How to Use Minitab:

Introduction and Basics

Noelle M. Richard
08/27/14

Contents

* Click on the links to jump to that page in the presentation. *

What is Minitab?

* For more Minitab tutorials, click the links below. *

- A program used for statistical analysis
  - As of June 2014, version 16
- Can be used for:
  - Basic Statistics
    - Graphs/Summary Statistics
    - Confidence Intervals
    - Hypothesis Tests
    - ANOVA Tables
  - Regression Analysis and Modeling
    - Simple Regression
    - General Linear Models

What is Minitab?

- Quality Control
  - Control Charts
  - Capability Analysis
  - Gauge R & R Studies
- Design of Experiments
  - Randomized Complete Block Designs
  - Factorial Designs
  - Fractional Factorial Designs
  - Split-Plot Designs
  - Response Surface

- Unfamiliar with these terms? No worries! They will be explained later on.

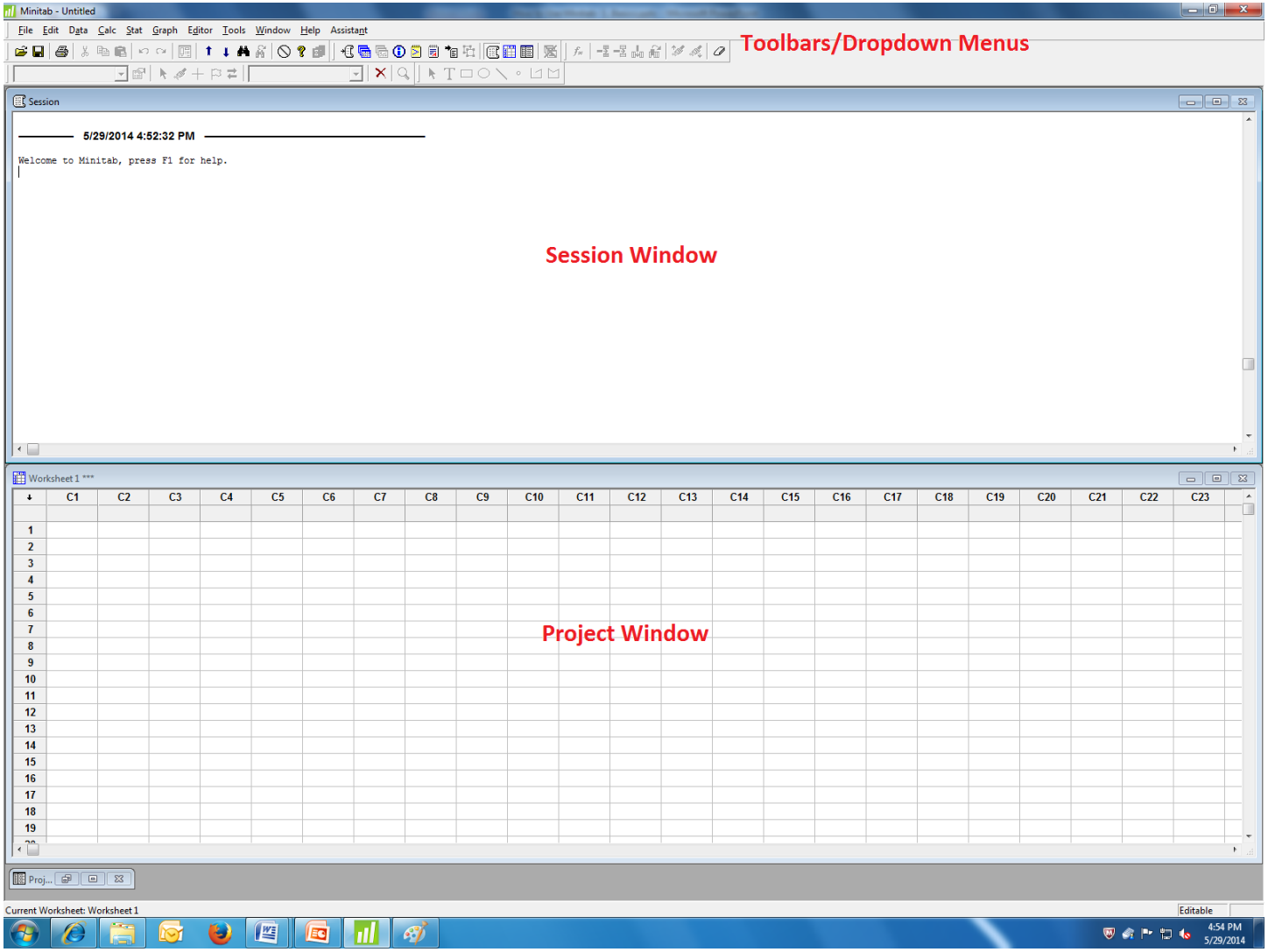
Return to Contents

Minitab Environment

- Project Window
  - Where you enter in your data
- Session Window
  - Where your results/output will be displayed
- Toolbar/Dropdown Menus

