

# **Statistical Applications**

## **Using Minitab**

May 14, 2014

Larry Bartkus

Los Angeles Section ASQ

## The Statistician

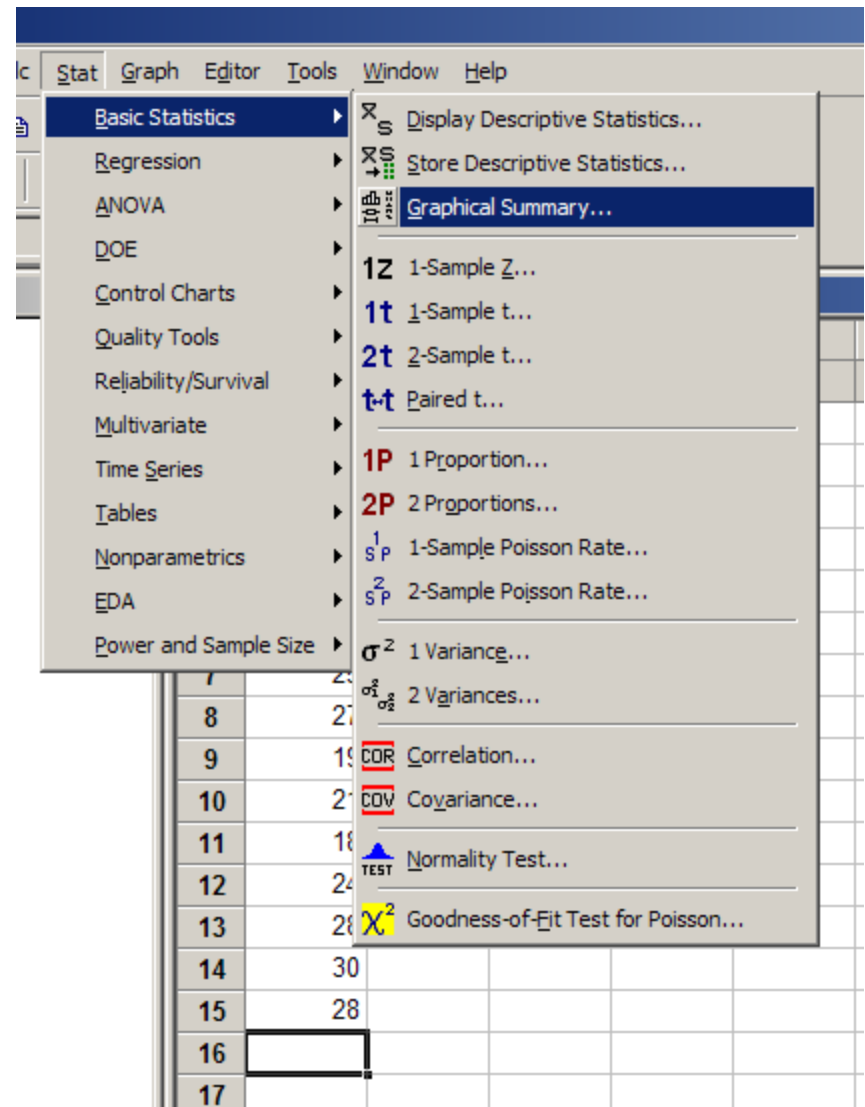
The Statistician is a person who poses as an exacting expert on the basis of being able to turn out with prolific fortitude infinite strings of incomprehensible formulae calculated with micrometric precision from vague assumptions which are based upon debatable figures taken from inconclusive experiments carried out with instruments of problematical accuracy by persons of doubtful reliability and questionable mentality.

# Key Take-Aways

- Minitab is a Simple and Powerful Tool in Data Analysis and Display
- Always try to Graphically Display your data
  - Easier to understand
  - Draws picture of what is really happening
  - Check for Patterns
    - Is it Normal?
    - Time Series for trends, changes, effects
- Organize and Plan you Data Collection first
- Analyze the results and recheck any assumptions
- Have Fun. Software has made data analysis easier, faster, and funner!

# A Warm-up Look

- First record 10-15 data points in the first column.
- Let's name the column "Results"
- Go to Stat>Basic Statistics>Graphical Summary



## How Do I Look At The Data ?

- It's PGA, Baby!!
  - Practical (Are there any obvious problems or patterns?) ANOB or ANOG
  - Graphical (Look at the Pictures)
    - Two different situations, two histograms requiring different approaches
  - Analytical (Check out the Statistics)

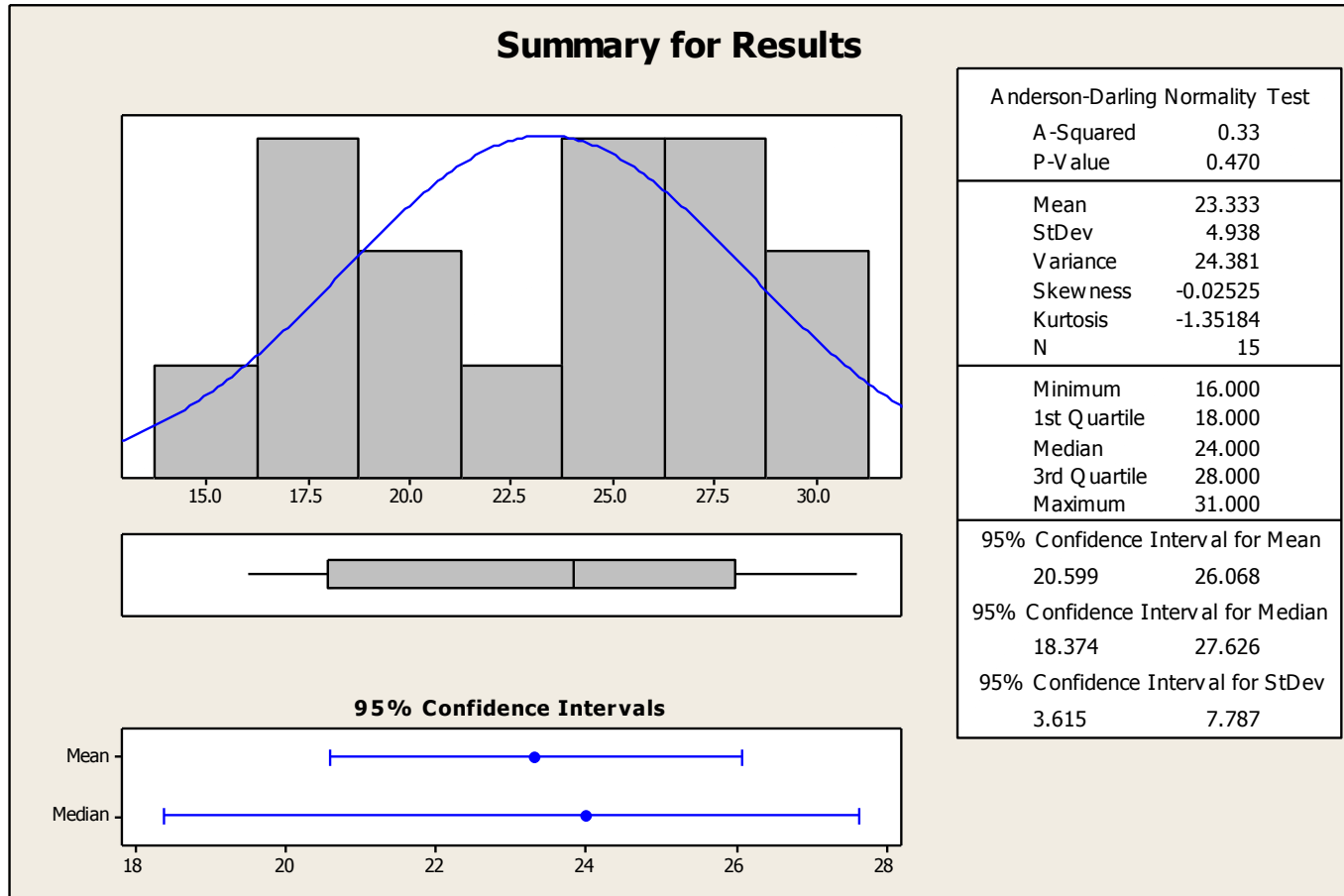
# You should end up with something like this

The screenshot shows the Minitab software interface. The main window is titled "Minitab - Untitled" and contains a worksheet named "Worksheet 1 \*\*\*". The worksheet has columns C1 through C17. Column C1 is labeled "Results" and contains the following data values: 18, 17, 25, 31, 16, 23, 25, 27, 19, 21, 18, 24, 28, 30, 28. A "Graphical Summary" dialog box is open, showing "C1 Results" in the "Variables:" field. The "Confidence level:" is set to 95.0. The dialog box has buttons for "Select...", "Help", "OK", and "Cancel".

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
	Results																
1	18																
2	17																
3	25																
4	31																
5	16																
6	23																
7	25																
8	27																
9	19																
10	21																
11	18																
12	24																
13	28																
14	30																
15	28																
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	

Display a graphical summary

# And a Graphical Summary



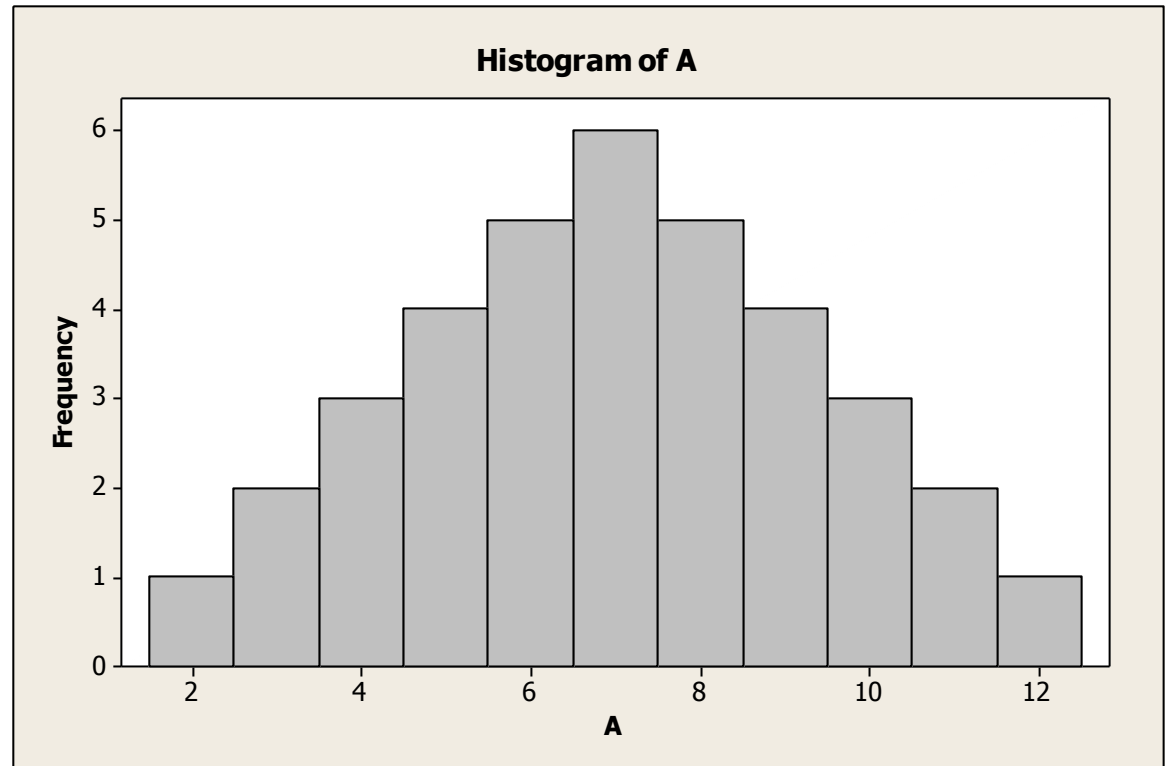
## More About Displaying Data

A histogram is a graphical display of tabulated frequencies, which are shown as bars. It illustrates what portion of cases fall into each of several categories.

Lets look at an example for the following data:

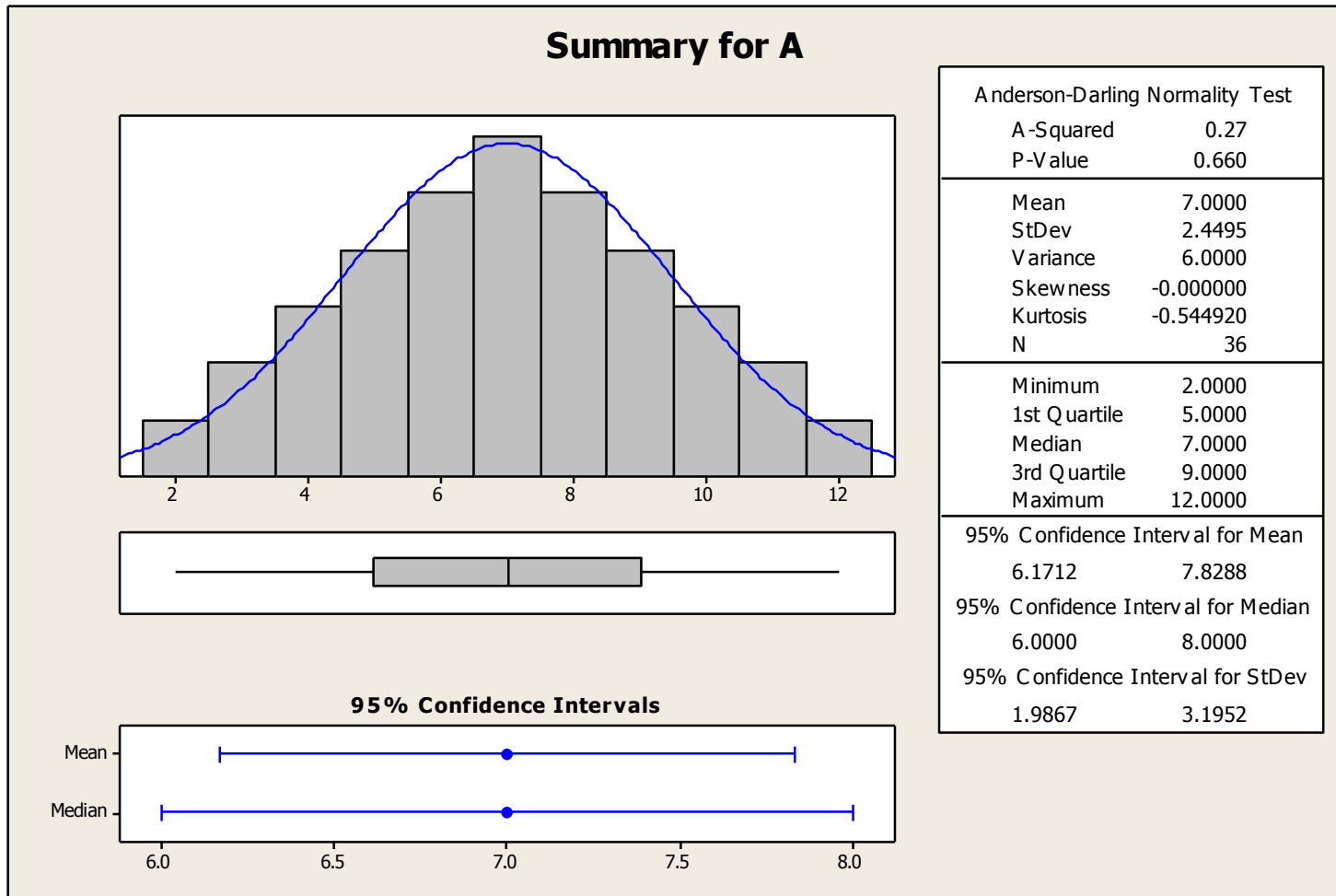
4,9,7,2,11,7,12,6,8,5,3,8,7,6,9,4,9,6,5,10,  
8,5,7,3,8,6,10,9,10,7,11,6,8,5,4,7.

The histogram would look like this:





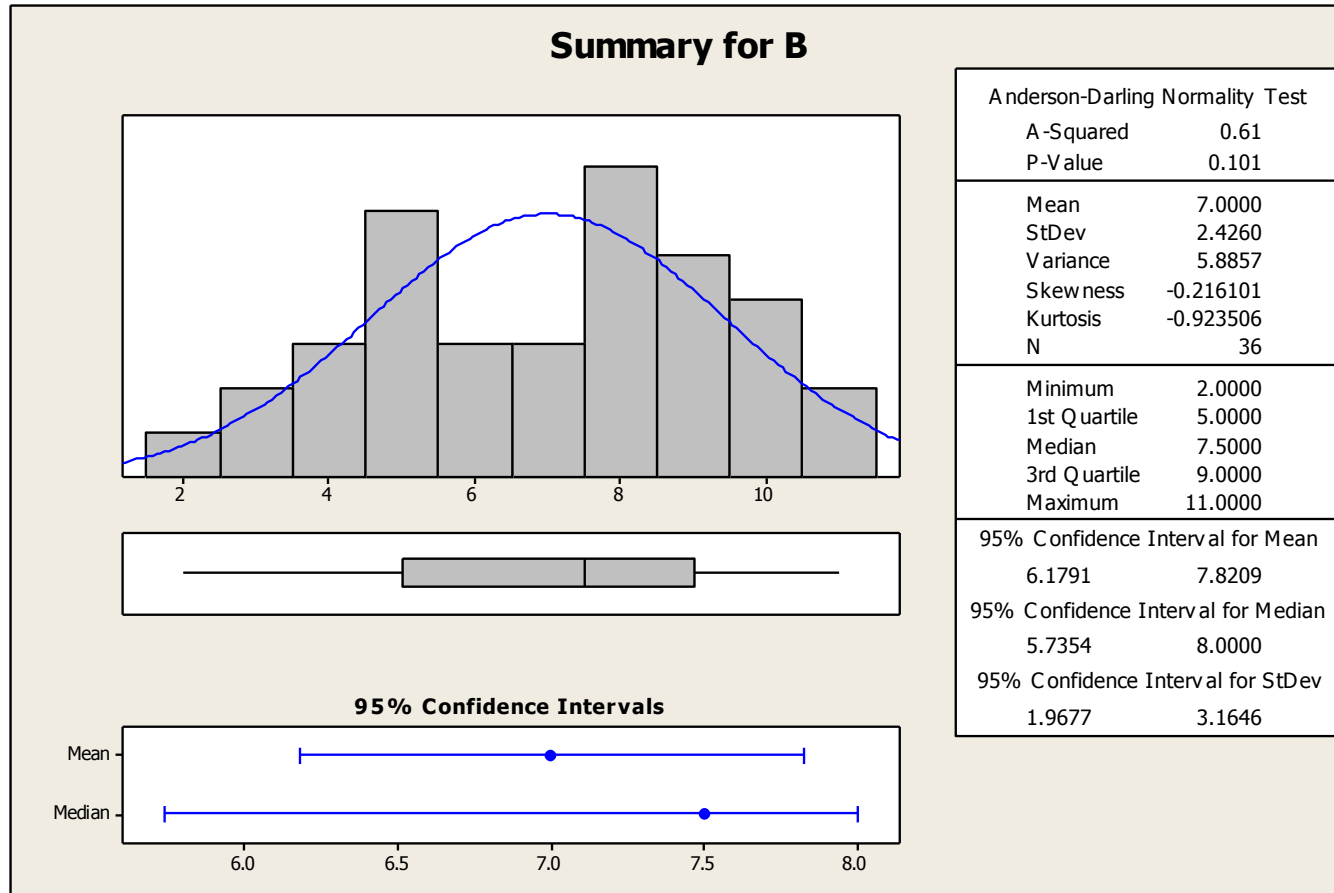
# A Quantitative Look At The Data



How about this data: 4,9,7,5,10,7,11,8,8,5,3,8,9,5,9,3,5,9,5,10

6,7,9,4,8,2,10,8,10,6,8,6,8,5,4,11.

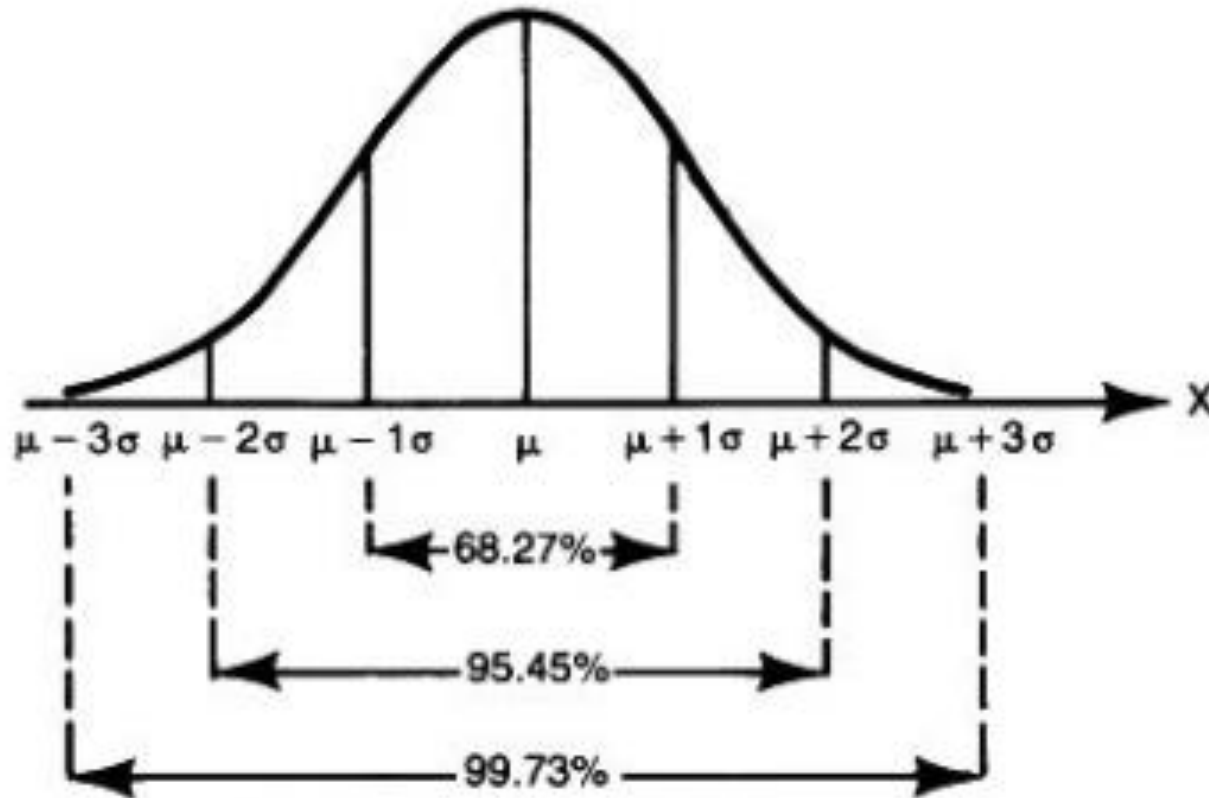
# How Does The New Data Set Look?



