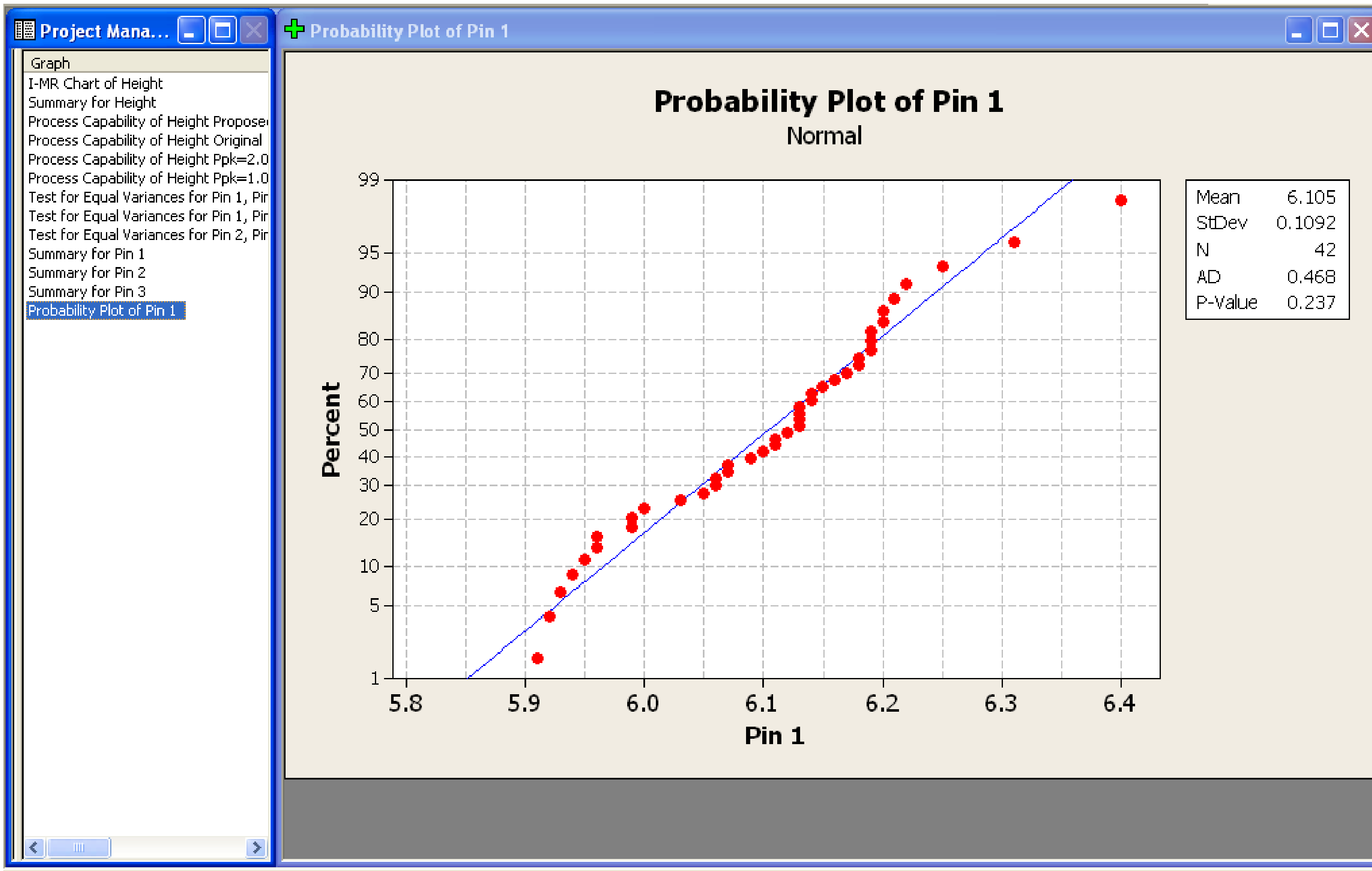


Graphs Folder

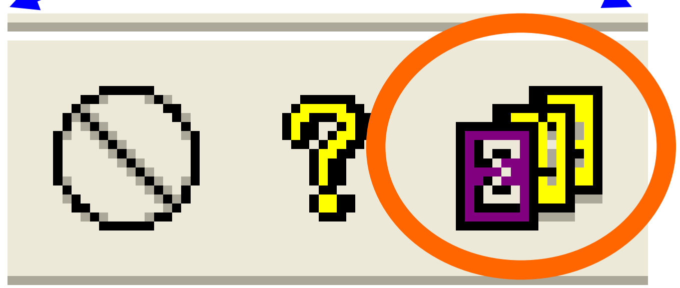


- Go to graph folder by clicking this button once
- Double-clicking on a specific graph brings it to the front
- right-mouse menu (tile, print, save, close, rename, copy, etc.)

Graphs Folder Example



StatGuide

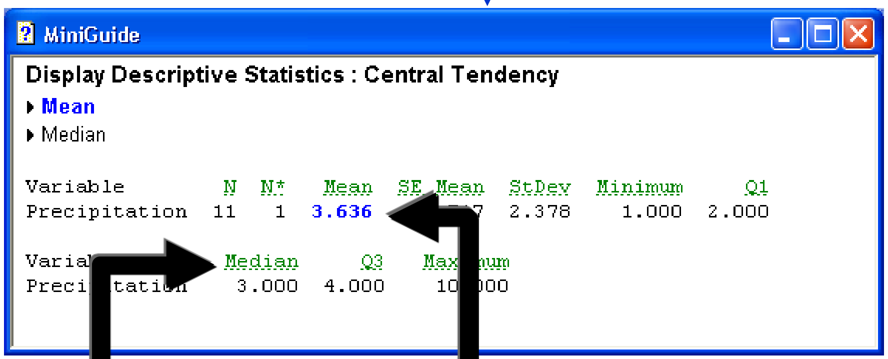


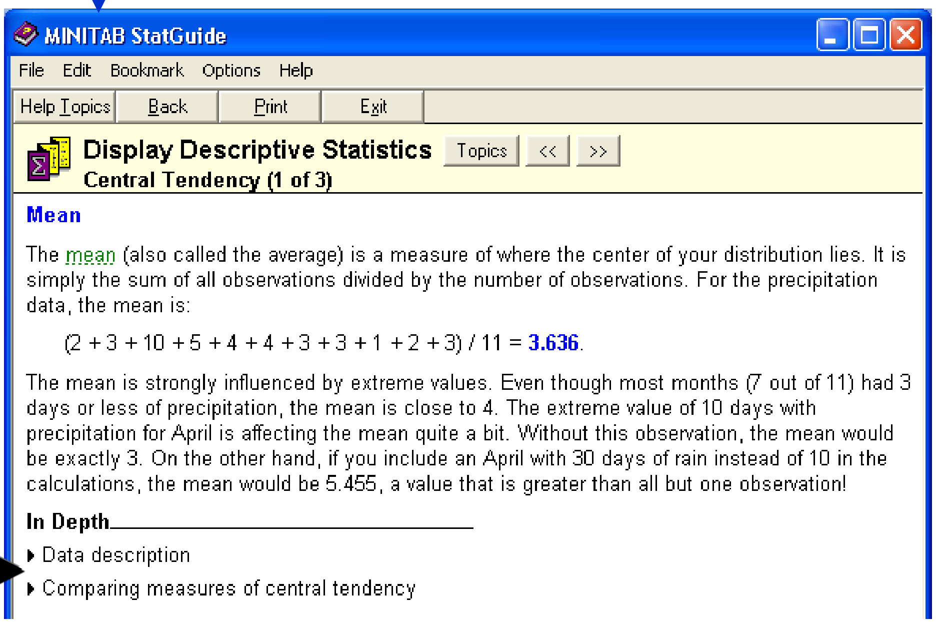
- Provides statistical guidance on your analysis after you run a STAT procedure
- The tone is informal and practical, not like a textbook

StatGuide Layout

Example

Detailed description





Relevant output highlighted

Hot-linked definitions

Links to additional topics

Breakout Exercise: Dominoes

- CFR : Thickness of dominoes
- Sample : a box of dominoes
- Objective : Study shape, center, spread, outliers of the box of dominoes you measure

How to estimate spread with a complex Y (variance components)

- What if the response is a function of several random variables?
- Suppose the critical parameter is Y = BSNOx + BSHC
- How can the parameters of Y be estimated?
- Mean(Y) = Mean(BSNOx) + Mean(BSHC)

Think Pythagorean!