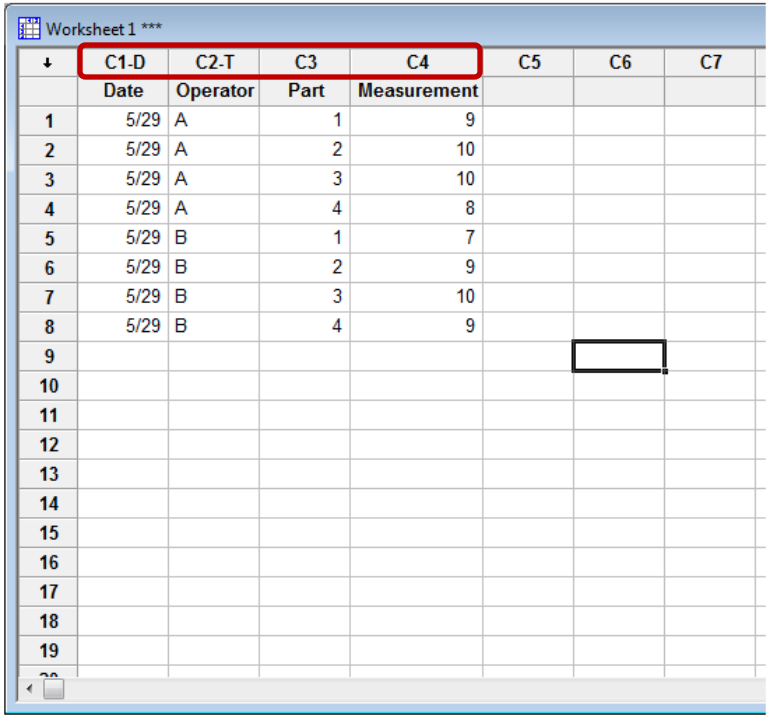
Return to Contents

Minitab Environment



Return to Contents

Project Window


Worksheet 1 ***

	C1-D	C2-T	C3	C4	C5	C6	C7
	Date	Operator	Part	Measurement			
1	5/29	A	1	9			
2	5/29	A	2	10			
3	5/29	A	3	10			
4	5/29	A	4	8			
5	5/29	B	1	7			
6	5/29	B	2	9			
7	5/29	B	3	10			
8	5/29	B	4	9			
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							

- Minitab analyses data in columns
- If data arranged in rows, transpose in excel first
  - Transposing in Minitab may mess up your data
- First Row (C1, C2, C3, etc.) → Column Labels
  D = Date T = Text Blank = Numeric
- Second Row → Variable Names