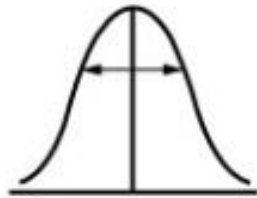
The Mean is the same and the Std Dev is almost identical, but the shape of their distribution is very different. What type of distribution is this?

# **Distributions**

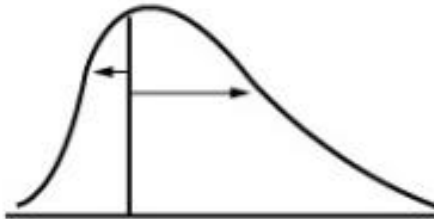
# The Normal Distribution aka The Bell Curve



# Non-Normal Distributions



A. Normal Distribution



B. Skewed Distribution



C. Exponential Distribution

# Checking Many Distributions At The Same Time

BES.MPJ

Calc Stat Graph Editor Tools Window Help

Basic Statistics  
Regression  
ANOVA  
DOE  
Control Charts  
**Quality Tools**  
Reliability/Survival  
Multivariate  
Time Series  
Tables  
Nonparametrics  
EDA  
Power and Sample Size

Run Chart...  
Pareto Chart...  
Cause-and-Effect...  
**Individual Distribution Identification...**  
Johnson Transformation...  
Capability Analysis  
Capability Sixpack  
Gage Study  
Attribute Agreement Analysis...  
Acceptance Sampling by Attributes...  
Acceptance Sampling by Variables  
Multi-Vari Chart...  
Symmetry Plot...

C3	C4	C5
Nat Log	Limits	Log10Lim
5.29832	1000	
6.08677	2000	
5.39363		
7.00307		
4.38203		
4.00733		
4.09434		
7.43838		
4.60517		
5.29832		
5.07517		
4.70048		
5.13580		

6  
7  
8  
9  
10  
11  
12  
13

170 2.23045

# Individual Distribution Identification

The screenshot shows the 'Individual Distribution Identification' dialog box in Minitab. On the left, a list of columns is shown: C1 Aerobes, C2 Log10, C3 Nat Log, and C4 Limits. A 'Select' button is below this list. The main area is titled 'Data are arranged as' and has two radio buttons. The first, 'Single column:', is selected and has a text box containing 'Aerobes'. Below it is a 'Subgroup size:' text box containing '1', with the instruction '(use a constant or an ID column)' underneath. The second radio button, 'Subgroups across rows of:', is unselected and has an empty list box below it. There are two more radio buttons: 'Use all distributions and transformations' (selected) and 'Specify' (unselected). Under 'Specify', there are four checked checkboxes for 'Distribution 1:', 'Distribution 2:', 'Distribution 3:', and 'Distribution 4:', each with a dropdown menu. The dropdowns are set to 'Normal', 'Exponential', 'Weibull', and 'Gamma' respectively. On the right side of the dialog, there are four buttons: 'Box-Cox...', 'Johnson...', 'Options...', and 'Results...'. At the bottom right are 'OK' and 'Cancel' buttons. At the bottom left is a 'Help' button. The dialog box has a title bar with the text 'Individual Distribution Identification' and a close button (X).

Individual Distribution Identification

C1 Aerobes  
C2 Log10  
C3 Nat Log  
C4 Limits

Select

Help

Data are arranged as

☒ Single column: Aerobes

Subgroup size: 1

(use a constant or an ID column)

☐ Subgroups across rows of:

☒ Use all distributions and transformations

☐ Specify

☒ Distribution 1: Normal

☒ Distribution 2: Exponential

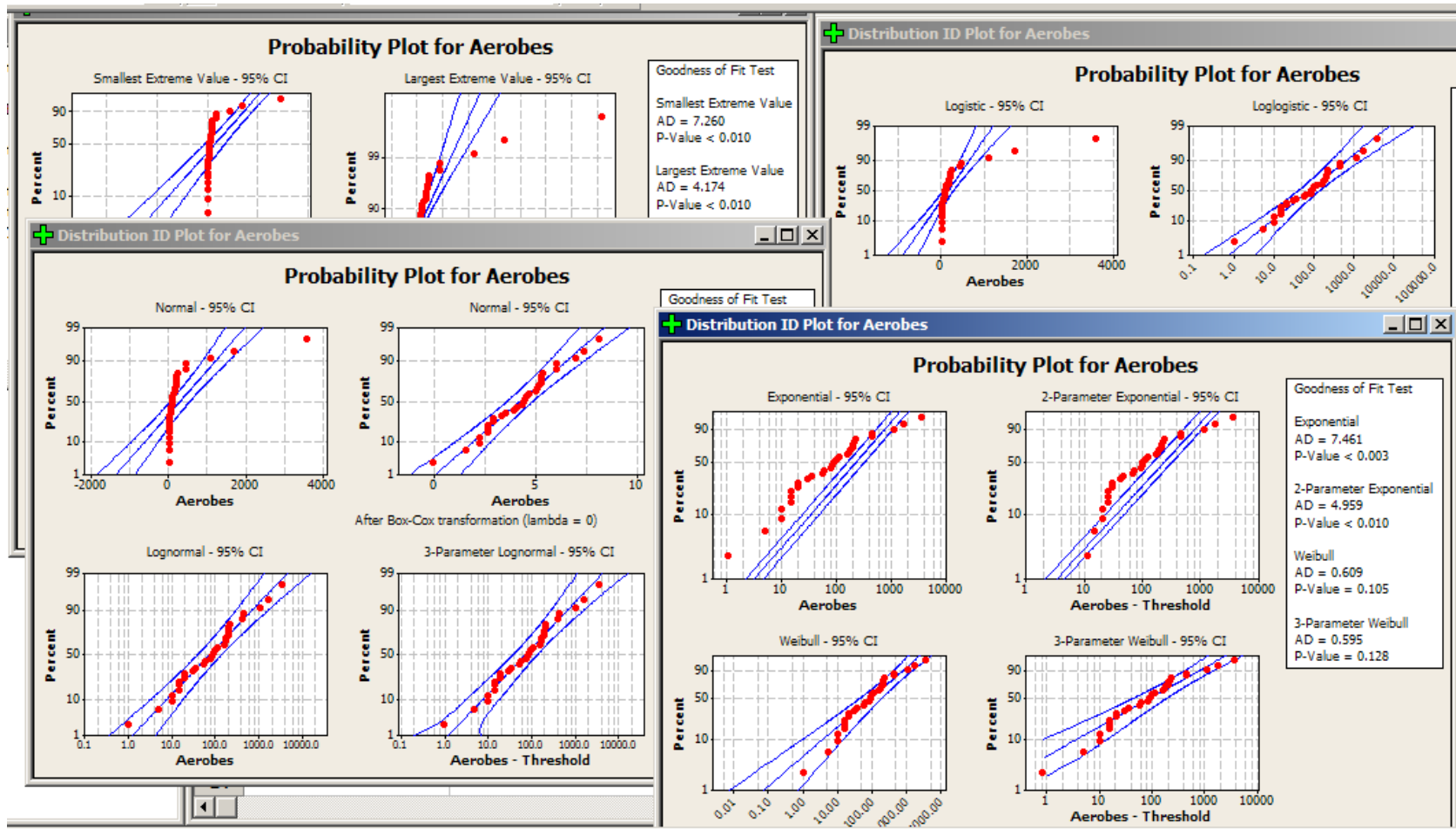
☒ Distribution 3: Weibull

☒ Distribution 4: Gamma

Box-Cox...  
Johnson...  
Options...  
Results...

OK  
Cancel

# Here Is What You Get



Holy Crap !



# Transformation

# Your Data Analysis Choices

- Normal Distribution (Bell Curve)
- Transform to Normalize the Data
- Utilize Non-parametric Statistics
- Treat as Discrete (Attribute Data)

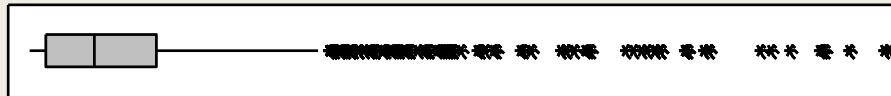
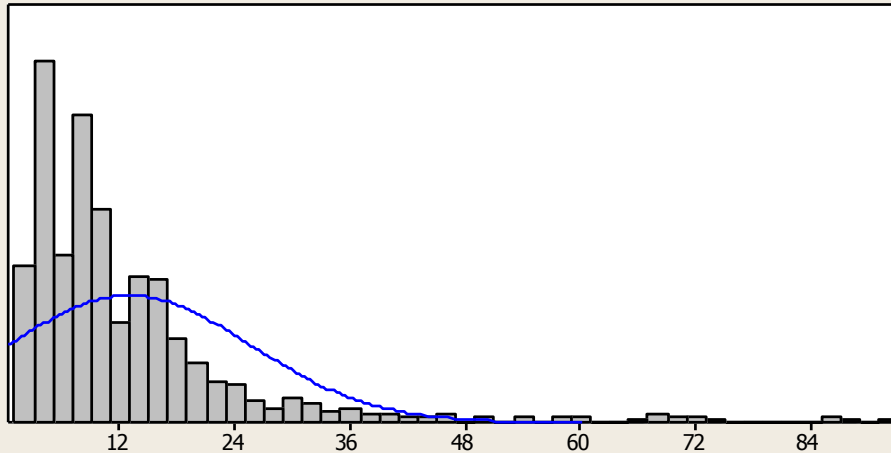
## Transformations

Make a transformation of the original characteristic to a new characteristic that is normally distributed. These transformations are useful for (a) achieving normality of measured results, (b) satisfying the assumption of equal sample variances required in certain tests, and (c) satisfying the assumption of additivity of effects in certain tests.

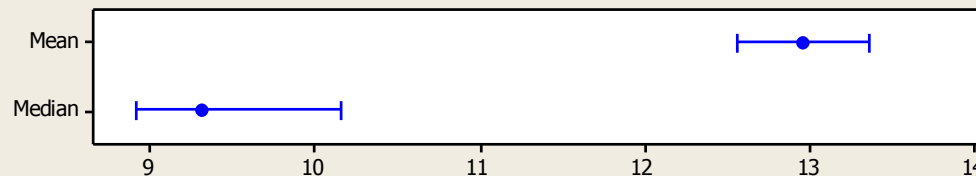
From Joseph Juran, Quality Control Handbook

# Some Things Just Aren't Normal

## Summary for July - Ground



### 95% Confidence Intervals



### Anderson-Darling Normality Test

A-Squared 285.69  
P-Value < 0.005

Mean 12.956  
StDev 12.674  
Variance 160.621  
Skewness 2.9189  
Kurtosis 11.0296  
N 3921

Minimum 2.650  
1st Quartile 4.210  
Median 9.320  
3rd Quartile 15.840  
Maximum 92.020

### 95% Confidence Interval for Mean

12.559 13.353

### 95% Confidence Interval for Median

8.920 10.160

### 95% Confidence Interval for StDev

12.399 12.961

# Testing For Normality

The screenshot shows the Minitab software interface. The 'Stat' menu is open, and the path 'Basic Statistics' > 'Normality Test...' is highlighted. The background worksheet contains data in columns C6 through C16. The session window on the left displays a histogram titled 'Results for: HISTO' and a probability plot.

**Stat Menu Options:**

- Basic Statistics
  - 1- Sample Z...
  - 1- Sample t...
  - 2- Sample t...
  - Paired t...
  - 1 Proportion...
  - 2 Proportions...
  - 1- Sample Poisson Rate...
  - 2- Sample Poisson Rate...
  - 1 Variance...
  - 2 Variances...
  - Correlation...
  - Covariance...
  - Normality Test...
  - Goodness-of-Fit Test for Poisson...
- Regression
- ANOVA
- DOE
- Control Charts
- Quality Tools
- Reliability/Survival
- Multivariate
- Time Series
- Tables
- Nonparametrics
- EDA
- Power and Sample Size

**Worksheet Data (Columns C6-C16):**

	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
9											
10											
11											
12											
13											
14											
15											
16	4	3	3								
17	9	5	-2								
18	6	9	1								
19	5	5	2								
20	10	10	-3								
21	8	6	-1								
22	5	7	2								
23	7	9	0								
24	3	4	4								
25	8	8	-1								
26	6	2	1								
27	10	10	-3								

**Session Window:**

Results for: HISTO

Probability Plot of

Percent

100

95

90

80

70

60

50

40

30

20

10

5

1

0 2

Test whether data follow a normal distribution

# A Study - Aerobes

## Aerobes

200

440

220

1100

80

55

60

1700

100

200

160

110

170

210

3600

90

85

35

430

180

5

15

1

30

10

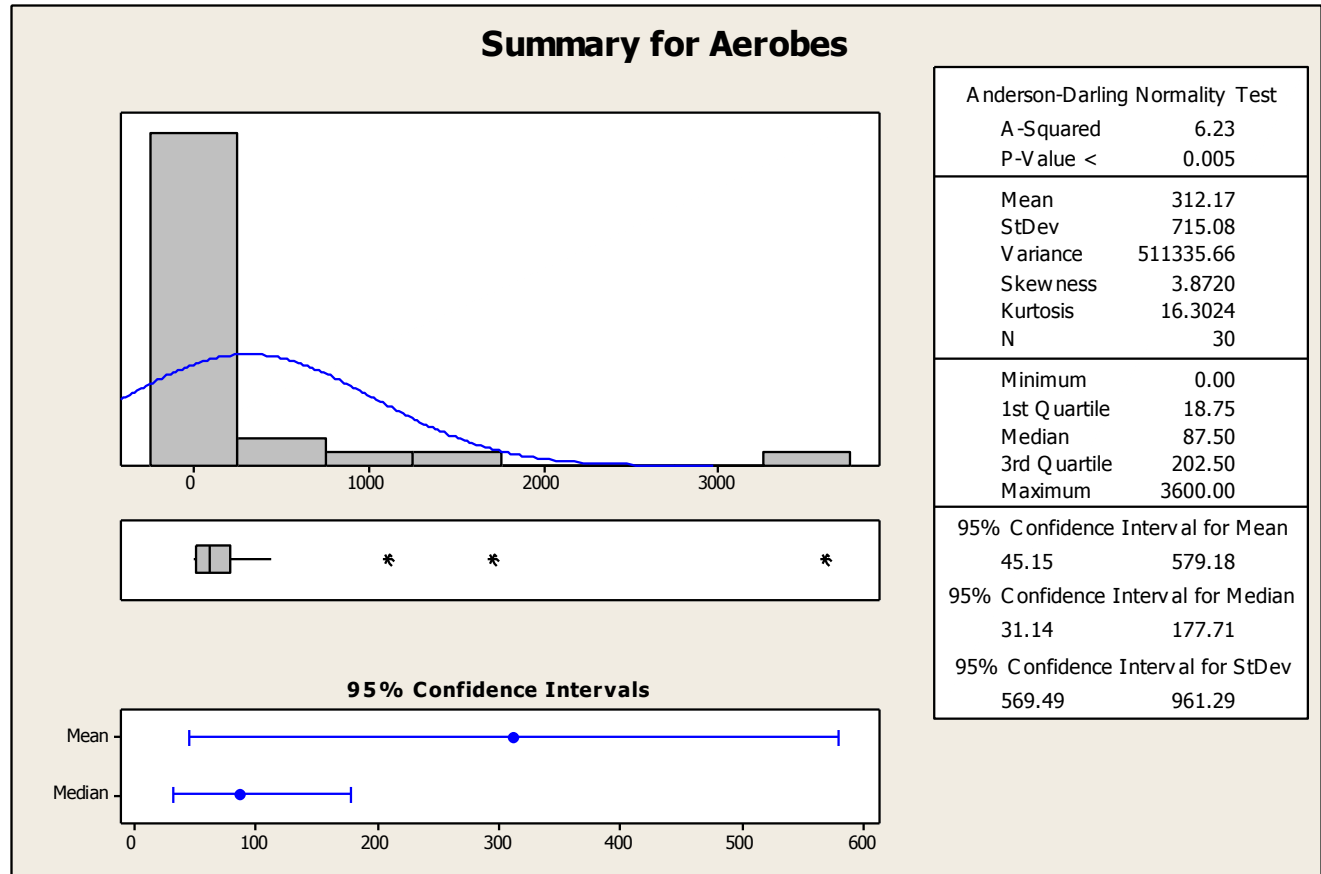
10

20

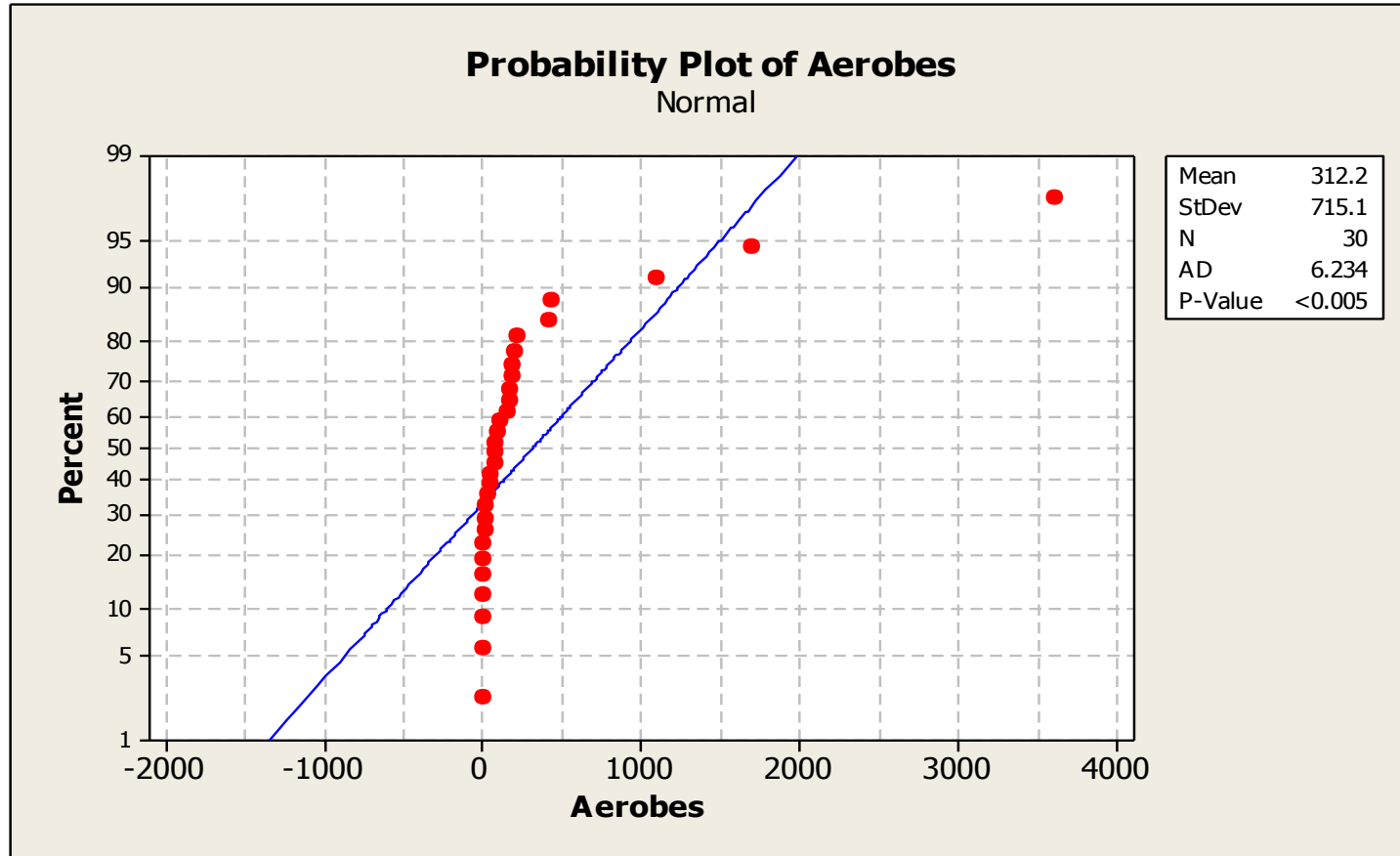
15

20

15



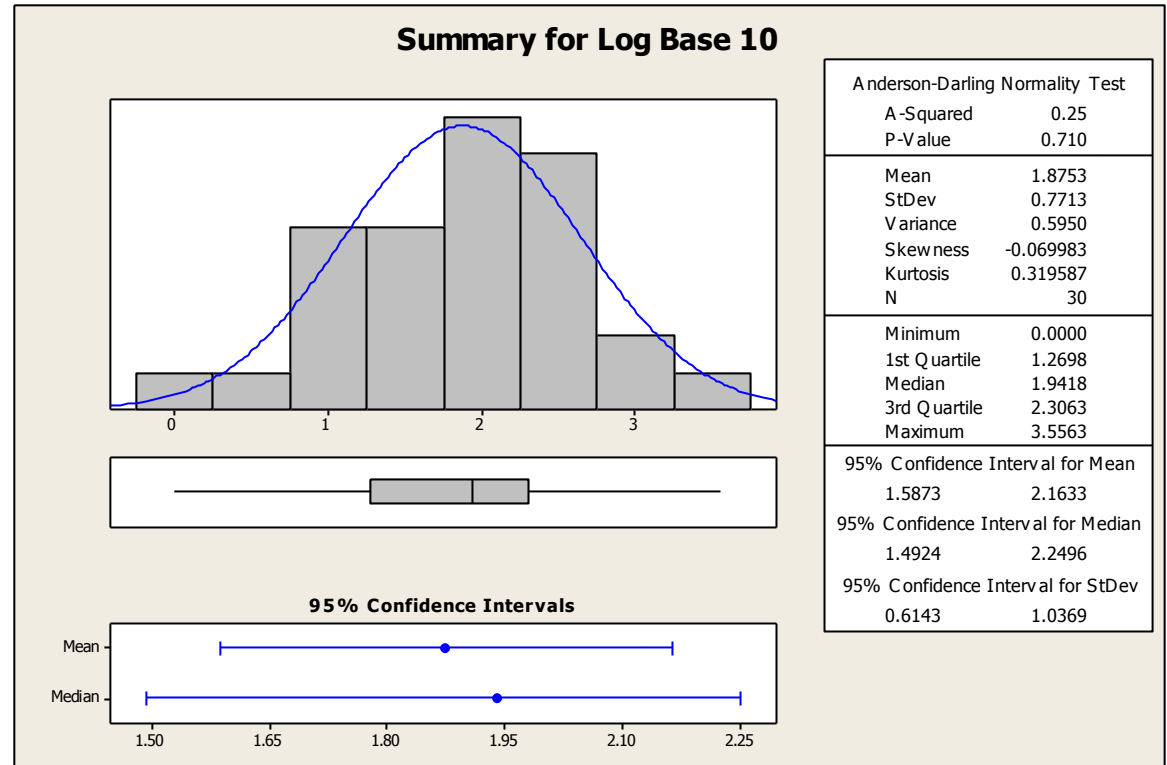
## Example of Need for Transformation



What the heck happened here?