- Do we add standard deviations? NO!
- **Standard deviations do not add**
- Need to add Variances (σ^2)!

$$\hat{\sigma}_Y^2 = \hat{\sigma}_{BSNOx}^2 + \hat{\sigma}_{BSHC}^2$$

$$\hat{\sigma}_Y = \sqrt{\hat{\sigma}_{BSNOx}^2 + \hat{\sigma}_{BSHC}^2}$$

Graphs : Priorities

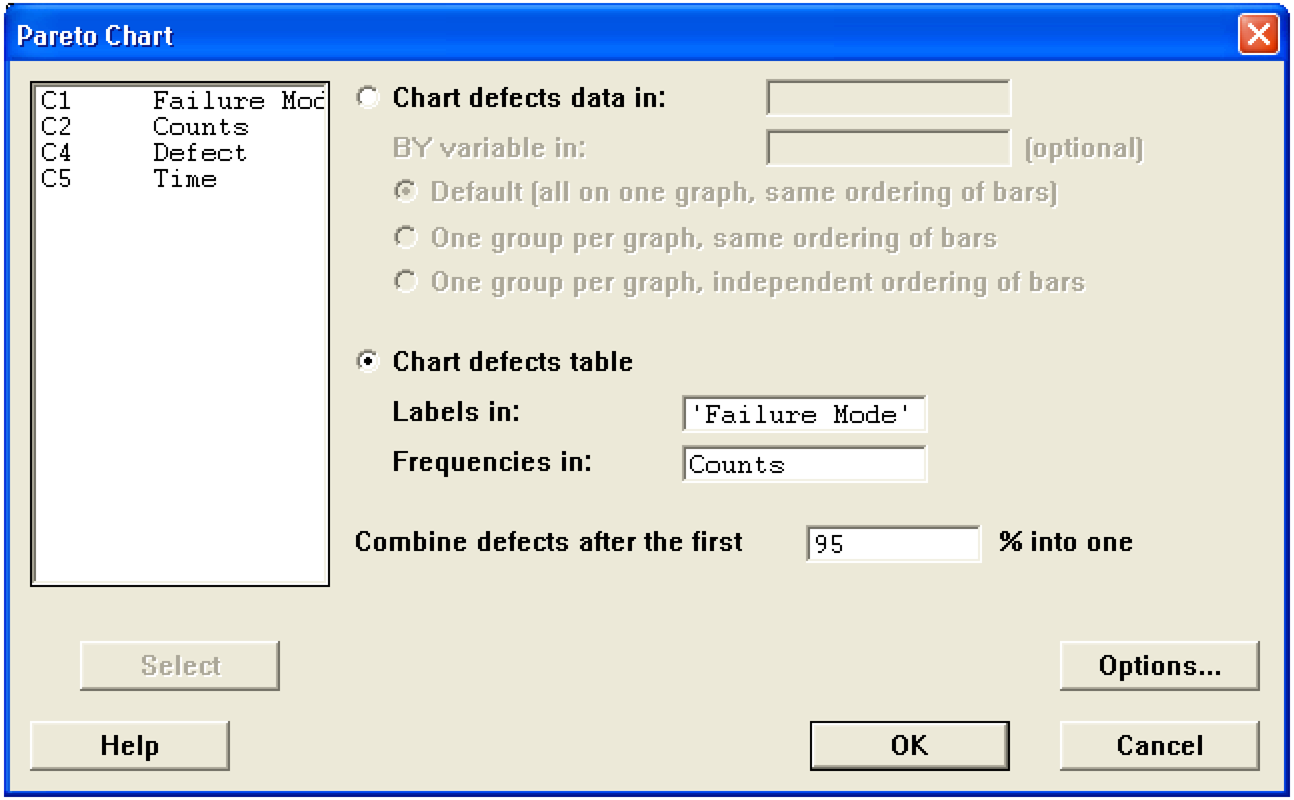
- This is the case where we have data in categories and frequency within each category
  - test cell data on type of failure modes and how many
  - RPN's of potential causes of failure in an FMEA
- Use graphs to :
  - determine how to prioritize the work
  - look for trends or systematic issues with the data
- The graphs we use to study this type of data are :
  - pareto chart