Return to Contents

Getting Data into Minitab

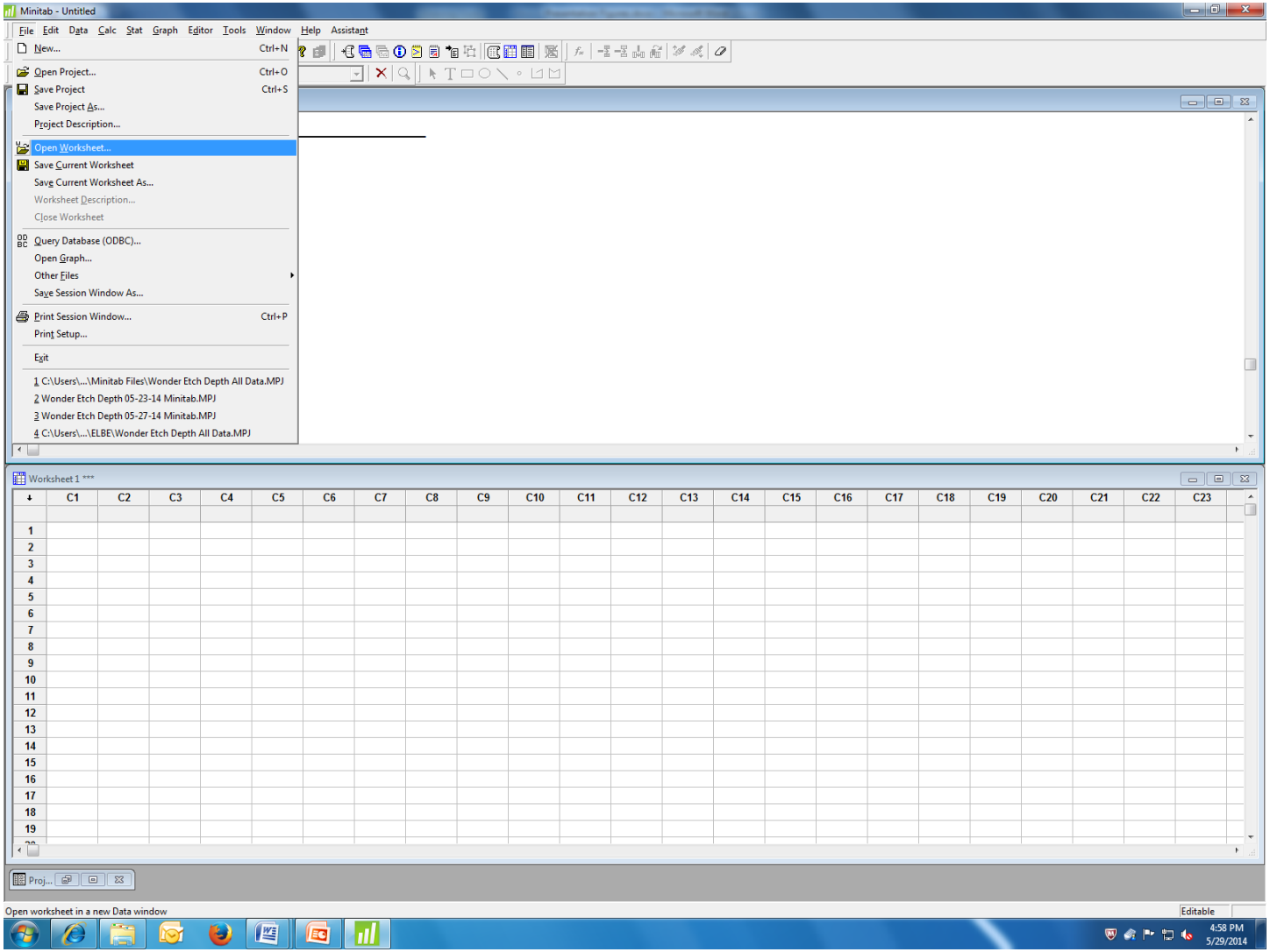
1. Type in yourself

2. Copy/Paste from Excel
   - Highlight/copy cells in Excel
   - Highlight exact # of cells in Minitab Project Window (MUST be exact dimensions)
   - Paste

3. Import
   - File→Open Worksheet
   - Select the Excel file and Open
   - This will import <u>all</u> worksheets from the file. Close out any you don't want.

Return to Contents

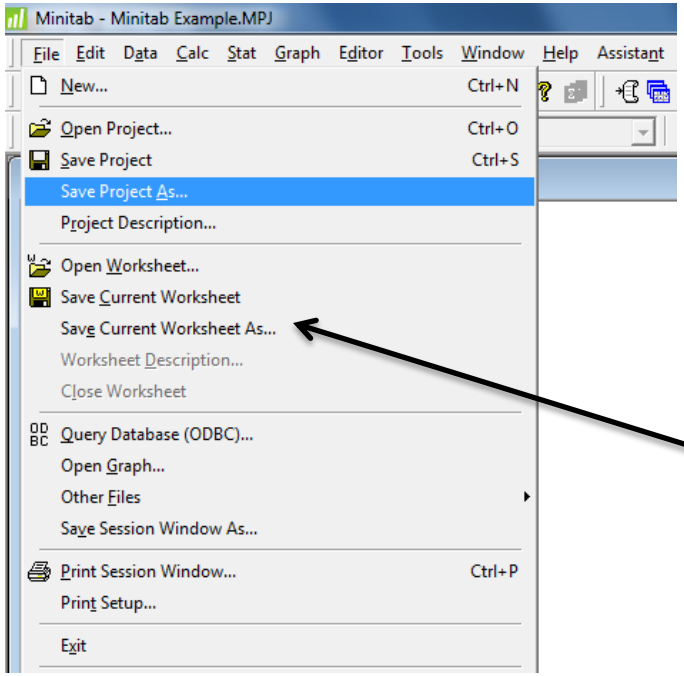
Importing Data From Excel



Return to Contents

Saving Project

- Saving your Project will:
  - Save all data entered into Project Window(s)
  - All output in the Session Window