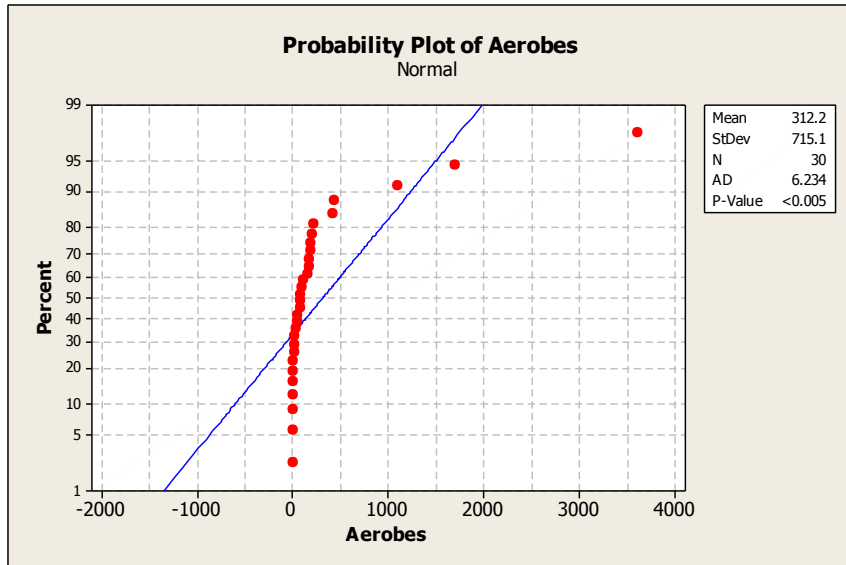
# Transformation to Log Values

	<u>Aerobes</u>	<u>Log10</u>	<u>Nat Log</u>
1.	200	2.30103	5.298317
2.	440	2.643453	6.086775
3.	220	2.342423	5.393628
4.	1100	3.041393	7.003065
5.	80	1.90309	4.382027
6.	55	1.740363	4.007333
7.	60	1.778151	4.094345
8.	1700	3.230449	7.438384
9.	100	2	4.60517
10.	200	2.30103	5.298317
11.	160	2.20412	5.075174
12.	110	2.041393	4.70048
13.	170	2.230449	5.135798
14.	210	2.322219	5.347108
15.	3600	3.556303	8.188689
16.	90	1.954243	4.49981
17.	85	1.929419	4.442651
18.	35	1.544068	3.555348
19.	430	2.633468	6.063785
20.	180	2.255273	5.192957
21.	5	0.69897	1.609438
22.	15	1.176091	2.70805
23.	1	0	0
24.	30	1.477121	3.401197
25.	10	1	2.302585
26.	10	1	2.302585
27.	20	1.30103	2.995732
28.	15	1.176091	2.70805
29.	20	1.30103	2.995732
30.	15	1.176091	2.70805

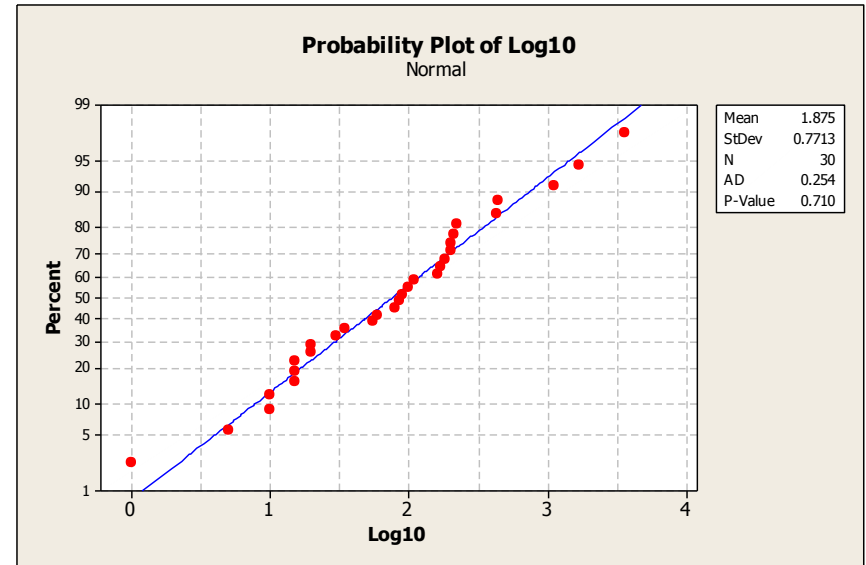




# Normality Test of a Transformation



Before



After

So how did we know to take the Log value in order to normalize our data?

# Minitab Application of Box Cox Transformation

Minitab - AEROBES.MPJ

File Edit Data Calc Stat Graph Editor Tools Window Help

Basic Statistics  
Regression  
ANOVA  
DOE  
Control Charts  
Quality Tools  
Reliability/Survival  
Multivariate  
Time Series  
Tables  
Nonparametrics  
EDA  
Power and Sample Size

Box-Cox Transformation...

Variables Charts for Subgroups  
Variables Charts for Individuals  
Attributes Charts  
Time-Weighted Charts  
Multivariate Charts

	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Nat Log	Limits	Log10Lim							
2.30103	5.29832	1000								
2.64345	6.08677	2000								
2.34242	5.39363									
3.04139	7.00307									
1.90309	4.38203									
55	1.74036	4.00733								
60	1.77815	4.09434								
1700	3.23045	7.43838								
9	100	4.60517								
10	200	5.29832								
11	160	5.07517								
12	110	4.70048								
13	170	5.13580								
14	210	5.34711								
15	3600	8.18869								
16	90	4.49981								
17	85	4.44265								
18	35	3.55535								
19	430	6.06379								
20	180	5.19296								
21	5	1.60944								
22	15	2.70805								
23	1	0.00000								
24	30	3.40120								
25	10	2.30259								
26	10	2.30259								
27	20	2.99573								
28	15	2.70805								
29	20	2.99573								
30	15	2.70805								
31										
32										
33										
34										
35										
36										
37										

Perform a Box-Cox transformation of nonnormal process data

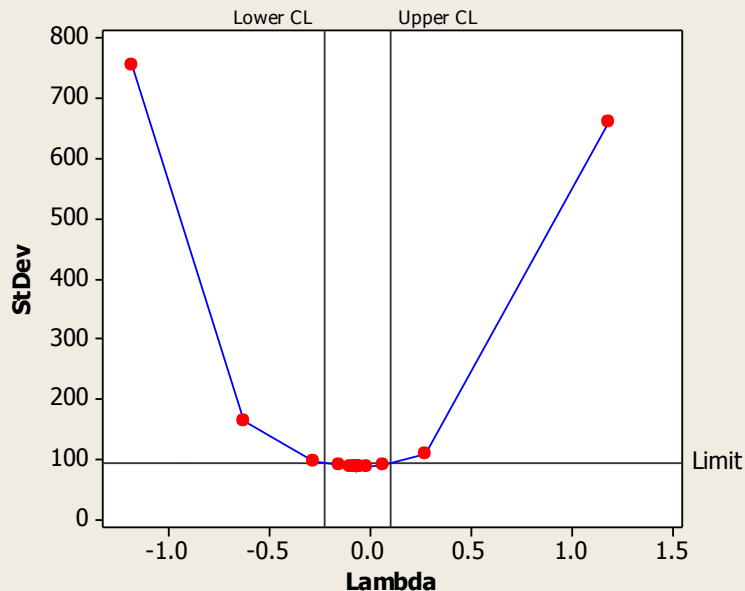
Start | Inbox - Microsoft ... | ev3Training | My Documents | Statistical Analysis ... | Microsoft PowerPoi... | Minitab - AEROB... | Minitab - AEROBES... | 9:50 AM

# Box Cox Transformation

A transformation was considered in order to normalize the data and produce a more accurate and predictable distribution pattern. The Box-Cox Plot of the data was generated to assist in the transformation process. The plot was produced for the data as follows:

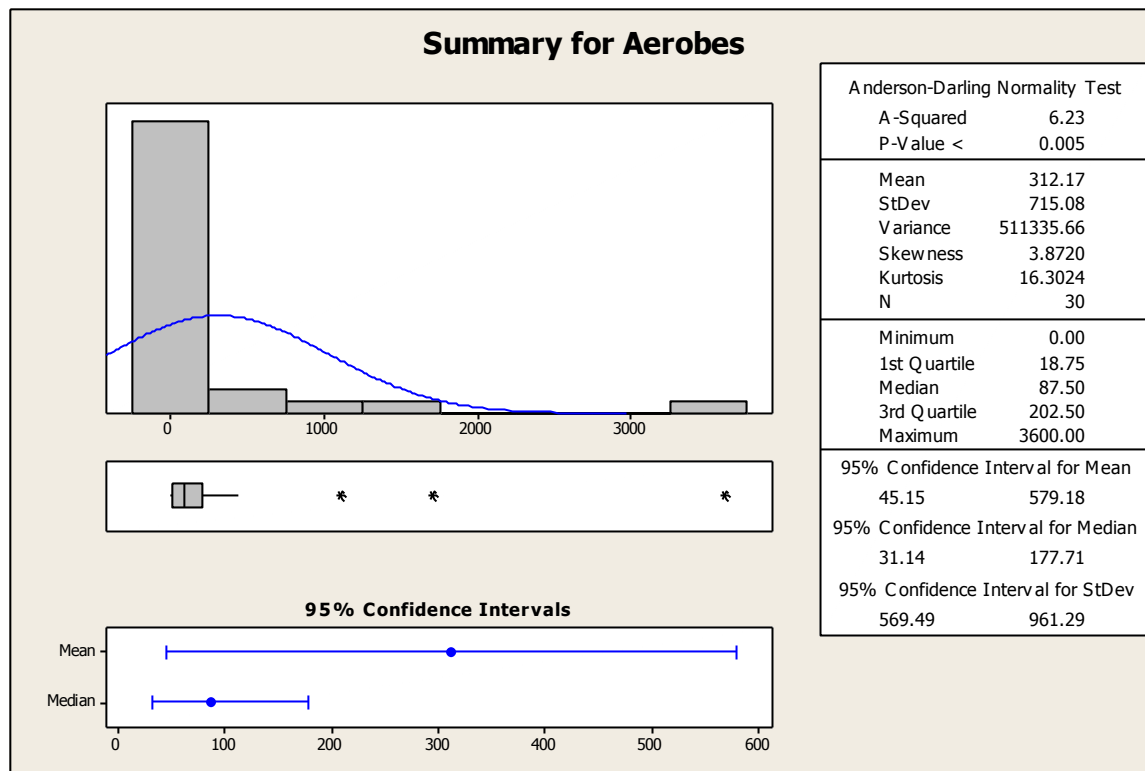
$\lambda$ (Power)	$Y^\lambda$	Common Names
-2	$\frac{1}{Y^2}$	Reciprocal (inverse) squared
-1	$\frac{1}{Y}$	Reciprocal (inverse)
-0.5	$\frac{1}{\sqrt{Y}}$	Reciprocal square root (inverse)
0	$\ln(Y)$	Log
0.5	$\sqrt{Y}$	Square root
1	No transformation	---
2	$Y^2$	Squared

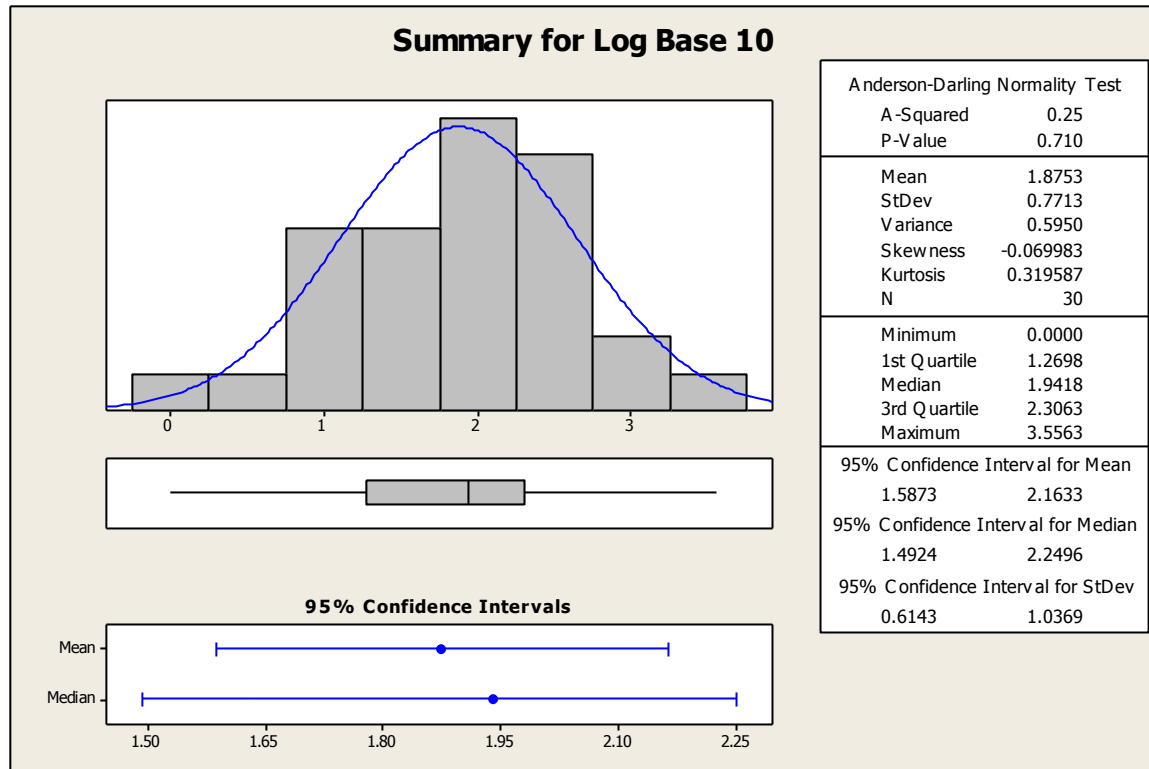
**Box-Cox Plot of Aerobes**



The Rounded Value for this transformation was calculated at 0.00 thus suggesting a Log ( $\ln(Y)$ ) Transformation per the Box-Cox Methodology. This again was not atypical since microbial data frequently exhibits itself in an exponential manner. Both a Natural Log and a Log Base 10 transformation were calculated on the original data set using the calculator function in Minitab 15.

Data set – 200, 440, 220, 1100, 80, 55, 60, 1700, 100, 200, 160, 110, 170, 210, 3600, 90, 85, 35, 430, 180, 5, 15, 1, 30, 10, 10, 20, 15, 20, 15.





## Conclusion

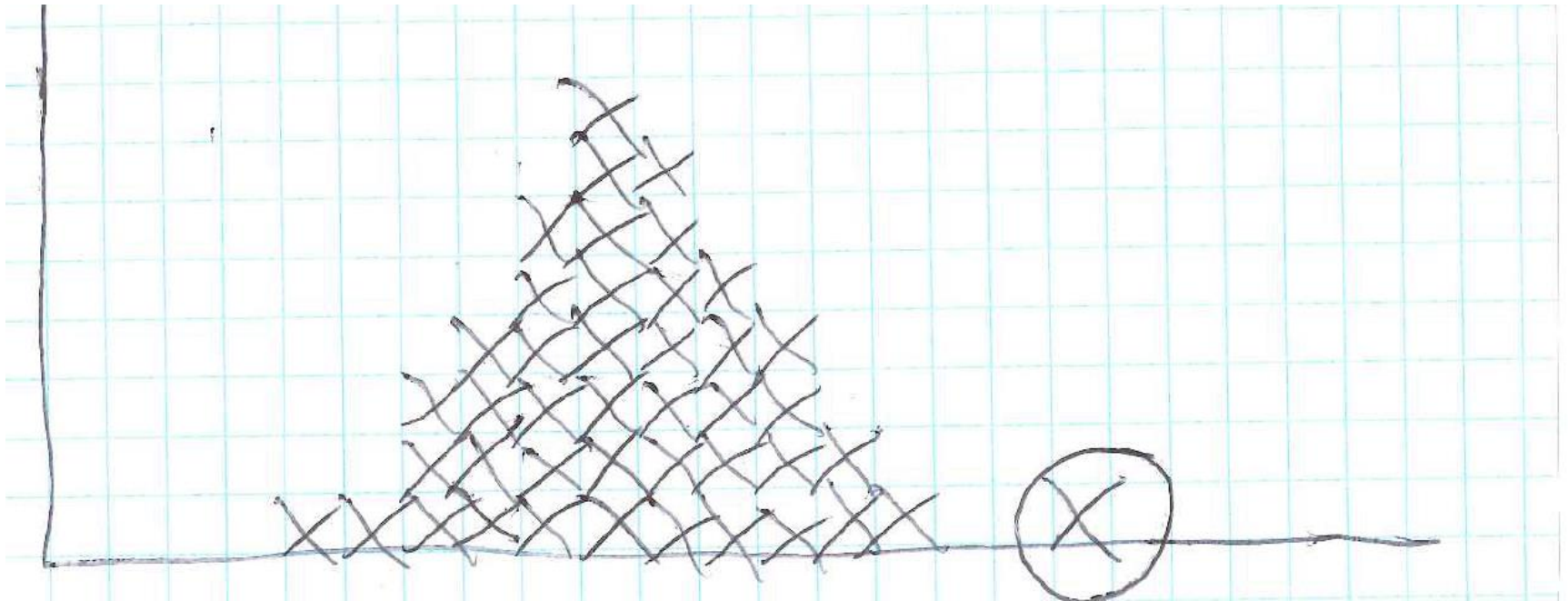
The sample with a reading of 3600 is expected to occur approximately 1 out of every 67 readings on an average and should not be considered extreme based on the distribution of the data. The sample with this count of Aerobes should be identified through a gram stain to determine the nature of its source, but unless abnormal concerns are raised based on the type of microorganism, no additional action should be taken.

## And if you can't do a valid transformation

### Common Nonparametric Tests

NONPARAMETRIC	FUNCTION	PARAMETRIC
1-Sample Sign*	Test the median for 1 sample Test difference in <b>dependent</b> samples	t test or z test
1-Sample Wilcoxon*	Test the median difference in <b>dependent</b> samples	t test or z test
Mann-Whitney*	Test the median difference in 2 <b>independent</b> samples	t test or z test
Chi-Square/ Spearman's Rank	Test <b>relationships</b>	Correlation / Regression
Kruskal-Wallis/ Mood's Median*	Test the median difference in <b>many independent</b> samples	ANOVA
Siegel-Tukey (not in Minitab)	Test differences in spread of 2 <b>independent</b> samples	F test
Friedman	Test if <b>two factors</b> are significant	2 – way ANOVA
Runs Test	Test for <b>Randomness</b>	No equivalent

# Outliers



The “Oh Crap” data point aka the Outlier



- An outlier is an observation point that is distant from other observations.
- An outlier may be due to variability in the measurement or it may indicate experimental error: the latter are sometimes excluded from the data set.
- An outlier may be caused by a defective unit or a problem in the process.
- ISO 16269 defines it as “member of a small subset of observations that appears to be inconsistent with the remainder of a given sample.”