Example : Graphs - Priorities

- Critical Parameter : Counts by Failure Mode
- Objective : To identify failure modes which account for 80% of the failures
- Question : What failure modes need to be addressed?
- Open worksheet “Pareto Data”

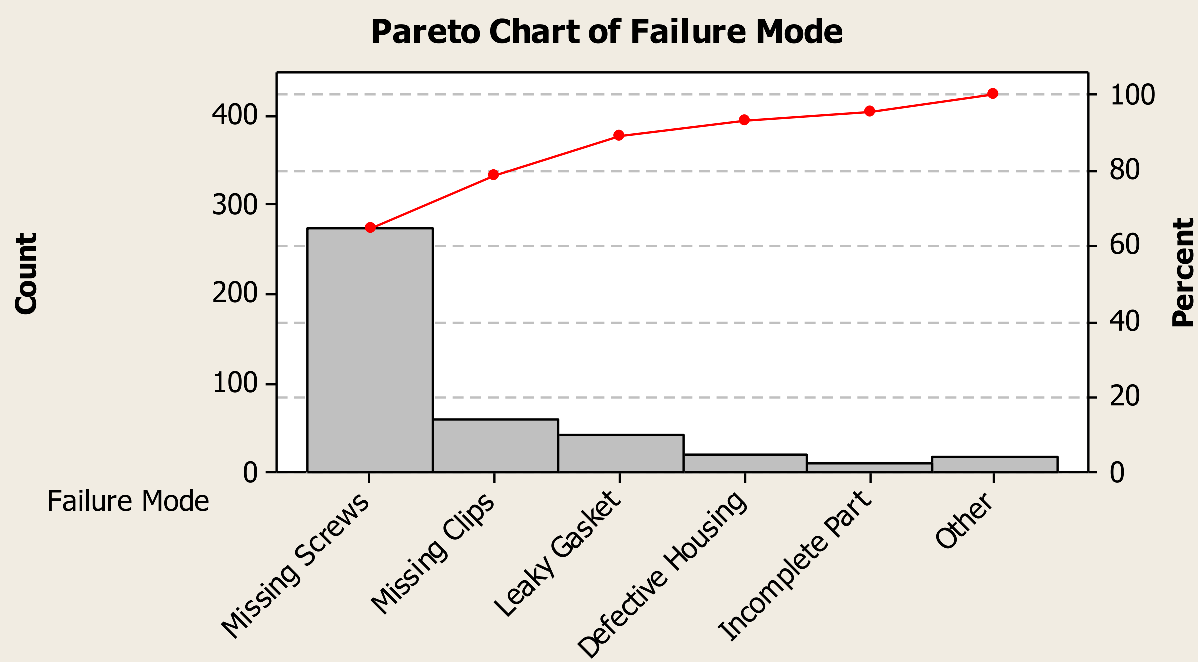
Example : Graphs - Priorities

Stat > Quality Tools > Pareto Chart
use defect table (C1 & C2)



Example : Graphs - Priorities

Failure Mode	Counts
Missing Screws	274
Missing Clips	59
Defective Housing	19
Leaky Gasket	43
Scrap	4
Unconnected Wire	8
Missing Studs	6
Incomplete Part	10