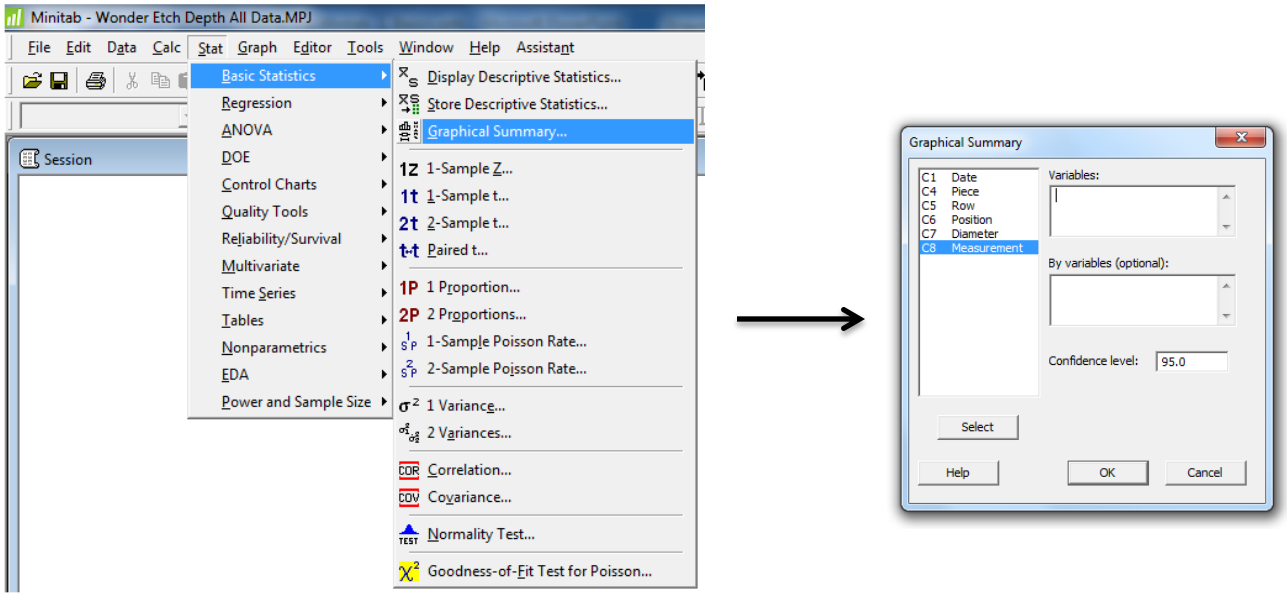
Can also save individual worksheets to open in other projects

Return to Contents

Graphical Summary

Data in the following examples: There are rectangle pieces with die on each. The die have different diameter sizes. We are measuring the width of a porous material surrounding each die. Pieces were etched using different methods.

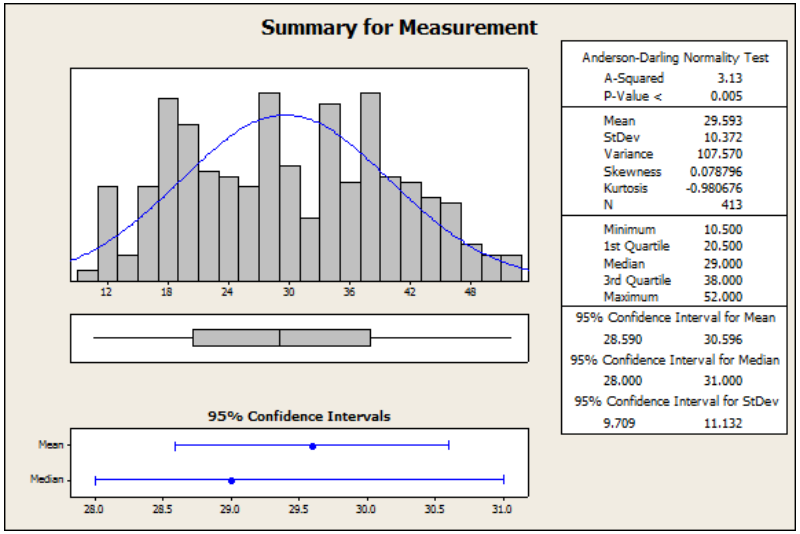
- Great "first look" at your data
- Add the variable(s) you want to look at in "Variables" box
- Can also split data up using "By variables"
  - Will produce separate graphs of "Variable" for each "By Variable"