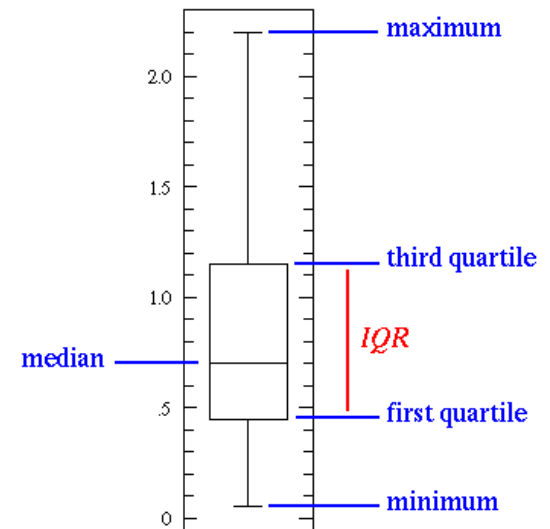
- There is no rigid mathematical definition of what constitutes an outlier; determining whether or not an observation is an outlier is ultimately a subjective exercise.
- Every effort must be made to determine what is causing the outlier to exist. Testing, equipment, operator error, materials, etc. all must be reviewed and assessment made.

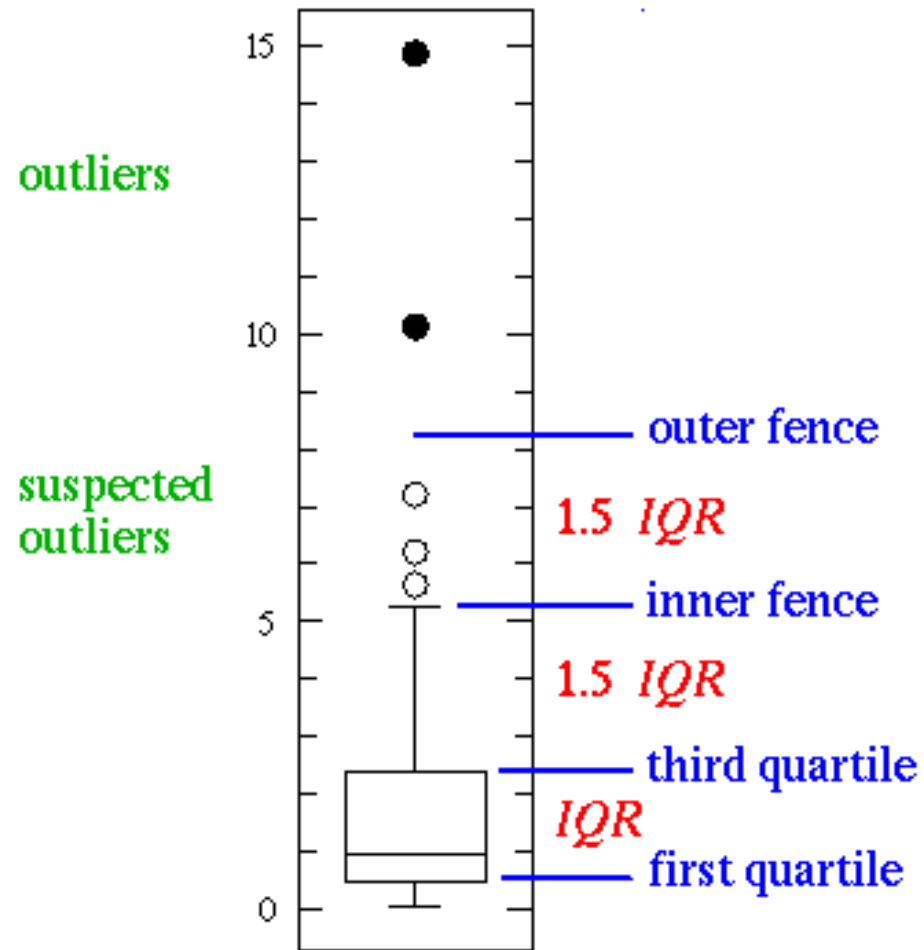
- ASTM E178-08 “Standard Practice for Dealing with Outlying Observations”
- ISO 16269-4:2010 “Statistical interpretation of data – Part 4: Detection and treatment of outliers”

- Model-based methods which are commonly used for identification assume that the data are from a normal distribution, and identify observations which are deemed “unlikely” based on mean and standard deviation:
  - Chauvenet’s criterion
  - Grubbs’ test for outliers
  - Peircece’s criterion

- Other methods flag observations based on measures such as the interquartile range. For example, if  $Q_1$  and  $Q_3$  are the lowest and upper quartiles respectively, then one could define an outlier to be any observation outside the range:

$$[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$$





<http://www.physics.csbsju.edu/stats/box2.html>

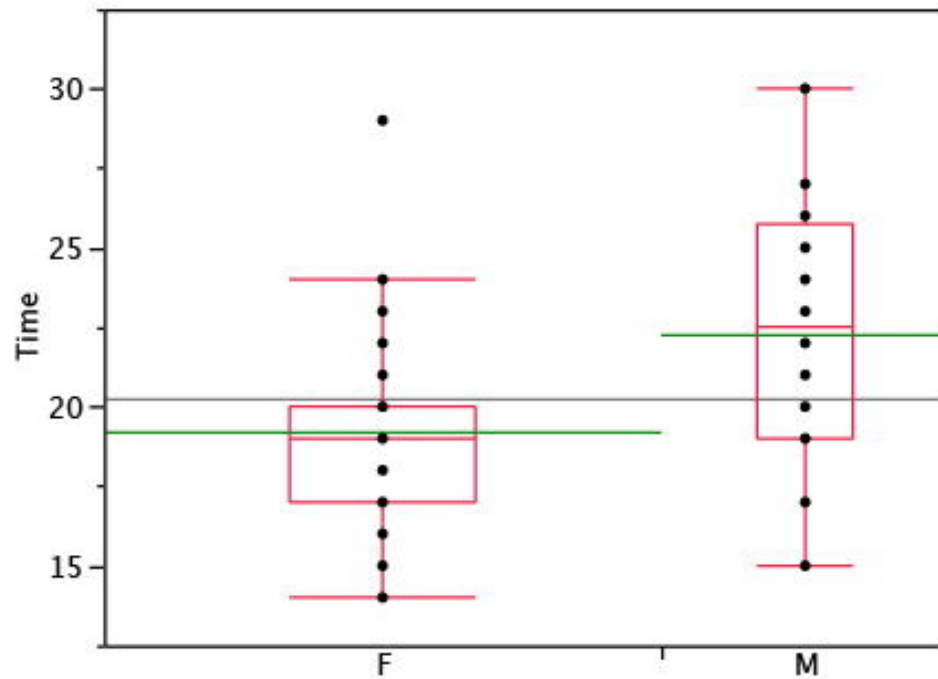
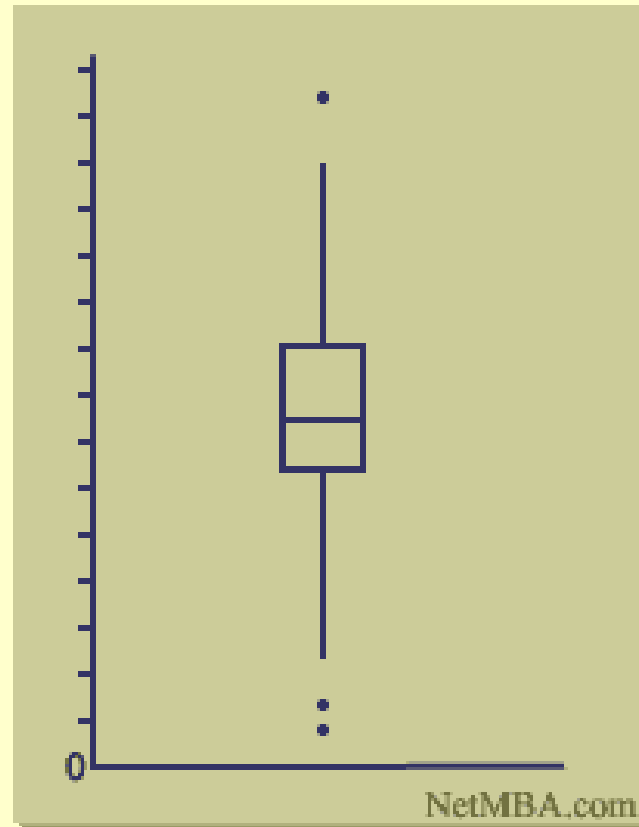


Figure 6. Box plots showing the individual scores and the means.

- [http://onlinestatbook.com/2/graphing\\_distributions/boxplots.html](http://onlinestatbook.com/2/graphing_distributions/boxplots.html)



The plot may be drawn either vertically as in the above diagram, or horizontally.

<http://www.netmba.com/statistics/plot/box/>

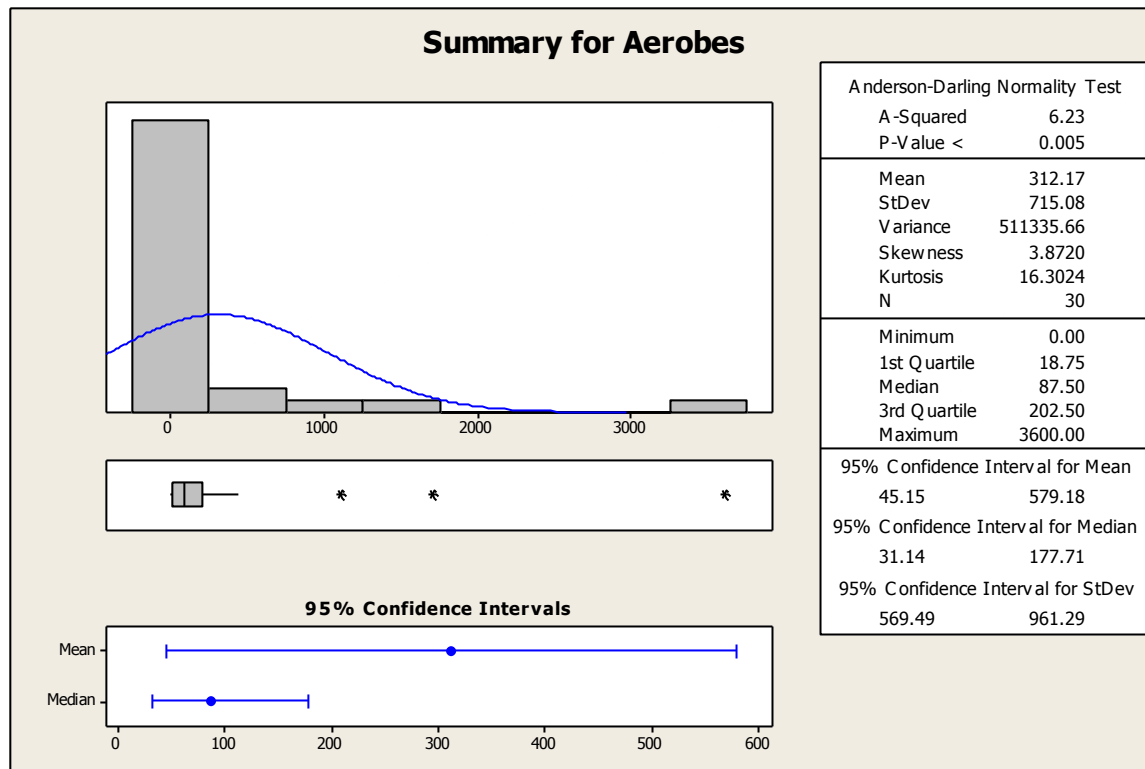


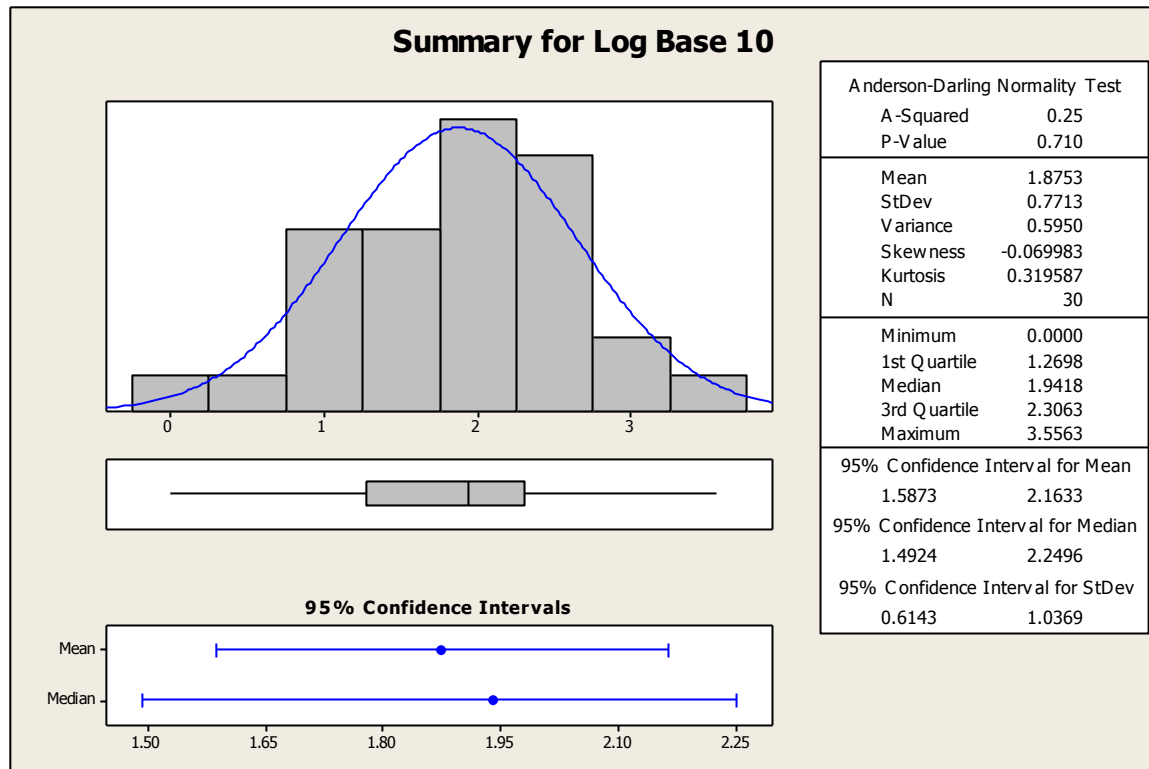
The boxplot is interpreted as follows:

- The box itself contains the middle 50% of the data. The upper edge (hinge) of the box indicates the 75th percentile of the data set, and the lower hinge indicates the 25th percentile. The range of the middle two quartiles is known as the inter-quartile range.
- The line in the box indicates the median value of the data.
- If the median line within the box is not equidistant from the hinges, then the data is skewed.
- The ends of the vertical lines or "whiskers" indicate the minimum and maximum data values, unless outliers are present in which case the whiskers extend to a maximum of 1.5 times the inter-quartile range.
- The points outside the ends of the whiskers are outliers or suspected outliers.

# Outlier determination can be tricky!

Data set – 200, 440, 220, 1100, 80, 55, 60, 1700, 100, 200, 160, 110, 170, 210, 3600, 90, 85, 35, 430, 180, 5, 15, 1, 30, 10, 10, 20, 15, 20, 15.





## **Conclusion**

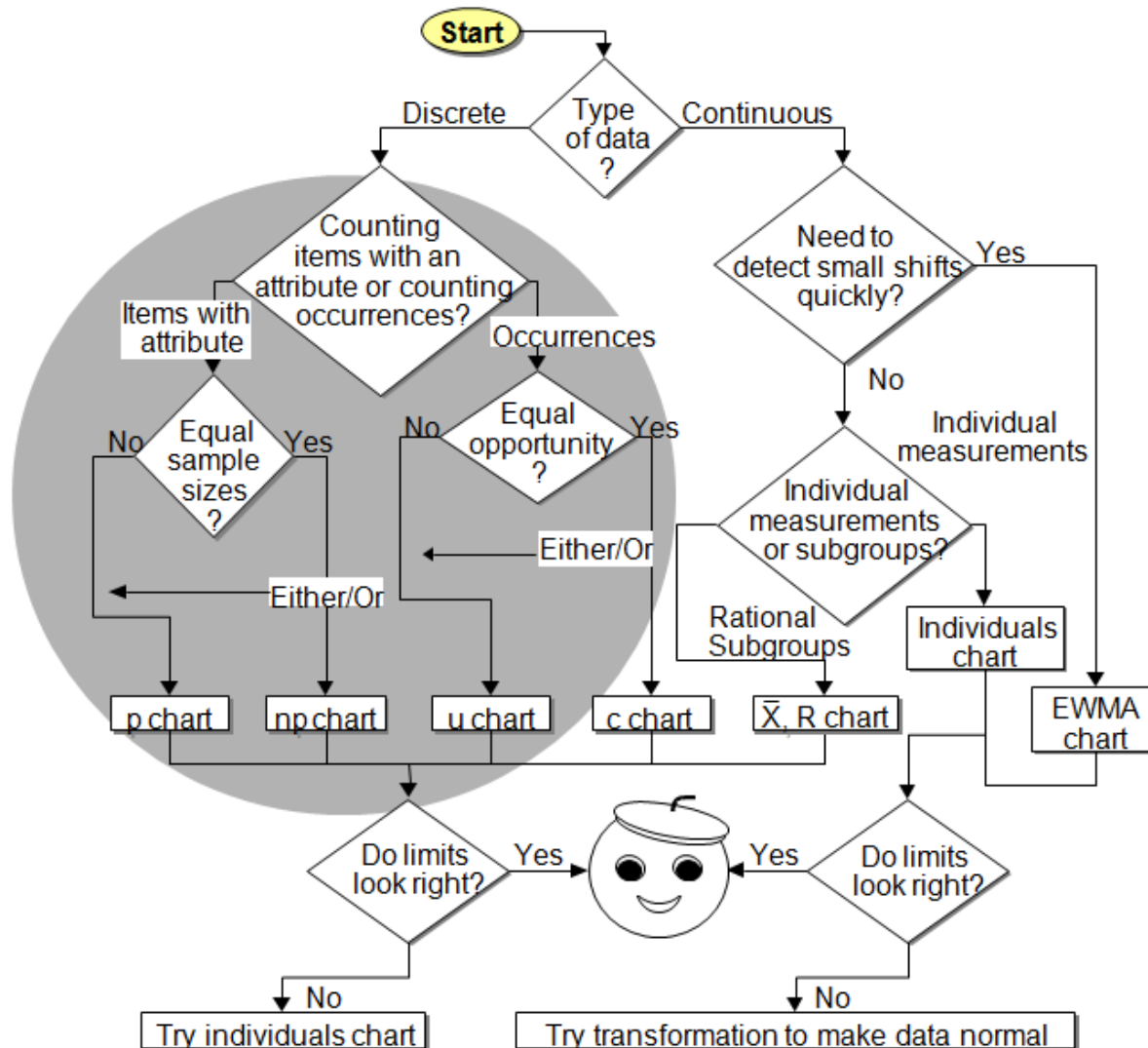
The sample with a reading of 3600 is expected to occur approximately 1 out of every 67 readings on an average and should not be considered extreme based on the distribution of the data. The sample with this count of Aerobes should be identified through a gram stain to determine the nature of its source, but unless abnormal concerns are raised based on the type of microorganism, no additional action should be taken.

- Outliers must be investigated
- Use the tools found in root cause analysis
  - Interview operators
  - Evaluate the measuring system
  - Try to duplicate the reading
- Don't make assumptions or jump to conclusions
- Do not remove outliers simply because they are outliers

Note that outliers are not necessarily "bad" data-points; indeed they may well be the most important, most information rich, part of the dataset. Under no circumstances should they be automatically removed from the dataset. Outliers may deserve special consideration: they may be the key to the phenomenon under study or the result of human blunders.

# Control Charts

# The World of Control Charts



# Control Charts

The screenshot shows the Minitab software interface. The menu bar includes File, Edit, Data, Calc, Stat, Graph, Editor, Tools, Window, and Help. The 'Stat' menu is open, showing options like Basic Statistics, Regression, ANOVA, DOE, Control Charts, Quality Tools, Reliability/Survival, Multivariate, Time Series, Tables, Nonparametrics, EDA, and Power and Sample Size. The 'Control Charts' menu is further expanded, showing 'Variables Charts for Individuals' selected, which then opens a sub-menu with 'I-MR...', 'Z-MR...', 'Individuals...', and 'Moving Range...'. The main workspace is a grid with columns labeled C5 through C16 and rows labeled 8 through 25. A small rectangular box is present in the grid at row 18, column C8. The status bar at the bottom indicates 'Welcome to Minitab, press F1 for help.' and shows the taskbar with various open applications including 'Minitab - Untitled', 'ev3Training', 'Black Belt', 'ControlCharts.ppt', and 'PE\_BBT\_5.1\_Monitor\_11...'. The system clock shows 7:18 PM.