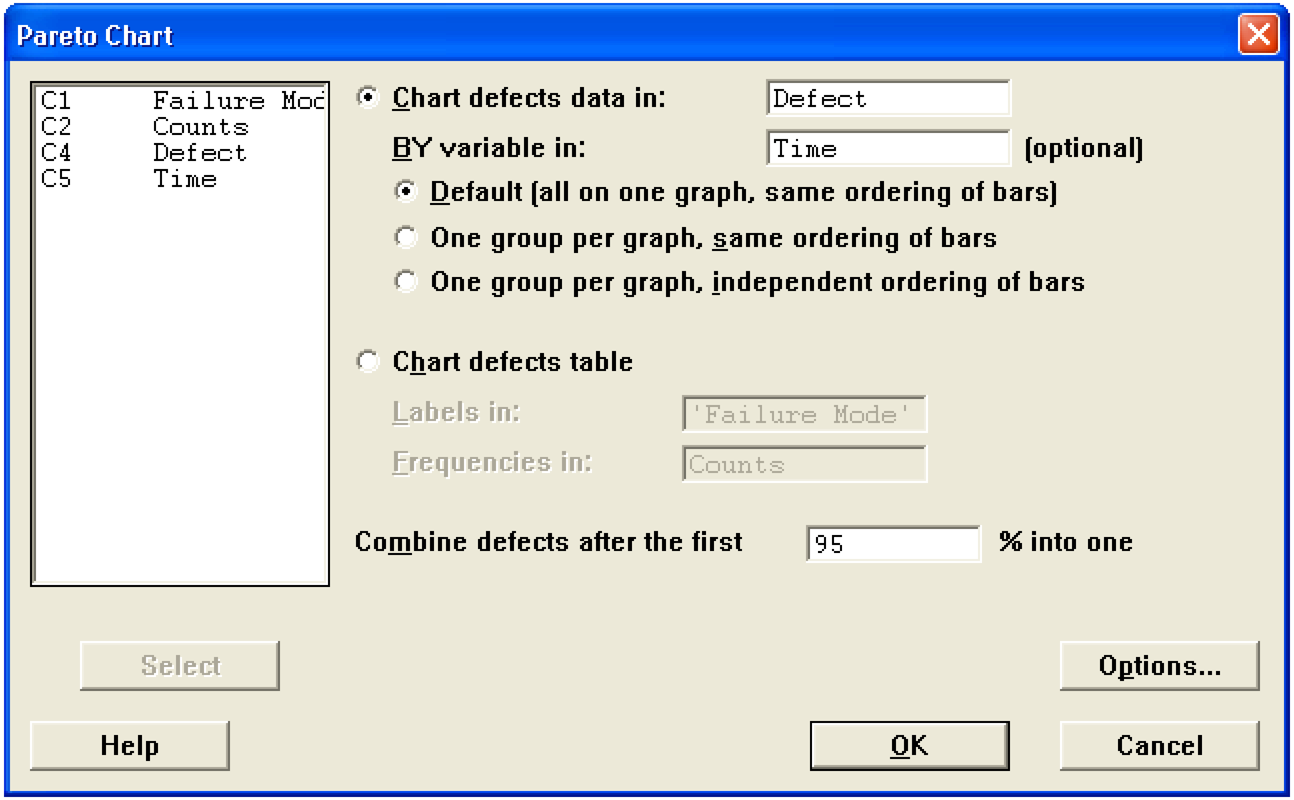
Count	274	59	43	19	10	18
Percent	64.8	13.9	10.2	4.5	2.4	4.3
Cum %	64.8	78.7	88.9	93.4	95.7	100.0