Return to Contents

Graphical Summary/Normality

Output: Summary Statistics, Histogram, Test for Normality, etc.



Normality Test:

P-Value ≤ 0.05 Data not normally distributed

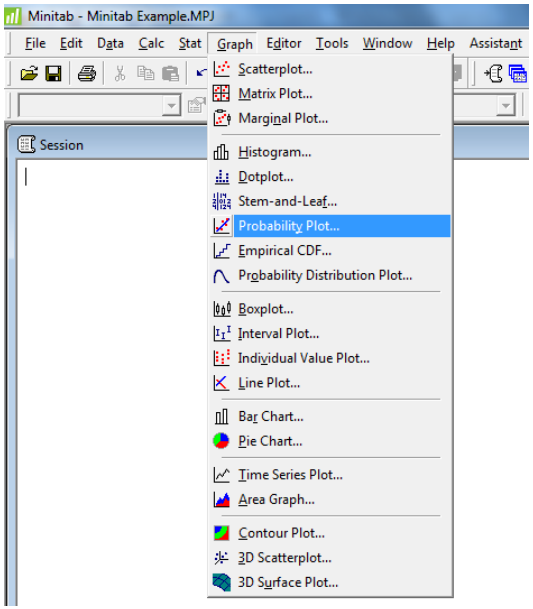
Why is this important? Many statistical techniques assume normality of your data

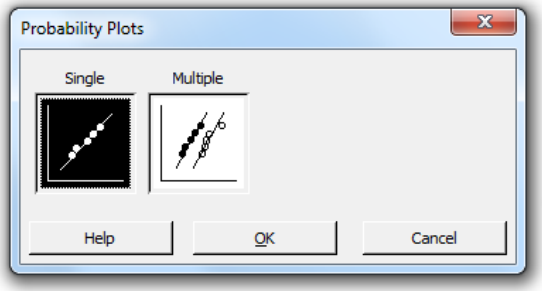
Anderson-Darling Test only good for larger samples (n > 20)

If n < 20, see next slide.