## Test to Determine Abnormal Situations (Special Causes)

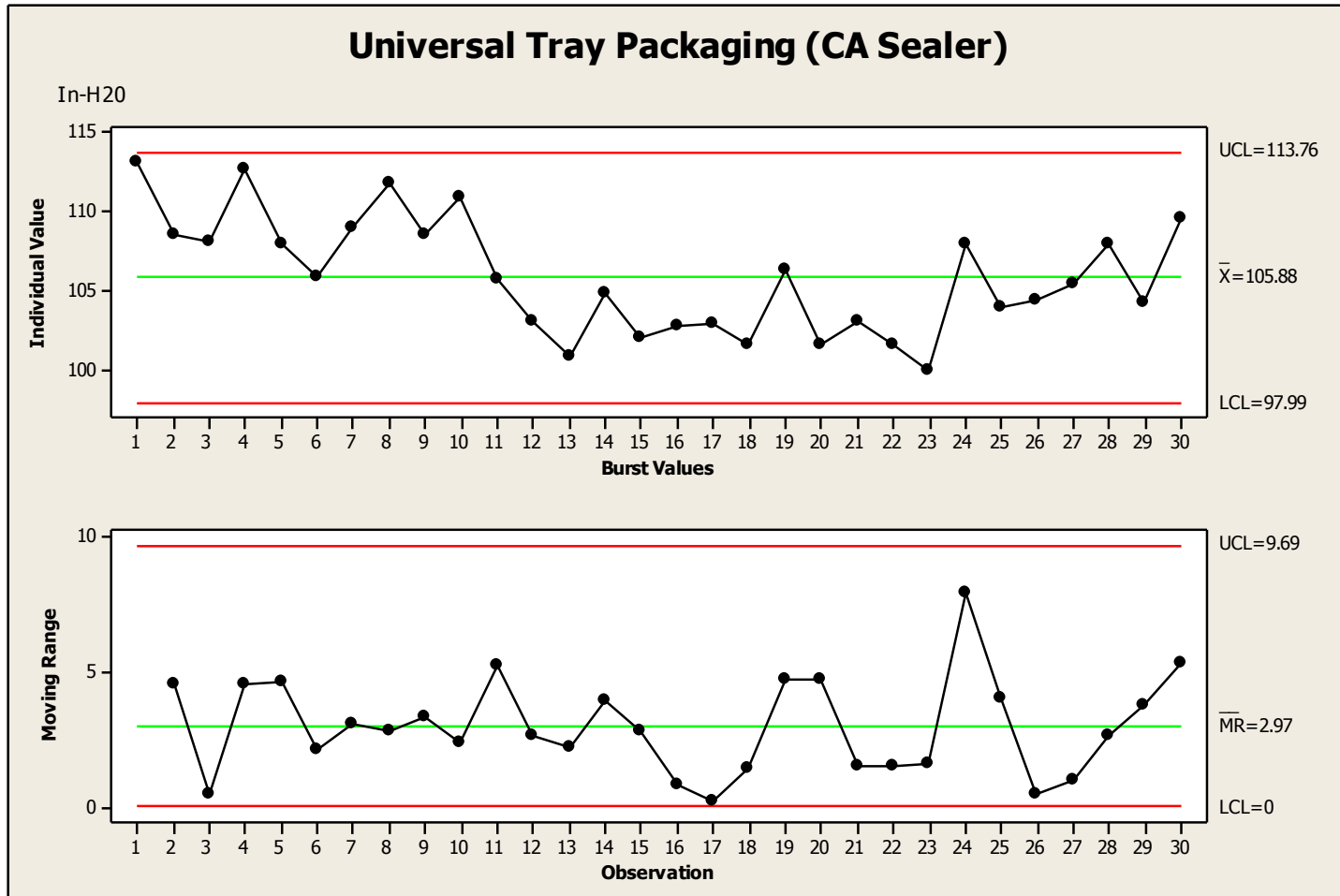
**Individuals Chart - Options** [X]

Parameters | Estimate | S Limits | Tests | Stages | Box-Cox | Display | Storage

Perform selected tests for special causes [v] K

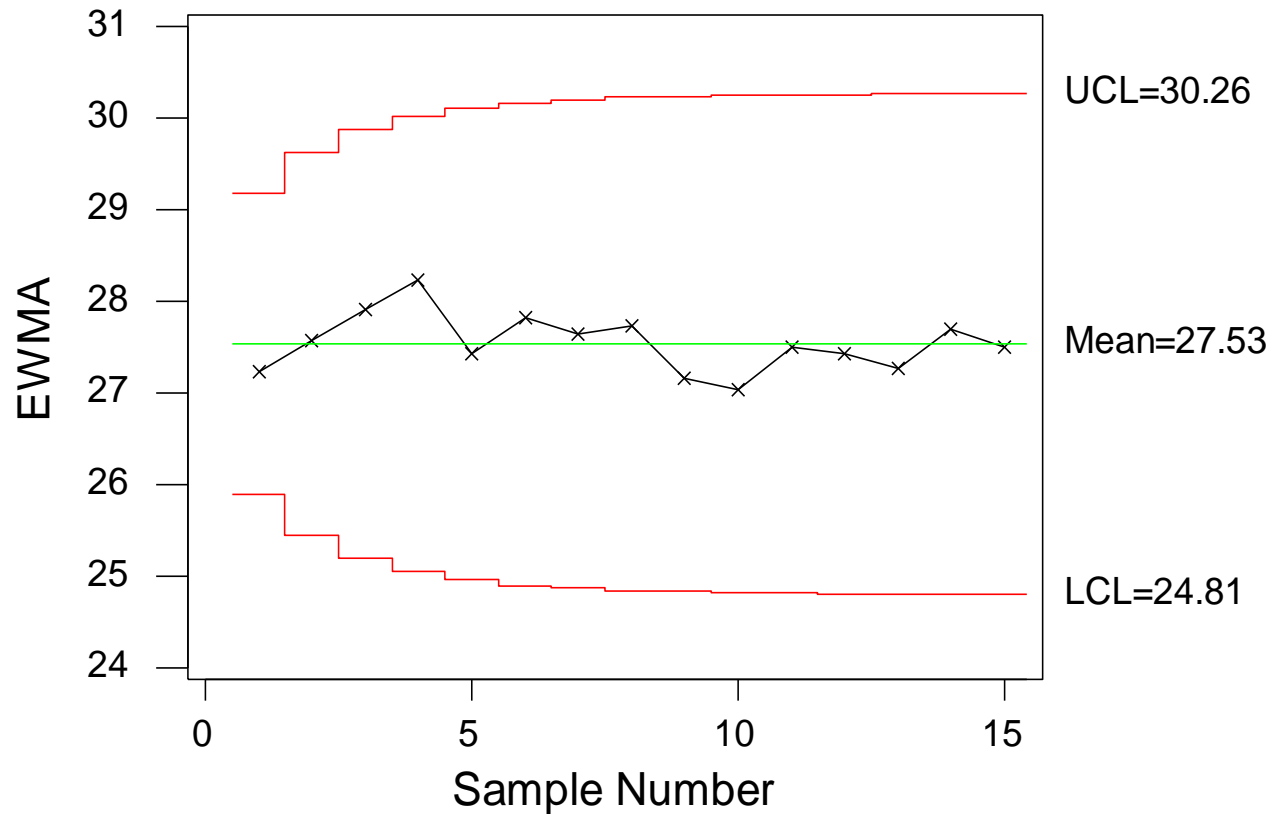
<input checked="" type="checkbox"/> 1 point > K standard deviations from center line	3.0
<input type="checkbox"/> K points in a row on same side of center line	9
<input type="checkbox"/> K points in a row, all increasing or all decreasing	6
<input type="checkbox"/> K points in a row, alternating up and down	14
<input type="checkbox"/> K out of K+1 points > 2 standard deviations from center line (same side)	2
<input type="checkbox"/> K out of K+1 points > 1 standard deviation from center line (same side)	4
<input type="checkbox"/> K points in a row within 1 standard deviation of center line (either side)	15
<input type="checkbox"/> K points in a row > 1 standard deviation from center line (either side)	8

# Individual and Moving Range Chart Example



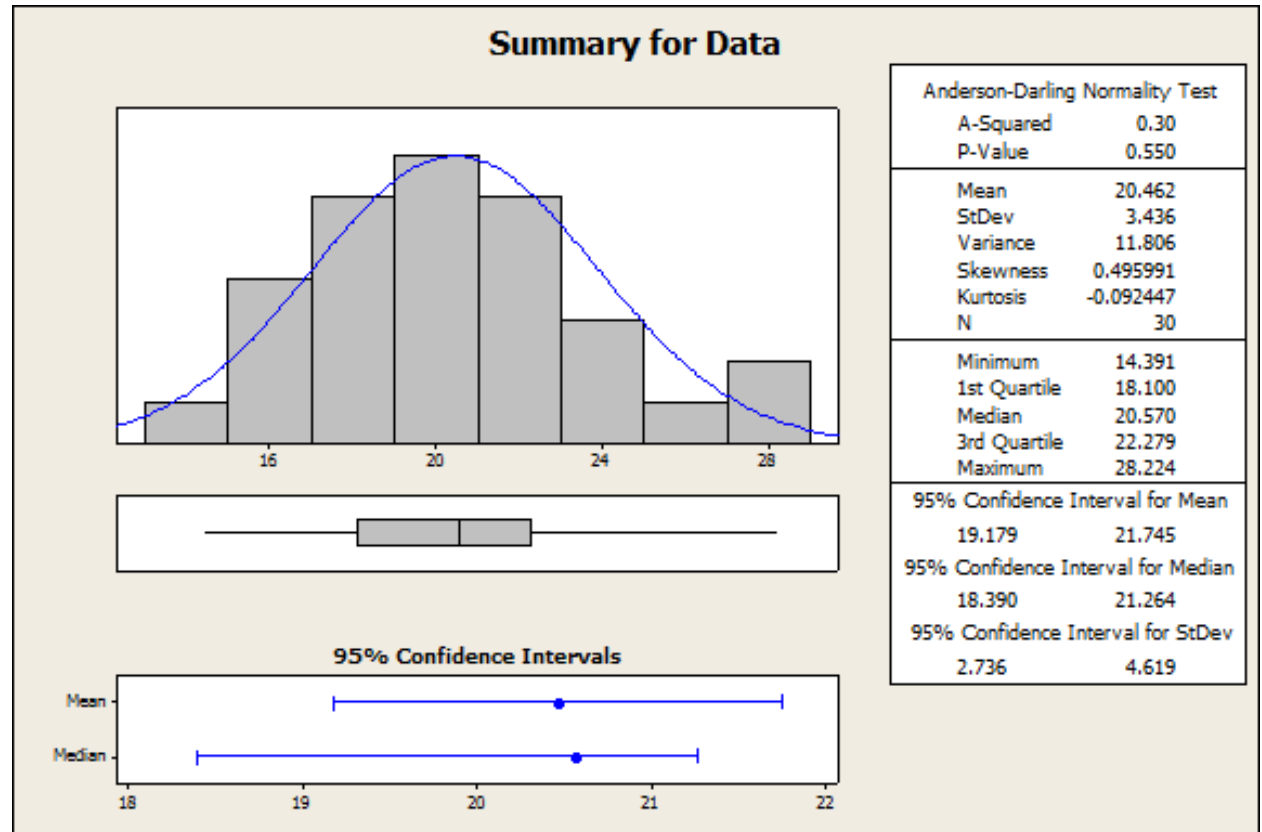
## Small Shifts – EWMA Chart

EWMA Chart for Contract Cycle Time

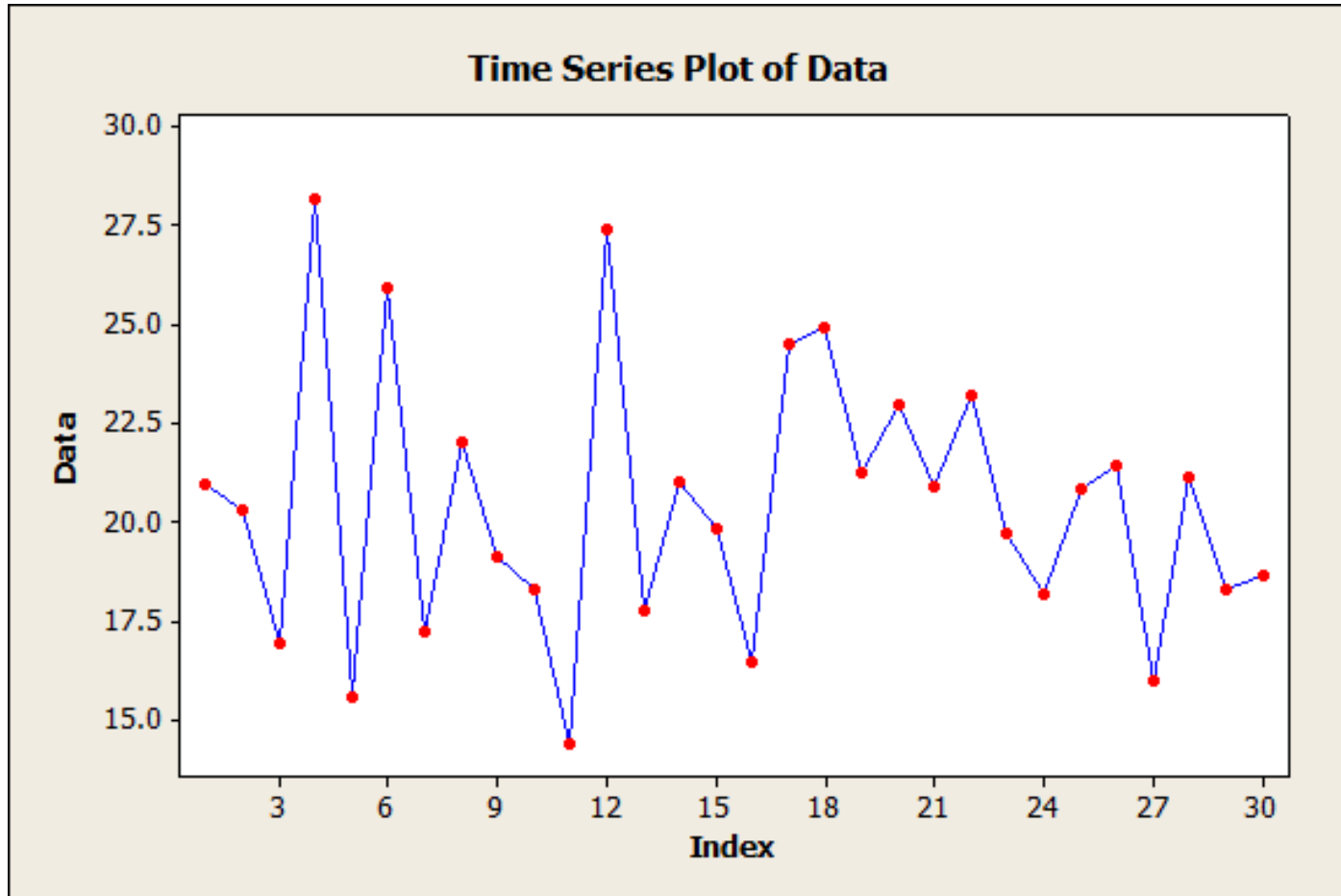


# Normal Data for Data Set 1

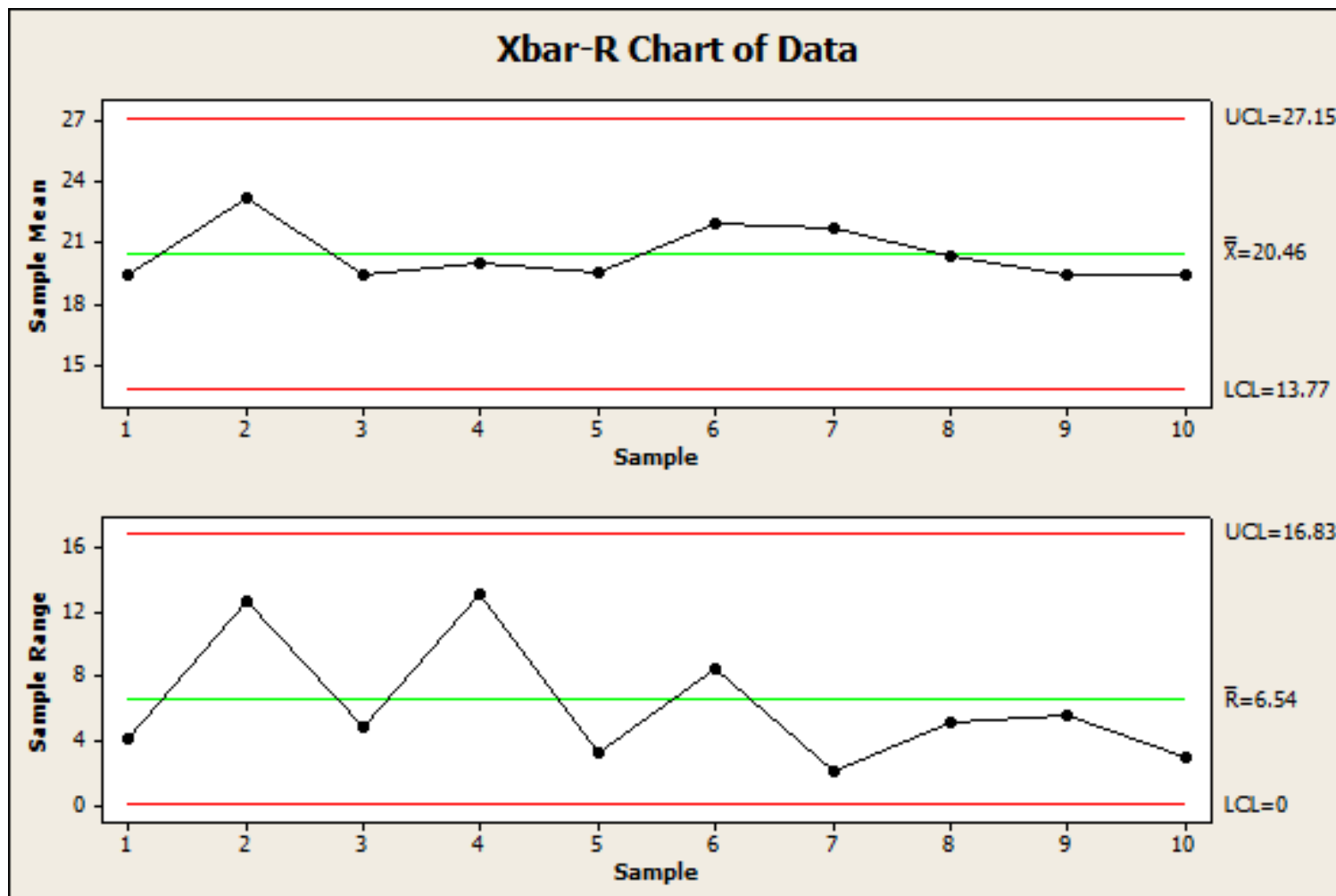
20.99	16.46
20.32	24.53
16.95	24.95
28.22	21.29
15.61	22.98
25.96	20.91
17.21	23.23
22.05	19.74
19.10	18.21
18.31	20.83
14.39	21.46
27.43	16.00
17.77	21.17
21.02	18.32
19.85	18.63