Example : Graphs - Priorities

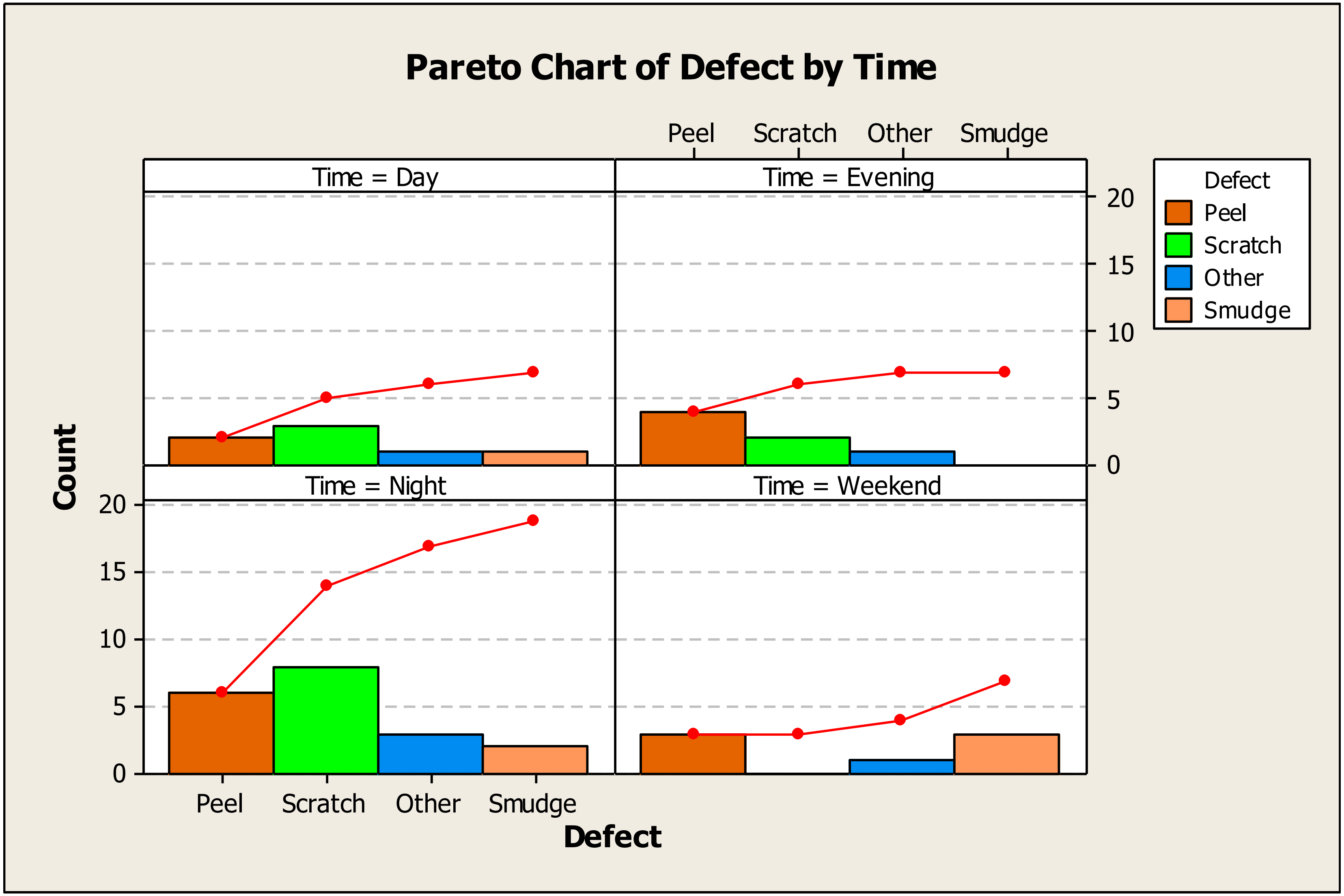
- Critical Parameter : Defect Count
- Objective : Determine if there are any systematic issues in the data that need to be addressed
- Question : Is there an effect due to Time?
- Open worksheet “Pareto Data”

Example : Graphs - Priorities

Stat > Quality Tools > Pareto Chart
use defect data (C4 & C5)



Example : Graphs - Priorities



Example Summary : Priorities

- For the Failure Mode data
  - need to eliminate missing screws & missing clips
  - perhaps these can go to zero
  - typically never fix a failure mode 100% unless you remove the component
  - would probably want to address the 3rd failure mode as well to ensure 80% reduction
- For the Defect data by Time
  - weekend had no scratches - WHY?
  - would want to identify what in the process is different
  - a watch-out is that the counts aren't normalized by opportunity (don't know total production by time)

Objectives Revisited

Should be able to:

- Use graphs to visually represent variation
- Assess normality using the probability plot
- Describe a variable's distribution using graphs as well as shape, center, and spread metrics
- Use the normal distribution to calculate the probability of an outcome