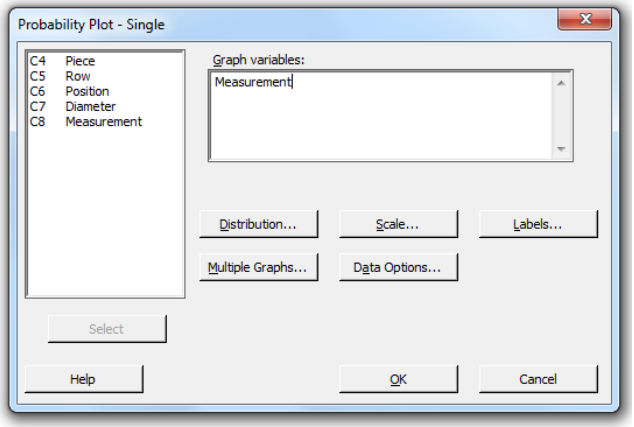
Return to Contents

Normality Test Using Probability Plot

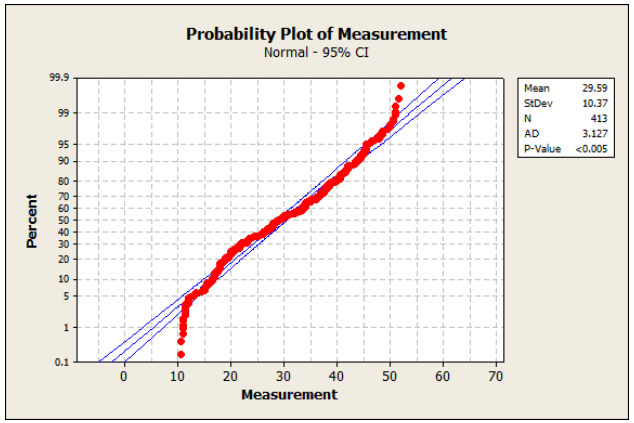




Use if sample size < 20



Output

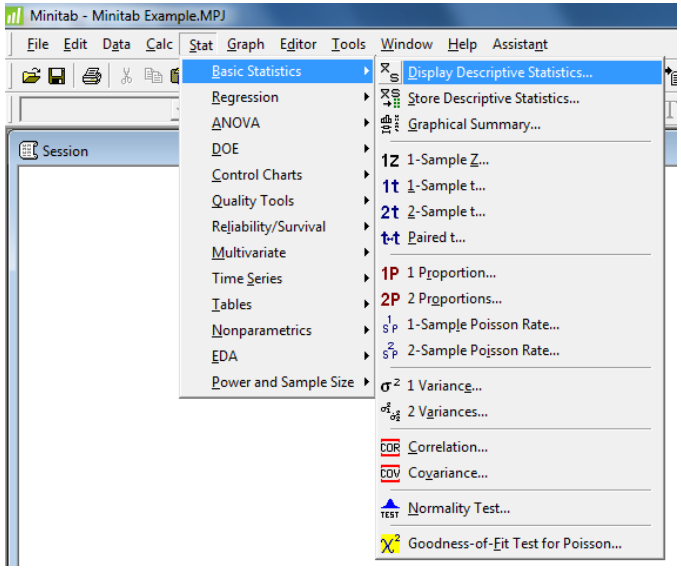


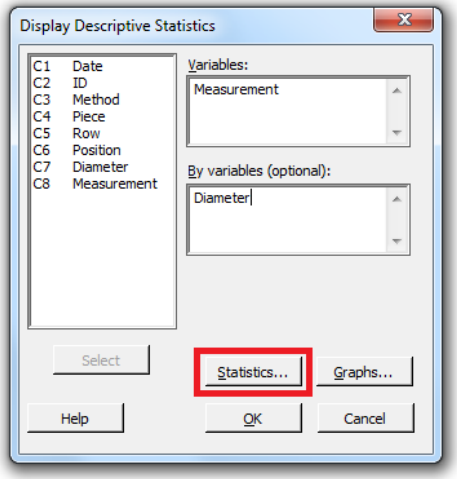
If the data are normally distributed, most red data points should lie on the center blue line. Almost all points should be within the outer blue lines.

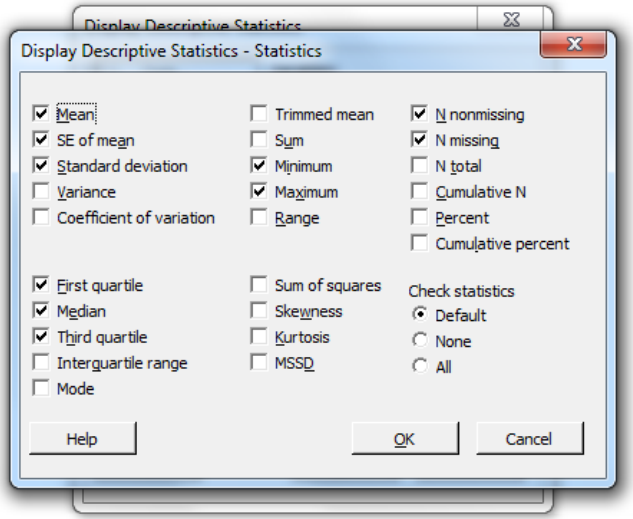
The plot to the left shows data that is NOT normally distributed.