# Time Series Plot for Data 1

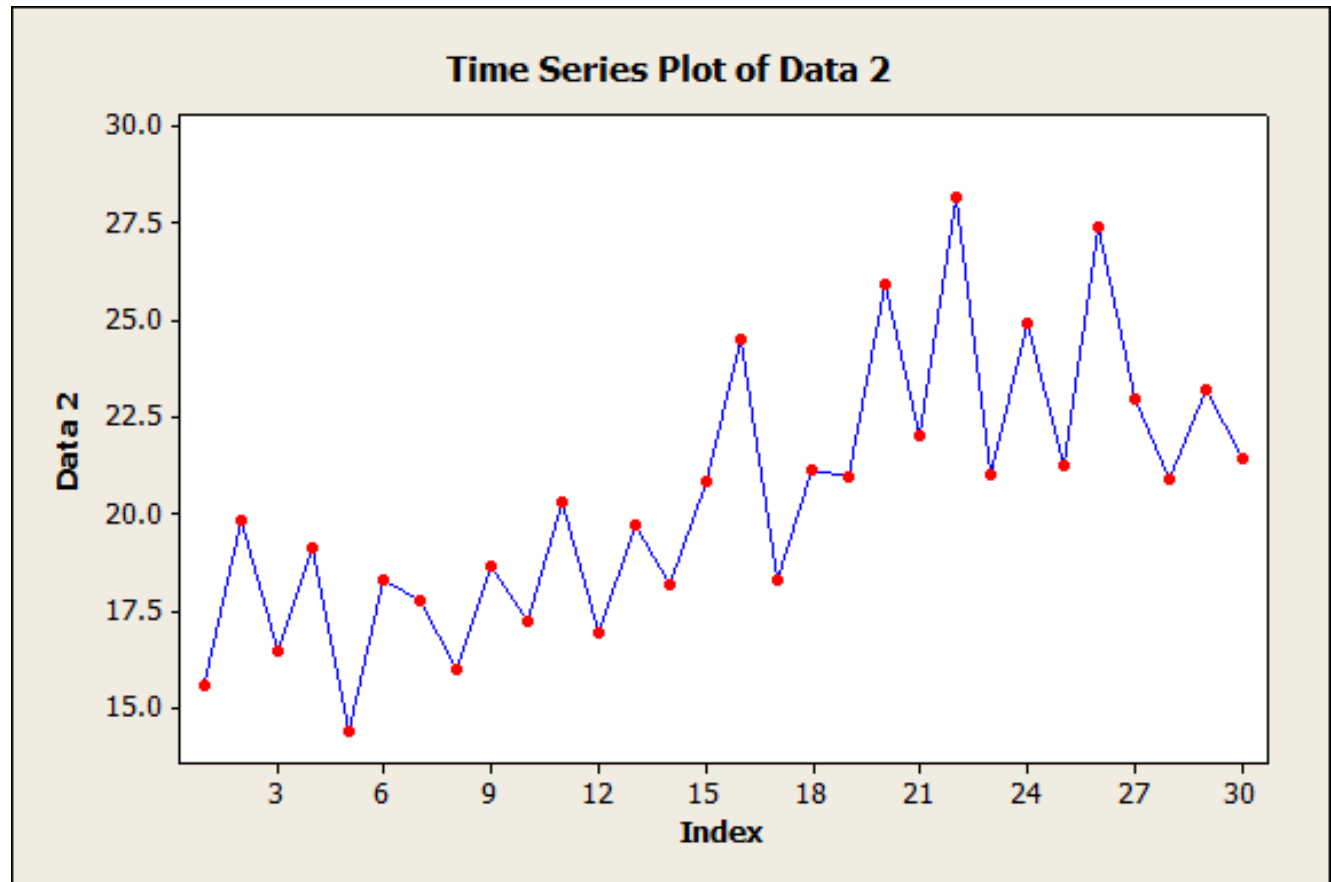


# X Bar & R for Data Set 1

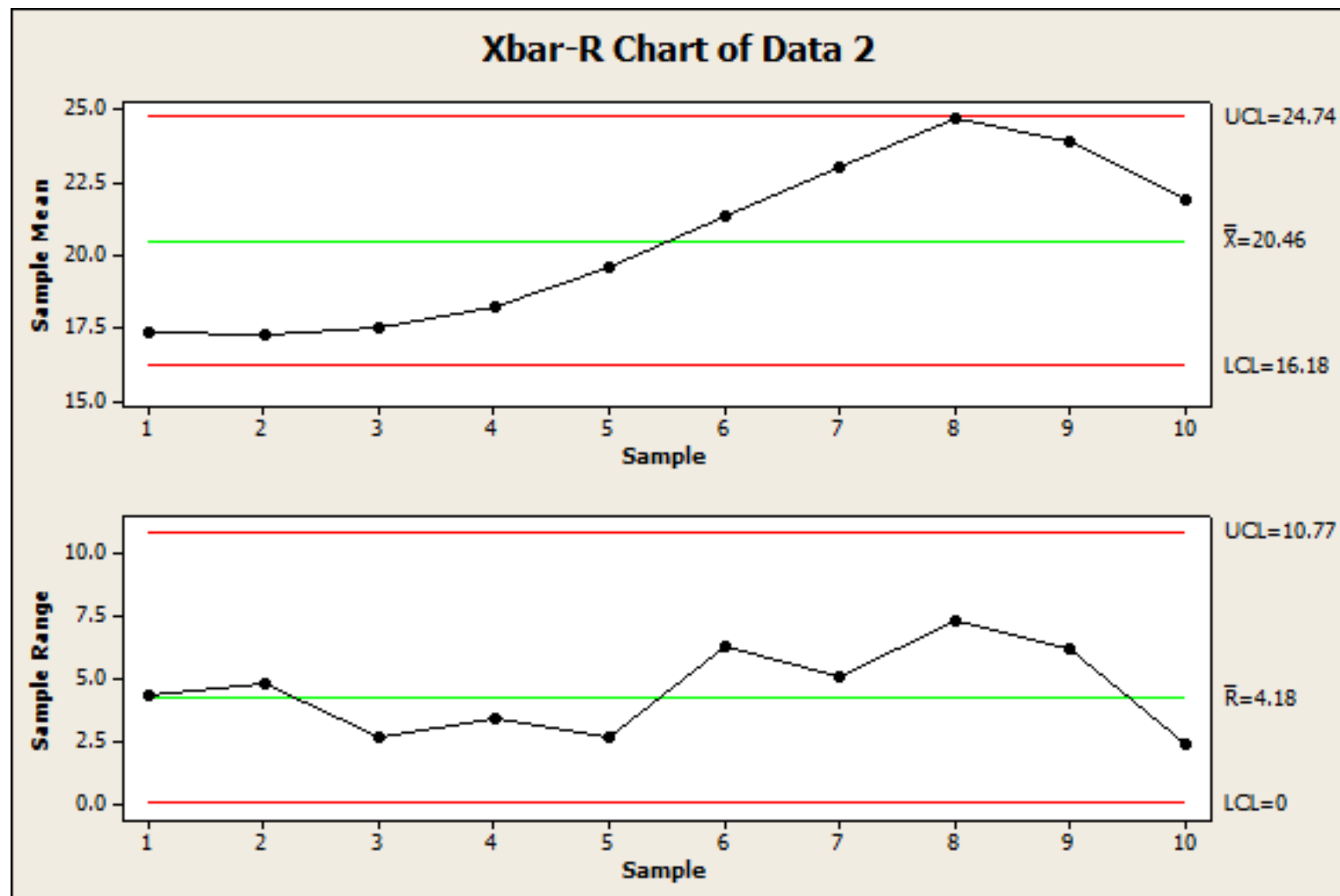


## Time Series for Data Set 2

15.61	24.53
19.85	18.32
16.46	21.17
19.10	20.99
14.39	25.96
18.31	22.05
17.77	28.22
16.00	21.02
18.63	24.95
17.21	21.29
20.32	27.43
16.95	22.98
19.74	20.91
18.21	23.23
20.83	21.46



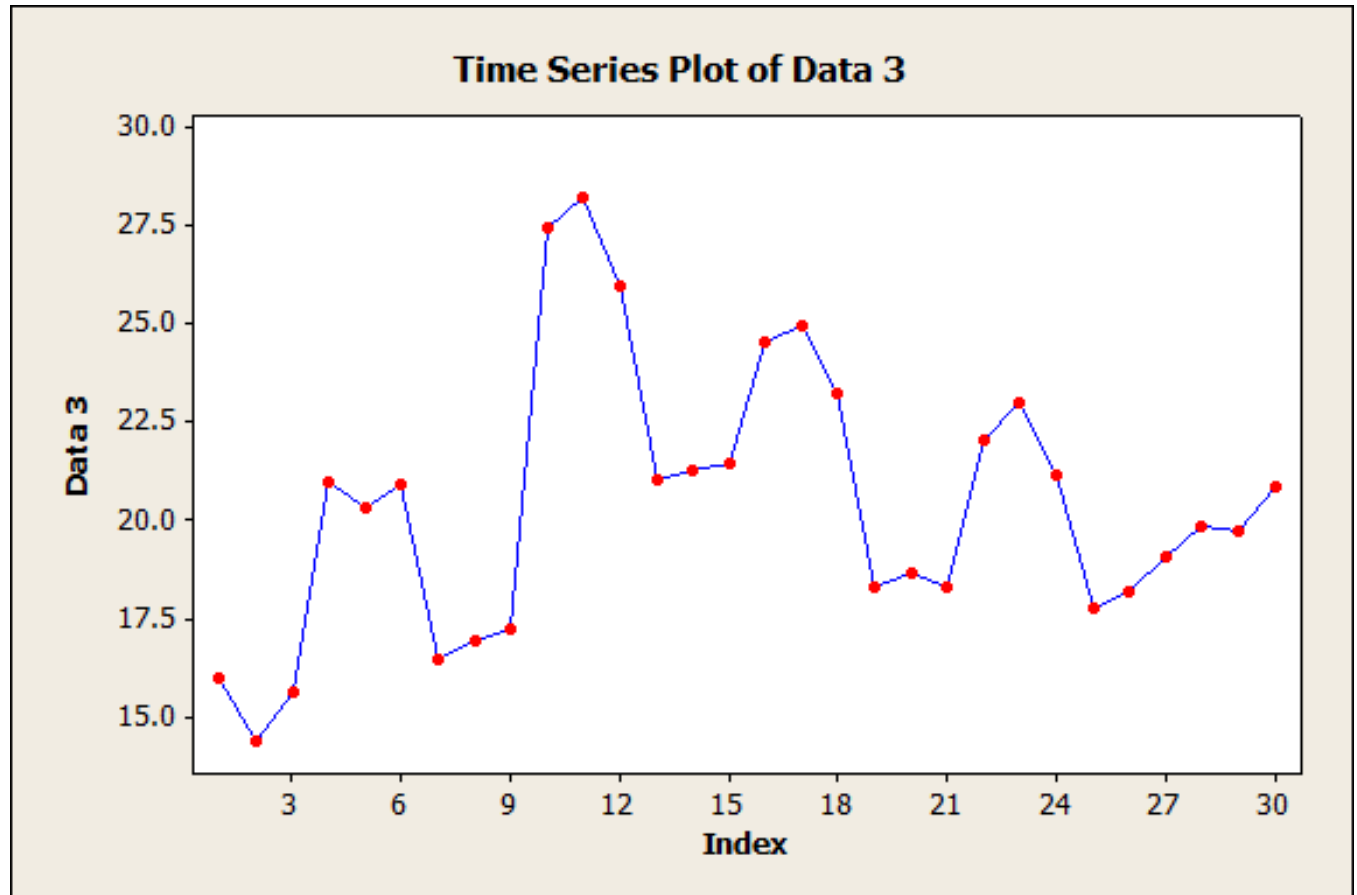
## X Bar & R for Data Set 2



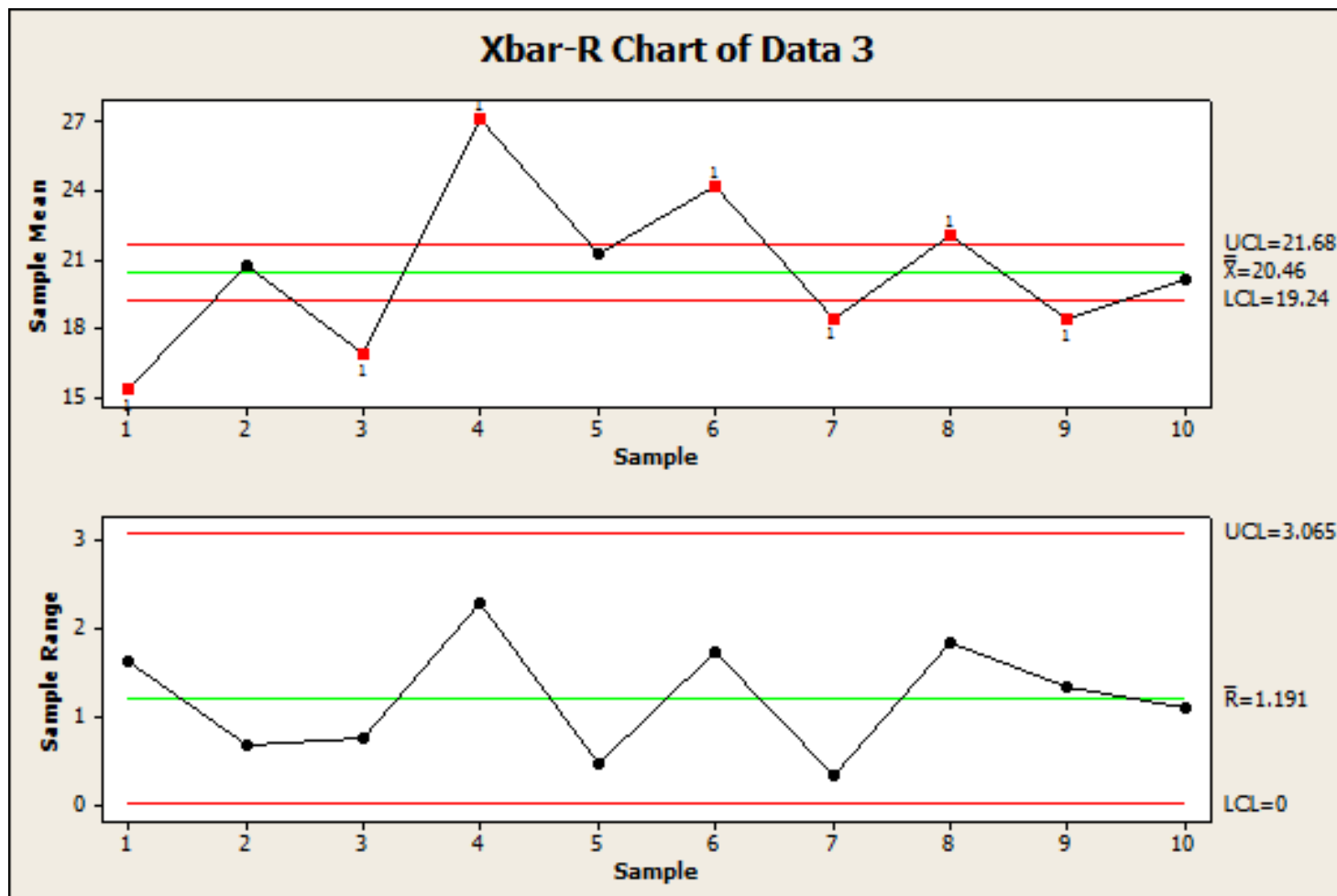


## Time Series Plot for Data 3

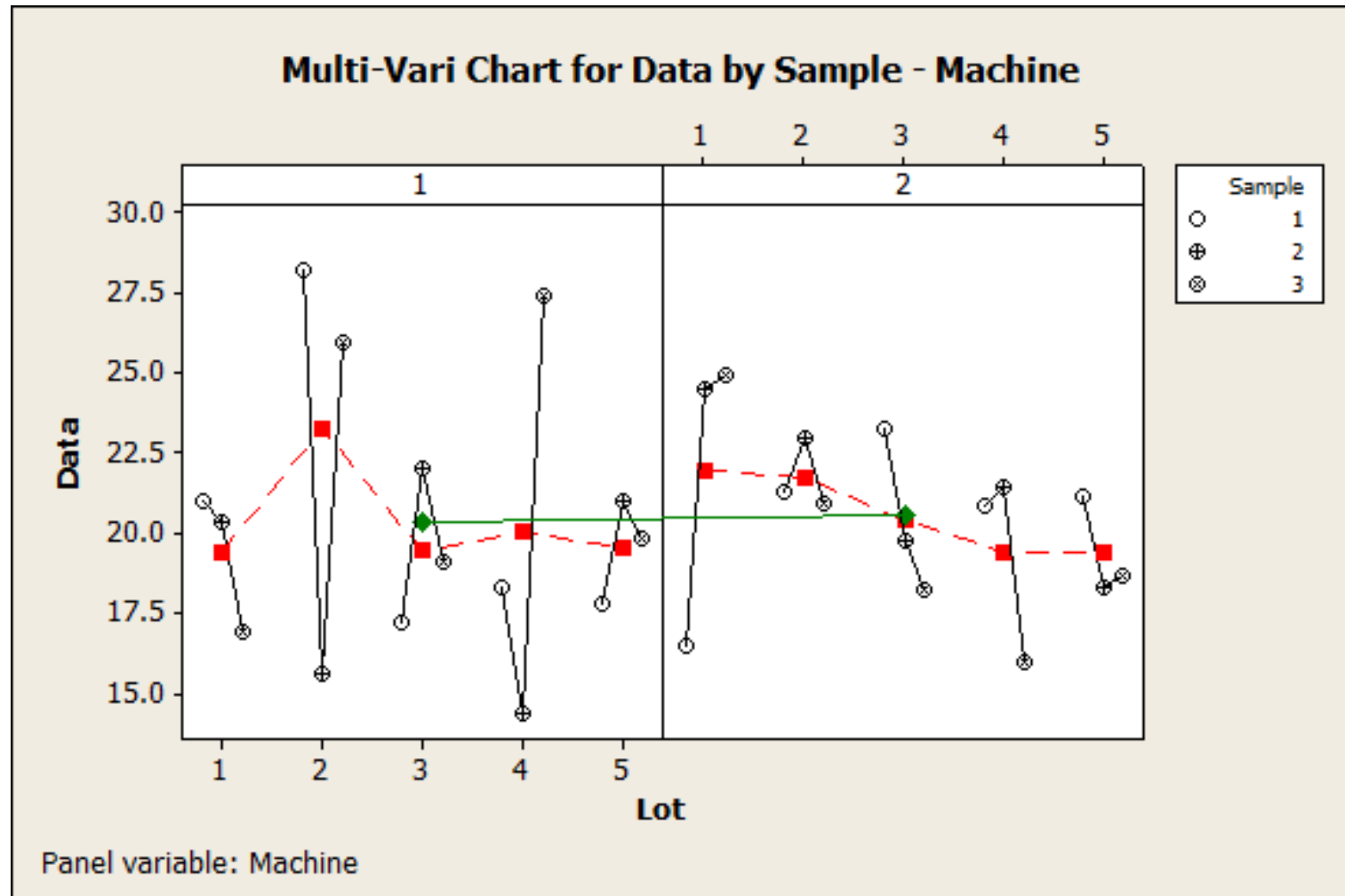
16.00	24.53
14.39	24.95
15.61	23.23
20.99	18.32
20.32	18.63
20.91	18.31
16.46	22.05
16.95	22.98
17.21	21.17
27.43	17.77
28.22	18.21
25.96	19.10
21.02	19.85
21.29	19.74
21.46	20.83



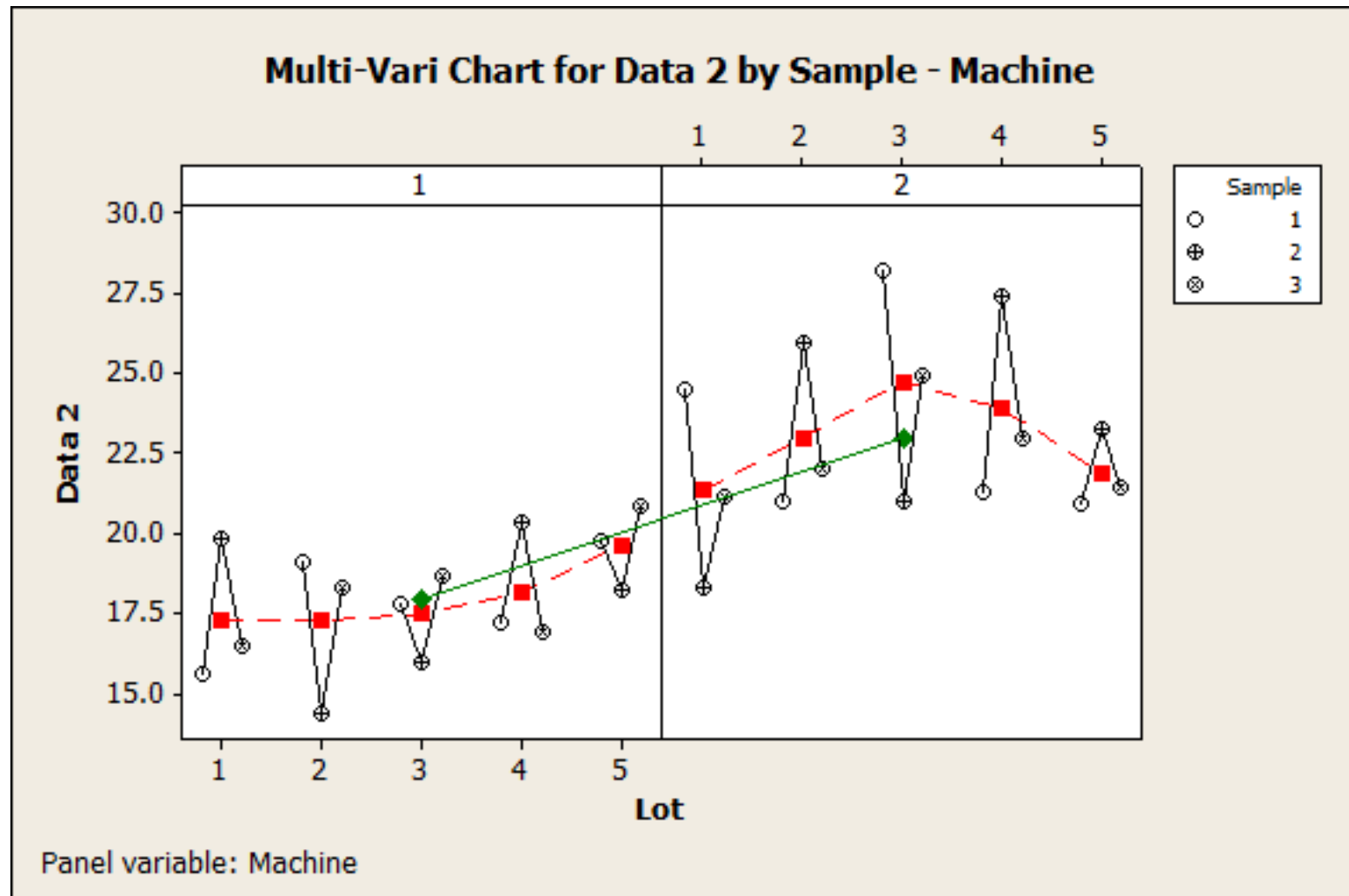
# X Bar & R for Data Set 3



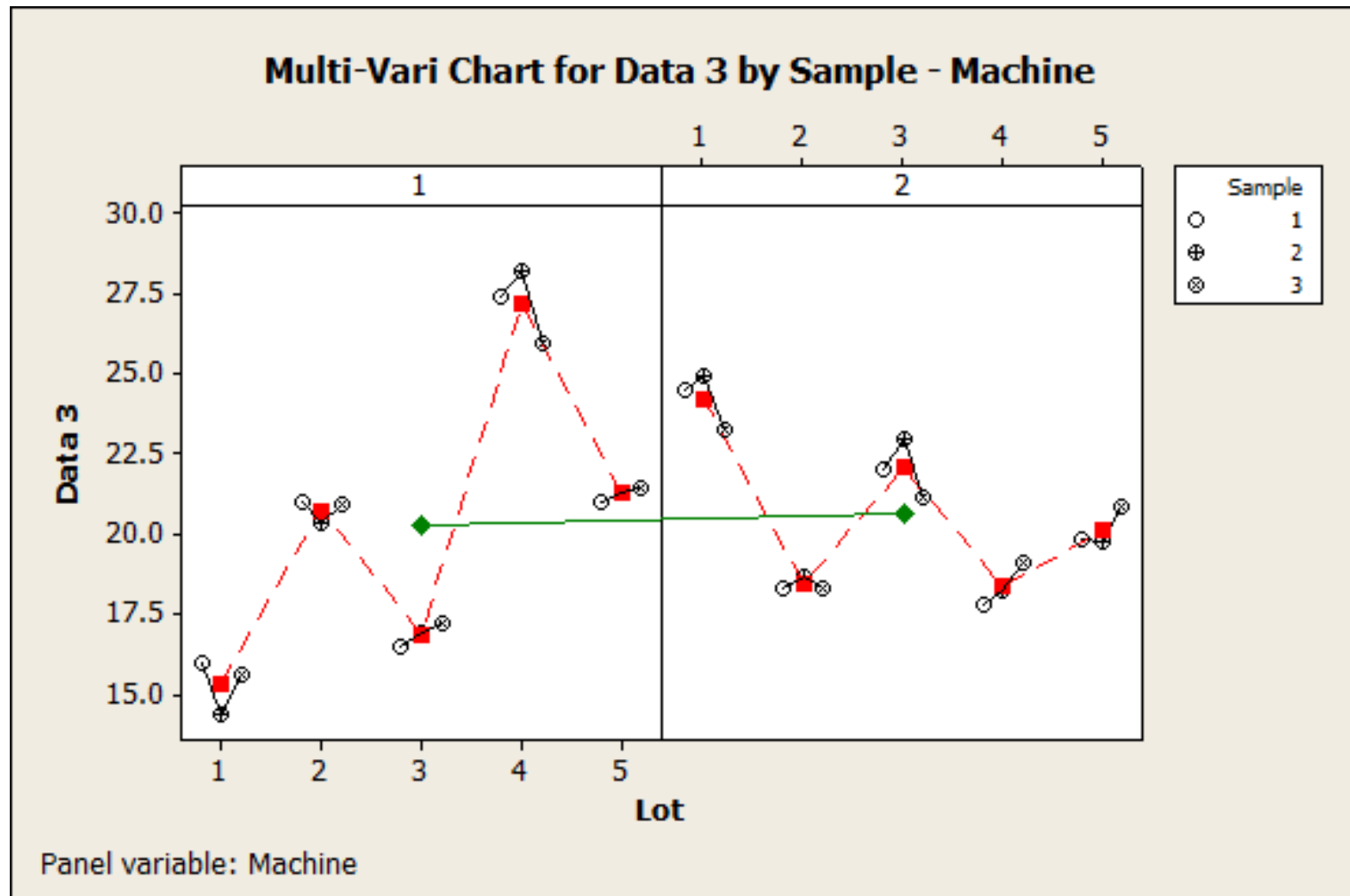
# Multi-Vari Chart for Data Set 1



## Multi-Vari Chart for Data Set 2



# Multi-Vari Chart for Data Set 3



# Questions

**Two questions:**

- 1) What have we learned from the data sets and graphs?**
- 2) Will the product resulting from these three different data sets be different in terms of quality or the same?**