Return to Contents

Descriptive Statistics







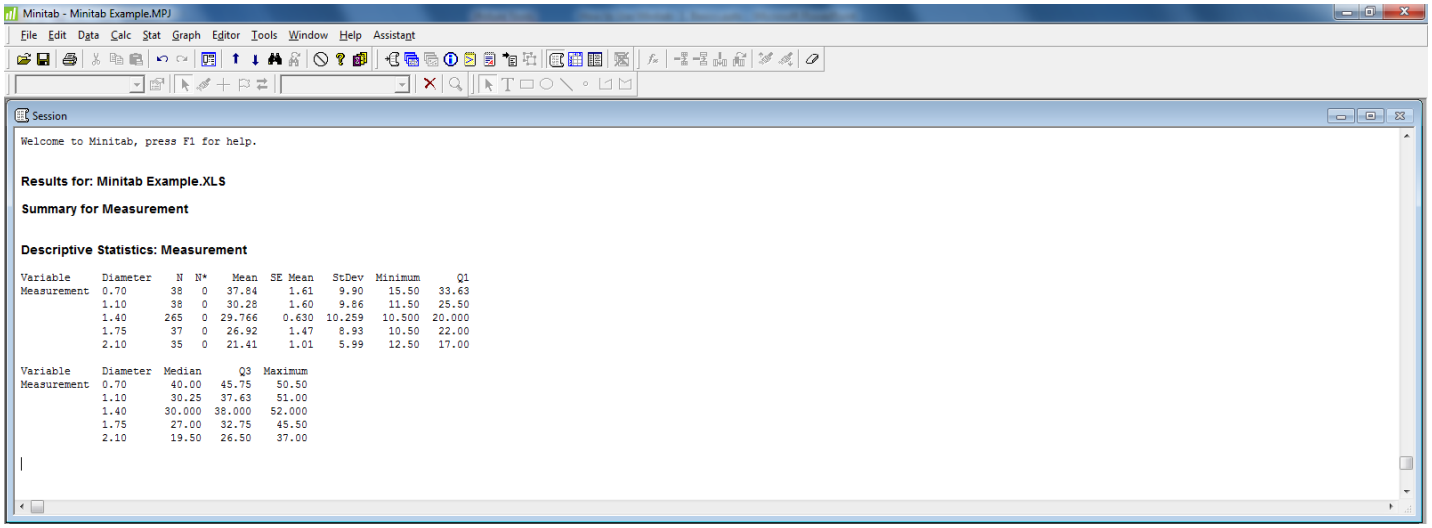
Shows various summary statistics like mean, variance, range, etc.

Can select what statistics you want to see

Return to Contents

Descriptive Statistics

- Output stored in Session Window
- Can Highlight and Copy all text



Return to Contents

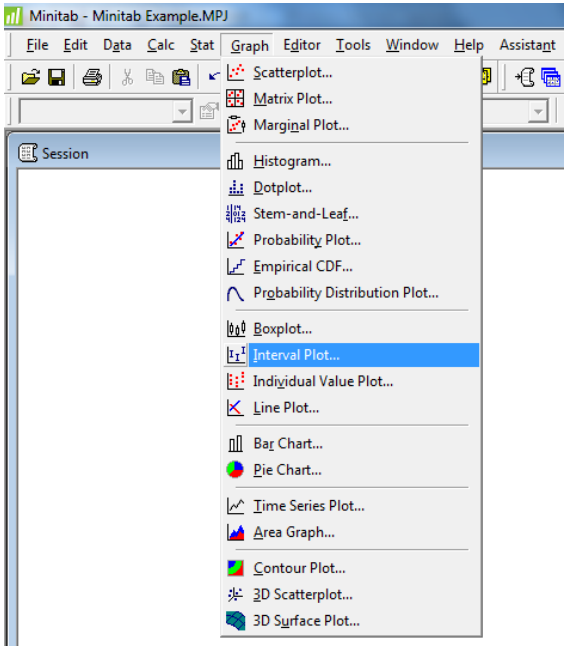
Confidence Intervals/Interval Plots

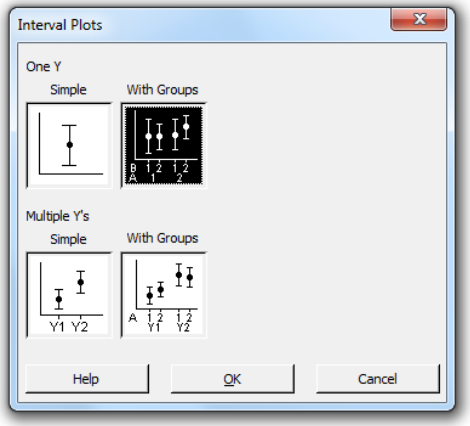
- 95% confidence intervals (C.I.) for mean
  - Your data is a sample from a large population
  - Use C.I. to estimate the "true" mean for the population
  - 95% = confidence level
- Graphical Summary produces C.I. for mean, median and standard deviation
  - Can change confidence level to what you want
  - Higher level means wider interval. Usually 95% is sufficient
  - Smaller sample sizes and larger std. deviation produce wider intervals
- Can draw interval plot to visualize the intervals