# Hypothesis Testing

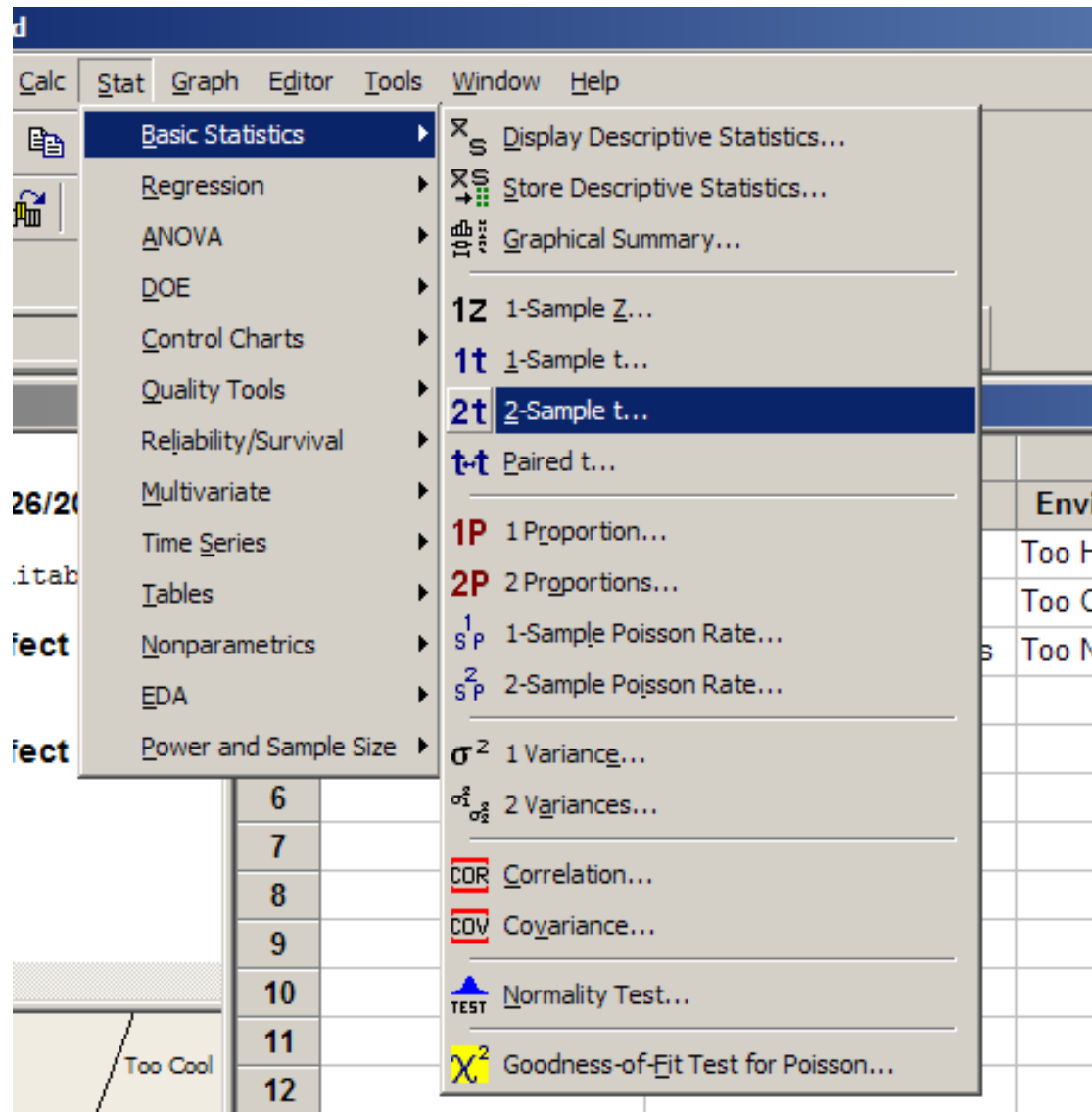
## Types of Hypothesis Tests

- The first three tests look at differences in group **averages**.
- The 4th looks for differences in **variance**.
- The last 3 look at differences in **proportions**.

Hypothesis Test	Purpose
<i>t</i> -test	Compare two group averages
Paired <i>t</i> -test	Compare two group averages when data is paired
ANOVA (Analysis of Variance)	Compare two or more group averages
Test of Homogeneity of Variance	Compare two or more group variances
Chi-Square test	Compare two or more group proportions
1-Proportion test	Compare one proportion to a Prescribed boundary
2-Proportion test	Compare two group proportions



# Hypothesis Testing

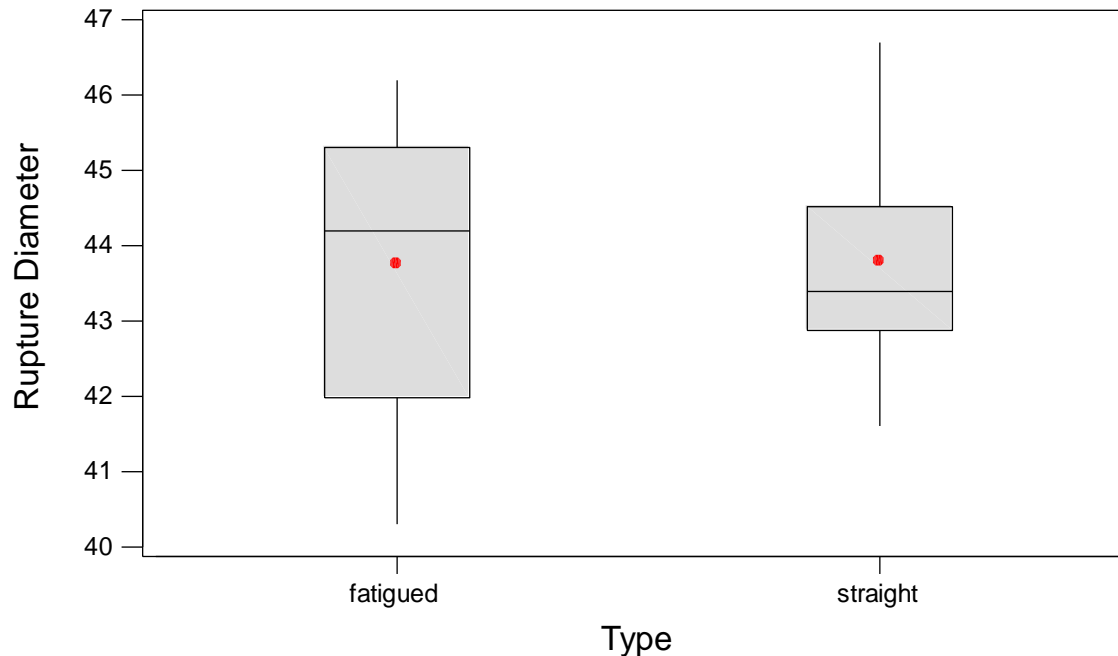


# Another Simple Combination

Does fatigue affect mean rupture diameter?

Boxplots of Rupture by Type

(means are indicated by solid circles)



t-test:  $p=0.96$       F-Test  $p=0.605$       Levene's Test  $p=0.382 \Rightarrow$

**NO.** No statistically significant difference in mean rupture diameter pre and post fatigue.....

## The Three t's

- t-tests come in 3 main types
  - One sample t-test
    - Compares a sample to a known value
    - Value can be based on history
  - Two sample t-test
    - Compares two samples
  - Paired t-test
    - Usually compares two treatments to the sample samples
    - Example: two people measuring the same parts

## Let's Talk About P-values

- There are many ways to state the conclusions reached based on a P-value. Which of these are easier to understand?
  1. A P-value is used to judge whether an observed difference between groups is significantly bigger than common-cause (random) variation (yes, if  $P < .05$ ).
  2. If  $P < .05$ , reject the  $H_0$  and conclude the  $H_a$ .
  3.  $P < .05$  means that there is less than a 5% chance that the groups came from the same distribution.
  4. A P-value determines whether the observed difference is a statistically significant difference (yes, if  $P < .05$ ).
  5. The P-value equals the probability of obtaining the observed difference given that the “true” difference is zero.
  6. If the P-value is low ( $< .05$ ), the observed difference must be significant since the probability is low that such a difference in samples could be observed, if indeed there was no “true” difference.
  7. A P-value is used to judge whether there is enough statistical evidence to reject the null hypothesis (yes, if  $P < .05$ ).

# One-sample t-test

The one-sample t-test assumes the population is normally distributed. However, it is fairly robust to violations of this assumption, provided the observations are collected randomly and the data are continuous, unimodal, and reasonably symmetric.

Why use a one-sample t-test?

A one-sample t-test can help answer questions such as:

Is the process on target?

Does a key characteristic of a supplier's material have the desired mean value?

Is a new material or process taking us in the right direction?

**Let's discuss this last one. Are there times I want to reject the Null Hypothesis?**

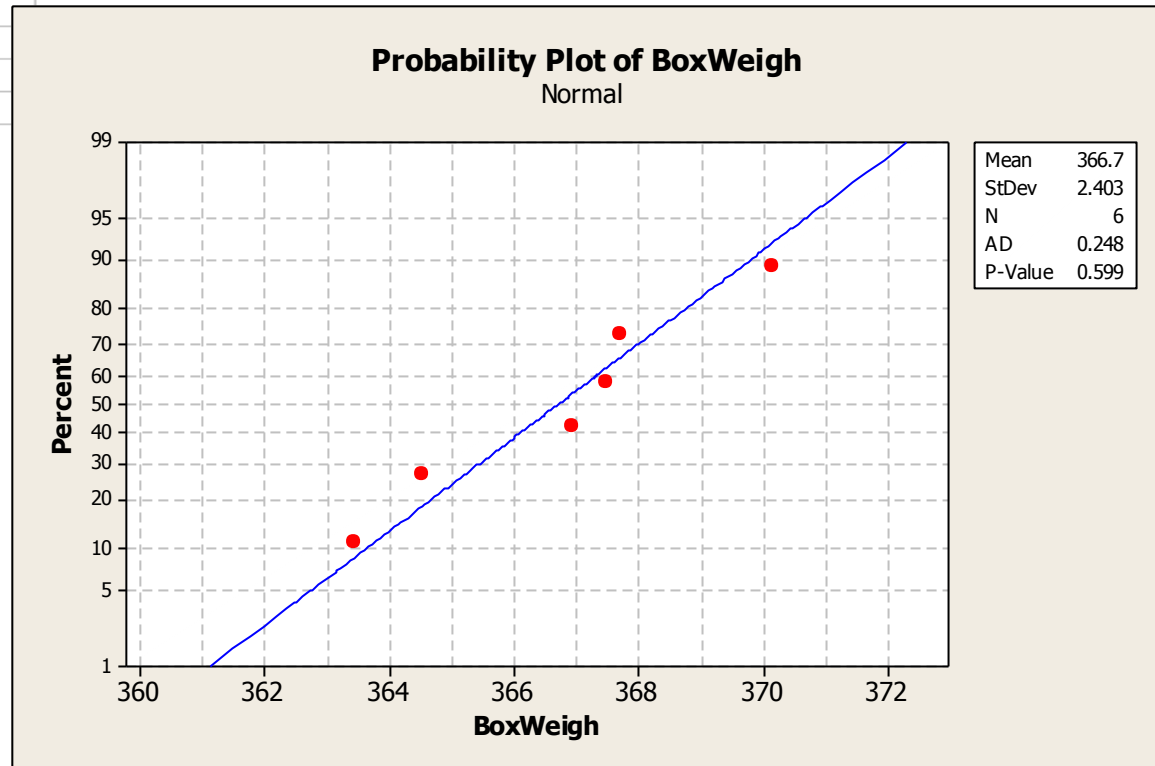
# One-sample t-test example

Cerealbx.MTW ***				
↓	C1	C2	C3	C4
	BoxWeigh			
1	370.129			
2	367.710			
3	366.948			
4	364.528			
5	363.429			
6	367.483			
7				
8				
9				
10				
11				
12				
13				
14				
15				

We want to determine if the box weight is at our target of 365 grams.

We need to start with a check of the normality assumption.

What is the hypothesis here?



# The Analysis

```
Welcome to Minitab, press F1 for help.  
Retrieving project from file: 'D:\MANUFACTURING\BASIC STATISTICS  
1.0\CEREALBX.MPJ'
```

## Results for: Cerealbx.MTW

### Probability Plot of BoxWeigh

### Probability Plot of BoxWeigh |

### One-Sample T: BoxWeigh

Test of  $\mu = 365$  vs not = 365

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
BoxWeigh	6	366.705	2.403	0.981	(364.183, 369.226)	1.74	0.143

## The Paired t-test

**The paired t-test uses matched data**

**Matched data can be for analysis of:**

Two Operators

Two Materials

Two Suppliers

Two Machines

Two Pieces of Measuring Equipment

**It is statistically more stringent than a  
two-sample t-test**

**Let's try to understand why.**



## An Exercise with the t test

SCRAP.MTW ***				
↓	C1	C2	C3	C4
	Cell	New\$Scrap	Std\$Scrap	
1	1	101.97	114.71	
2	2	126.17	134.40	
3	3	133.63	133.68	
4	4	173.98	185.04	
5	5	206.26	209.60	
6	6	161.11	168.23	
7	7	99.63	109.02	
8	8	133.82	131.06	
9	9	154.88	161.46	
10	10	141.72	146.78	
11				

Let's do treat this data as paired and also as unpaired to see if there is a difference.

Here's The Graphs of the unpaired approach

**2-Sample t (Test and Confidence Interval)**

☐ Samples in one column  
Samples:   
Subscripts:

☒ Samples in different columns  
First:   
Second:

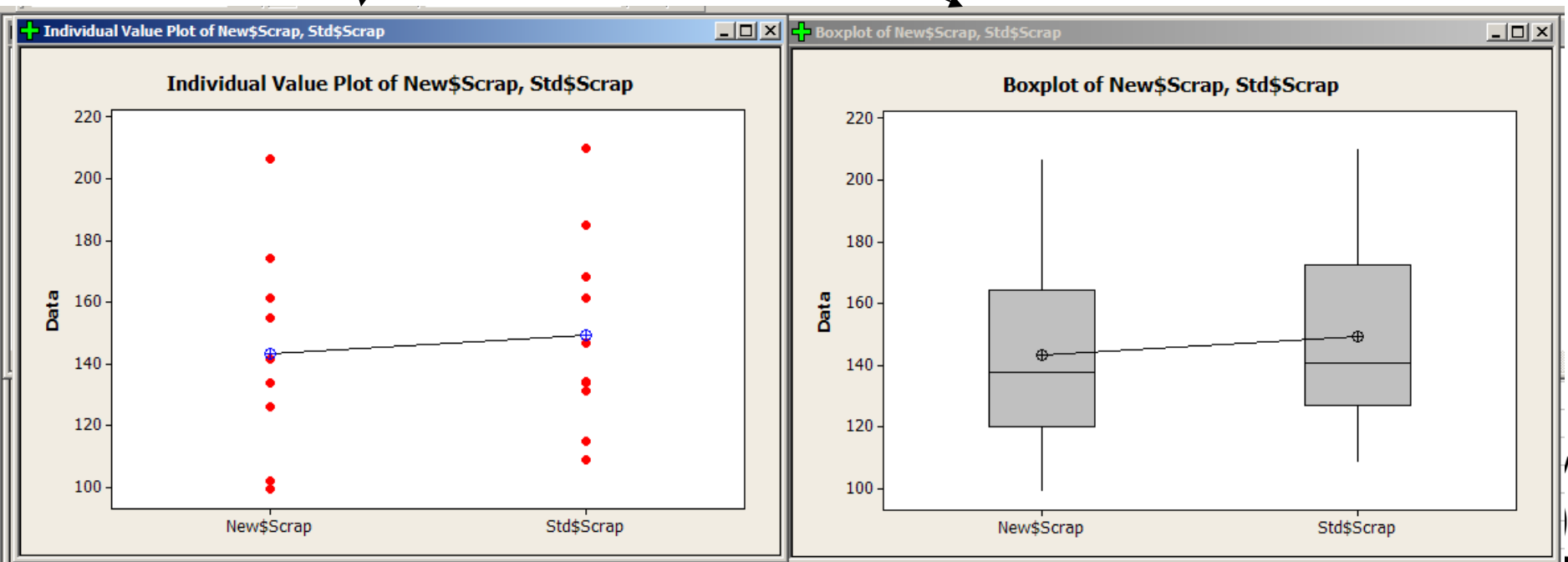
☐ Summarized data  
First:  Sample size:  Mean:  Standard deviation:   
Second:

☐ Assume equal variances

Select

Help

Graphs... Options... OK Cancel



## And the session windows

### 2-sample t

```
Difference = mu (New$Scrap) - mu (Std$Scrap)
Estimate for difference: -6.1
95% CI for difference: (-36.3, 24.2)
T-Test of difference = 0 (vs not =): T-Value = -0.42  P-Value = 0.677  DF = 17
```

### Paired t

Paired T for New\$Scrap - Std\$Scrap

	N	Mean	StDev	SE Mean
New\$Scrap	10	143.3	32.4	10.2
Std\$Scrap	10	149.4	31.7	10.0
Difference	10	-6.08	4.82	1.52

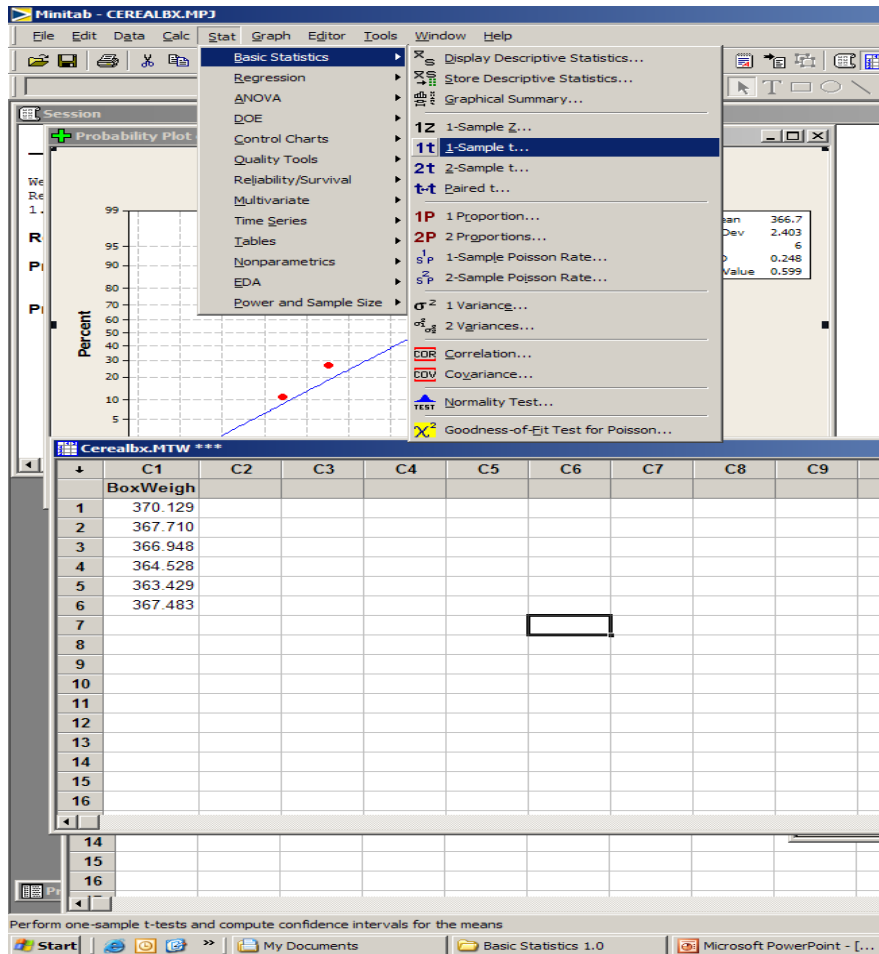
```
95% CI for mean difference: (-9.53, -2.63)
T-Test of mean difference = 0 (vs not = 0): T-Value = -3.99  P-Value = 0.003
```

What 2 Key indicators are there regarding the Null Hypothesis?

## But Why?

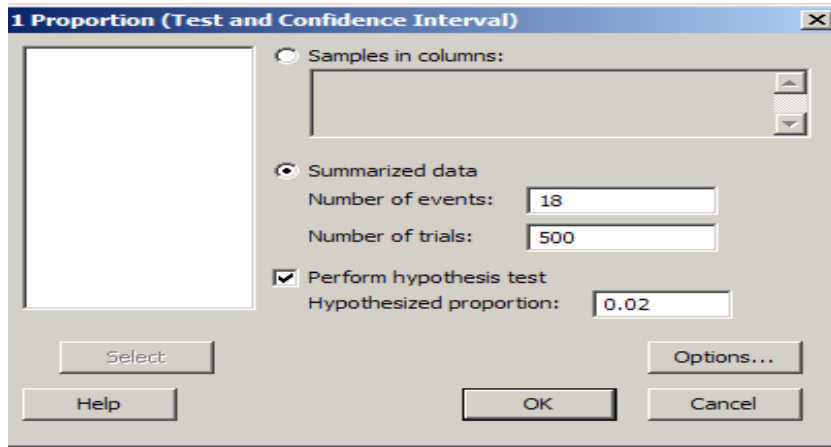
Cell	New \$Scrap	Standard \$Scrap	Paired Differences
1	101.97	114.71	12.74
2	126.17	134.40	8.23
3	133.63	133.68	0.05
4	173.98	185.04	11.06
5	206.26	209.60	3.34
6	161.11	168.23	7.12
7	99.63	109.02	9.39
8	133.82	131.06	-2.76
9	154.88	161.46	6.58
10	141.72	146.78	5.06
average	143.32	149.40	6.08
st.dev.	32.41	31.70	4.82

# Executing the test

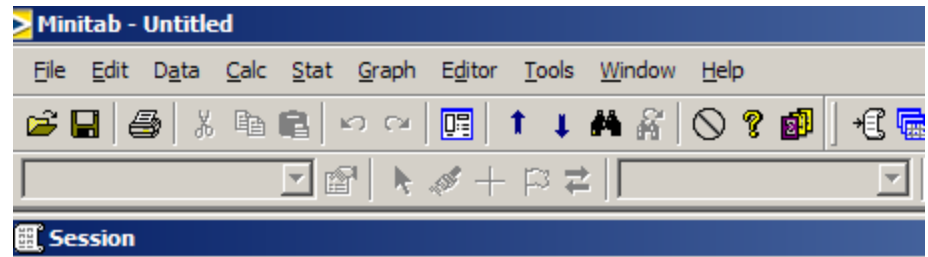


The '1-Sample t (Test and Confidence Interval)' dialog box is shown. The 'Samples in columns:' option is selected, and 'BoxWeigh' is entered in the text box. The 'Summarized data' option is unselected. The 'Perform hypothesis test' checkbox is checked, and the 'Hypothesized mean' is set to 365. The 'Select', 'Help', 'Graphs...', 'Options...', 'OK', and 'Cancel' buttons are visible at the bottom.

# One-portion test



We'll use summarized data for this one



4/29/2009 9:35:25 AM

Welcome to Minitab, press F1 for help.

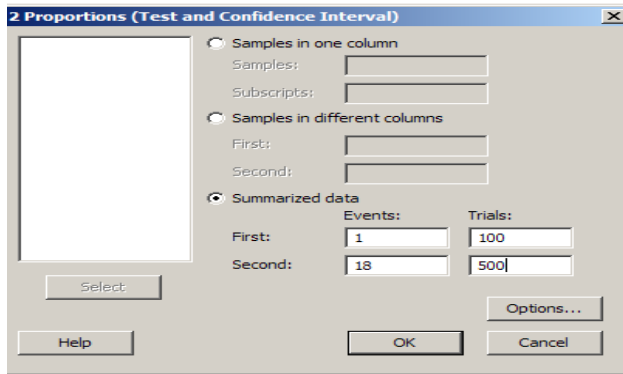
## Test and CI for One Proportion

Test of  $p = 0.02$  vs  $p \text{ not } = 0.02$

Sample	X	N	Sample p	95% CI	Exact P-Value
1	18	500	0.036000	(0.021473, 0.056300)	0.023

What statements can we now make about the process compared to a target of 2%?

# Two-portion test



We can compare two samples this way.

```
1      18  500  0.036000  (0.021473, 0.056300)  0.023
```

## Test and CI for Two Proportions

Sample	X	N	Sample p
1	1	100	0.010000
2	18	500	0.036000

Difference = p (1) - p (2)

Estimate for difference: -0.026

95% CI for difference: (-0.0514349, -0.000565134)

Test for difference = 0 (vs not = 0): Z = -2.00 P-Value = 0.045

Fisher's exact test: P-Value = 0.225

\* NOTE \* The normal approximation may be inaccurate for small samples.

What statements can we make about these two samples?

# **Process Capability**



# Capability Studies

- 1) The true purpose of a capability study is to determine exactly how well a process, product, component can meet a predetermined specification.
- 2) It can be measured in several different ways.
- 3) The values generated can be absolute or relative. We may want to compare or we may want to use it to check against a hard and fast standard.

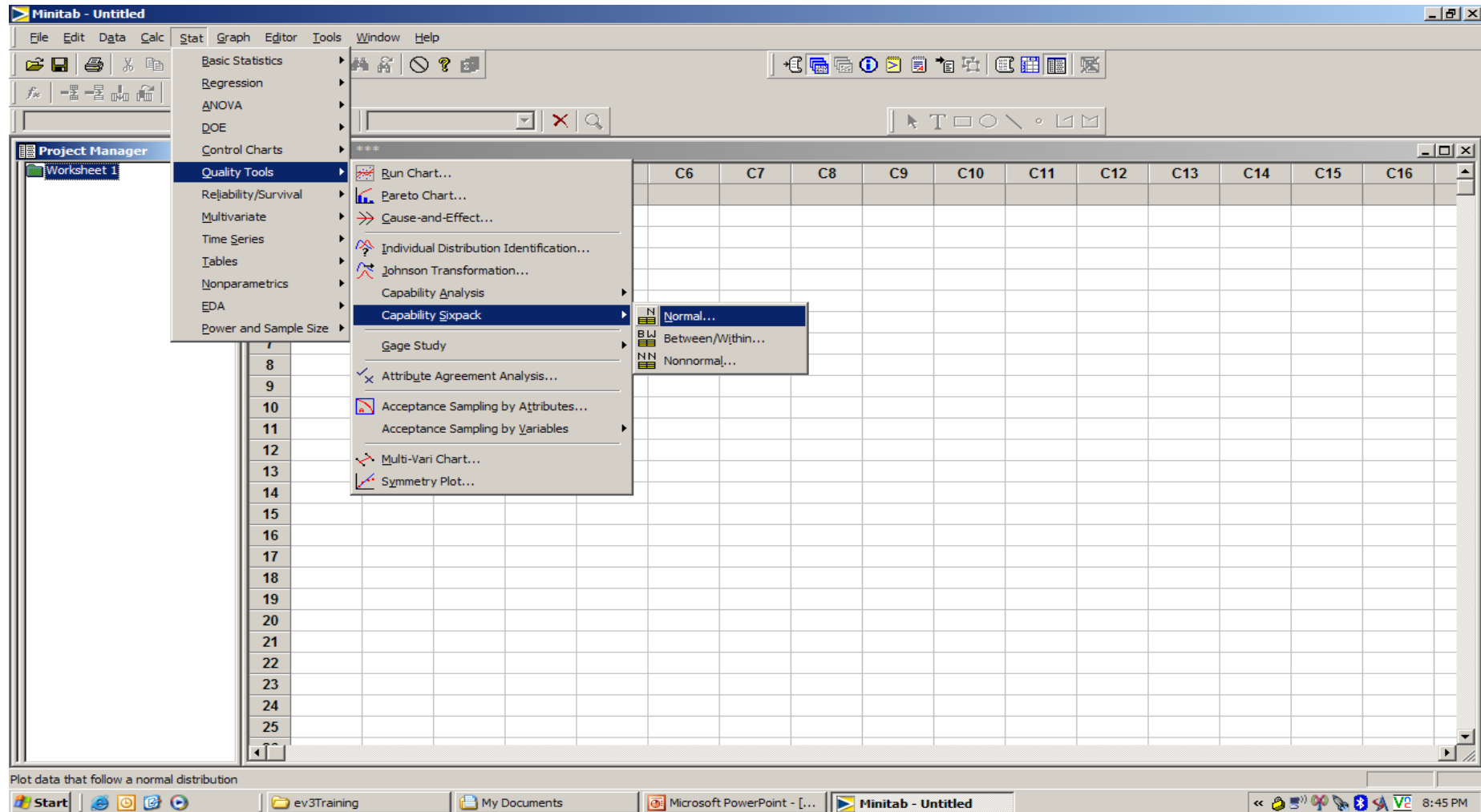
A Capability Study consists of:

Measuring Devices, Procedures, Definitions, People, Specifications and Processes.

## 7 Steps to a Good Process Capability Study

1. Obtain Process Data
2. Prepare Control Charts and Frequency Displays
3. Evaluate Process Stability
4. Estimate Short-Term and Long-Term Process Variation
5. Evaluate how well a Normal Distribution approximates the individual process observations
6. Prepare Summary Measures of Process Capability
7. Judge adequacy of Process Capability

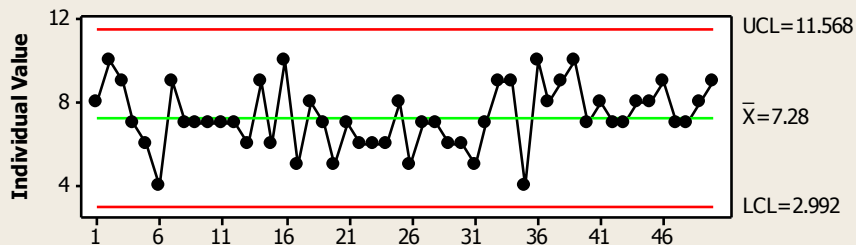
# Where is it in Minitab ?



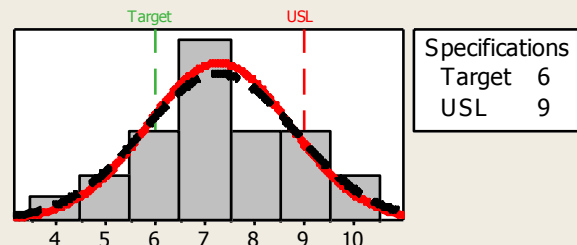
# Process Capability Example

## Process Capability Japan Production Line (Prior JIT)

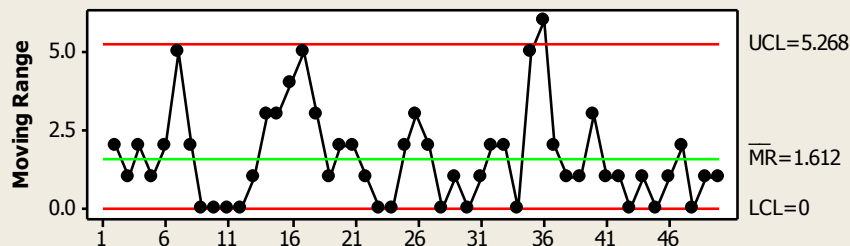
I Chart



Capability Histogram

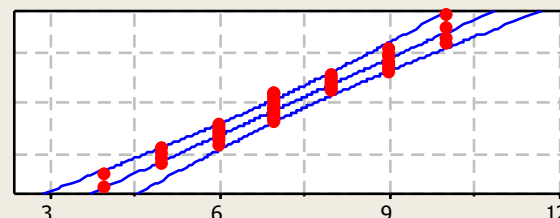


Moving Range Chart

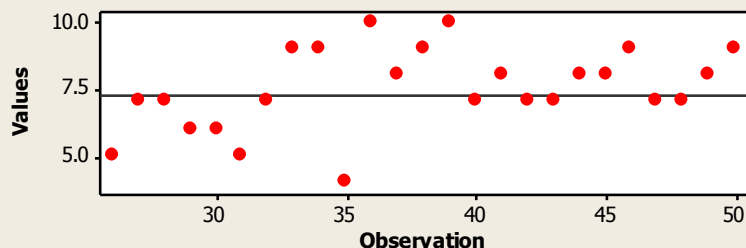


Normal Prob Plot

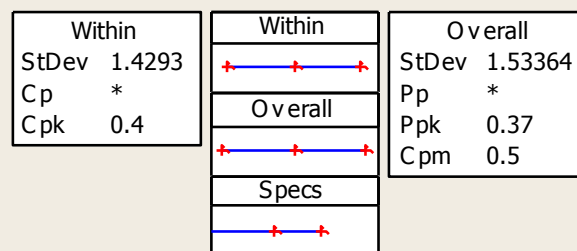
AD: 1.070, P: 0.008



Last 25 Observations

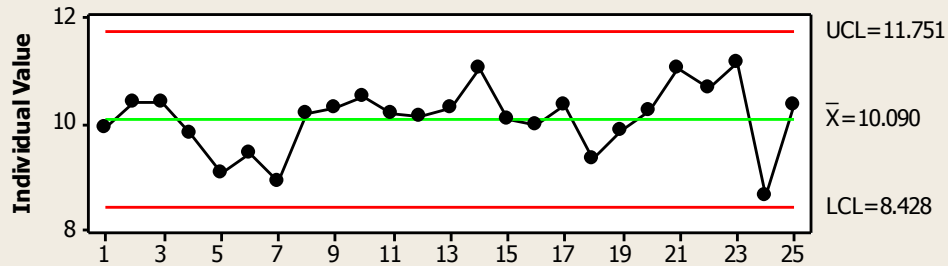


Capability Plot

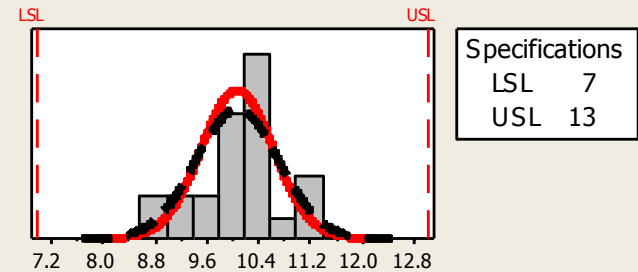


# Process Capability Sixpack of Data

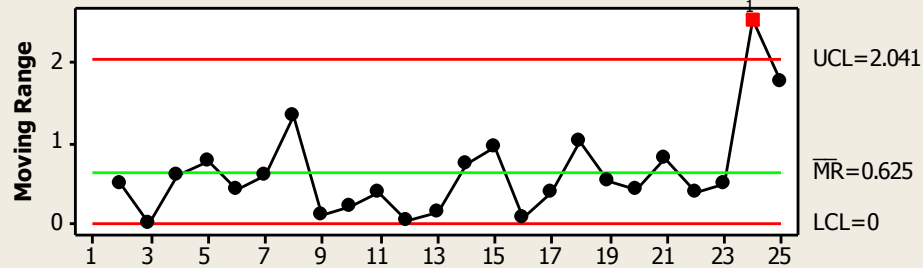
I Chart



Capability Histogram

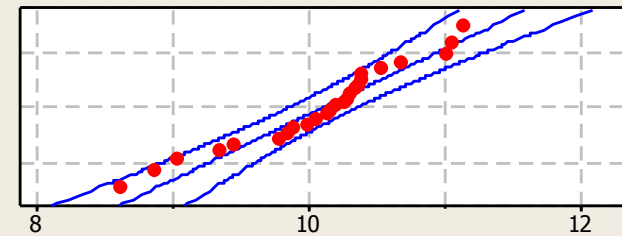


Moving Range Chart

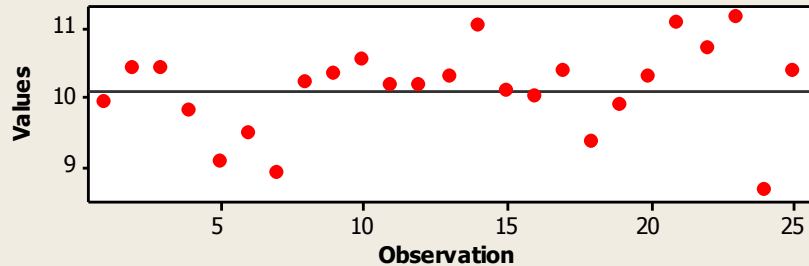


Normal Prob Plot

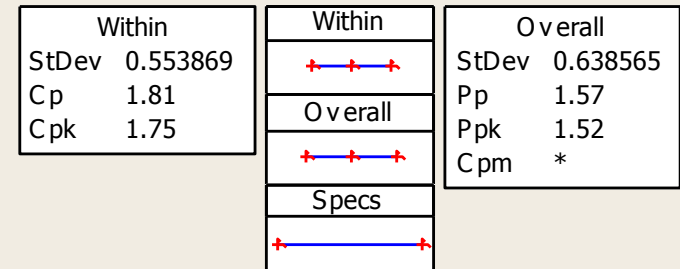
AD: 0.585, P: 0.116



Last 25 Observations

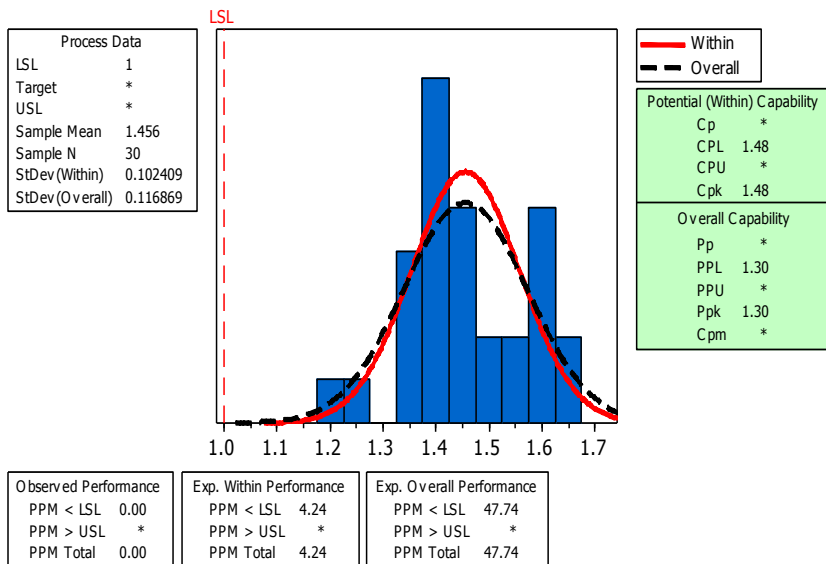


Capability Plot

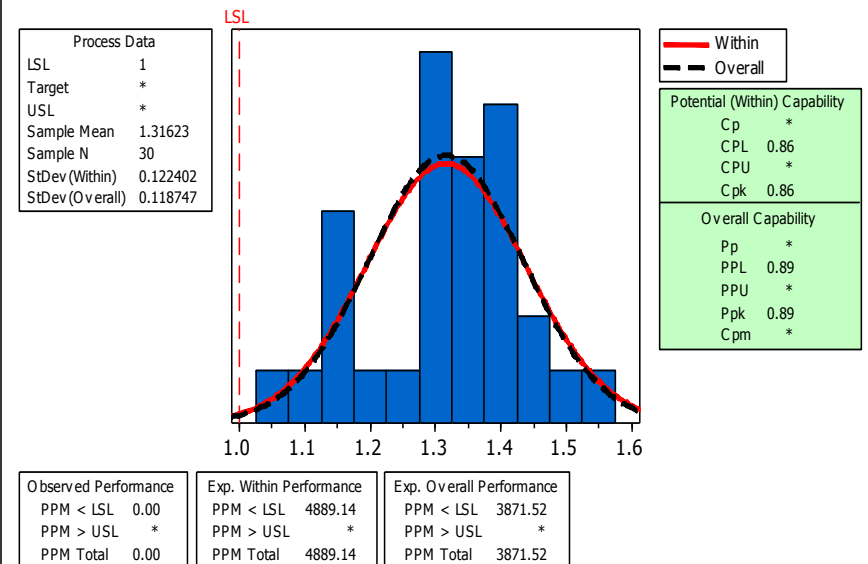


# How About Using Process Capability for Comparative Purposes?

Tray Packaging (CA Sealer)



Tray Packaging (Juarez Sealer)





Thank You For Your Time and Attention !  
Uncle Larry

Uncle Larry's  
a Cool Dude!