** http://en.wikipedia.org/wiki/Confidence_interval#Meaning_and_interpretation

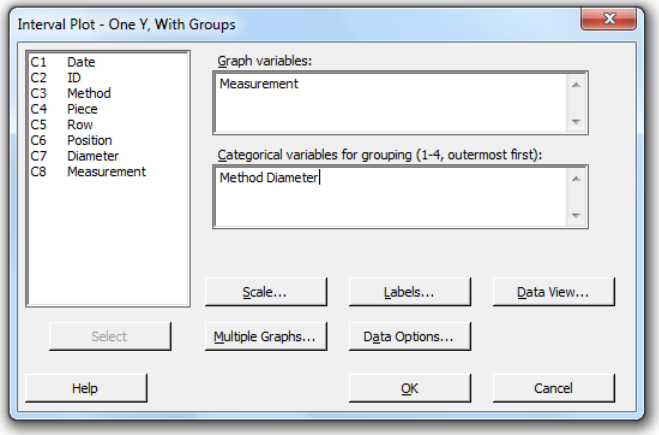
Return to Contents

Confidence Intervals/Interval Plots

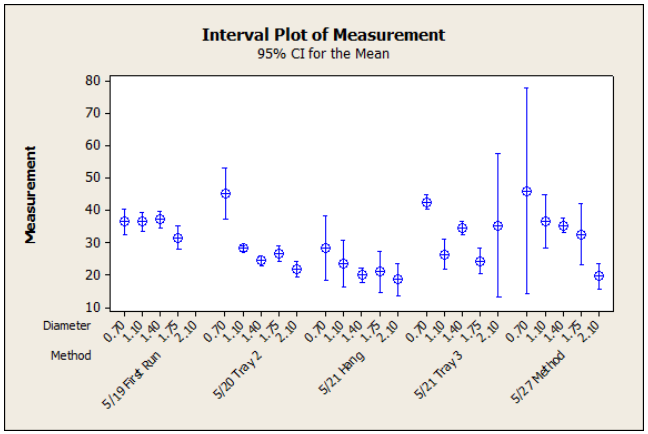




Can find intervals (for the mean) of specific group(s) in your data



Output:



Return to Contents

ANOVA Tables

- Analysis of Variance
  - How much of your variation is within groups?
  - How much of your variation is between groups?
  - Are the means of 2 or more samples statistically equivalent?

- ANOVA tables can answer these questions. They are used in various other areas of statistics too
  - Seeing how well a regression model fits data
  - Finding Repeatability and Reproducibility in Gage R & R Studies

- One-Way ANOVA tables are used to compare the means of more than 2 samples. In order to proceed with One-Way ANOVA, we need to understand Hypothesis Tests.

Return to Contents

Hypothesis Tests

- Answer questions like: Are population means equivalent? Is one greater than the other? Do populations have the same std. deviation?

- Each test contains two hypotheses:
  1. Null Hypothesis: This is the "default" hypothesis. Usually, it assumes there is no relationship between 2 phenomena
     ex.) 2 population means are equivalent.
  2. Alternative Hypothesis: Another hypothesis that could be true
     ex.) mean 1 is larger than mean 2

The goal is to either:

1. Not Reject the null hypothesis or
2. Reject the null hypothesis, in favor of the alternative

Return to Contents

Hypothesis Tests

- After the hypotheses have been stated, the test quantifies the evidence against the null hypothesis. This evidence is reflected in a value known as the test statistic.

- From this test statistic, we get a p-value, which is used determine whether to reject the null hypothesis or not.

Large p-value → large amount of evidence in favor of the null hypothesis
Small p-value → small amount of evidence in favor of the null hypothesis

Rule of Thumb: p-value ≥ 0.05 → do not reject the null hypothesis
p-value < 0.5 → reject the null hypothesis in favor of the alternative

Click here for more information about p-values, if interested.

Return to Contents

One-Way ANOVA

- As mentioned before, it is used to compare the means of more than 2 samples.

How does it work? It's a Hypothesis Test!

Null Hypothesis: the samples were drawn from populations with the same mean values.
Alternative Hypothesis: the samples were drawn from pops with different mean values.

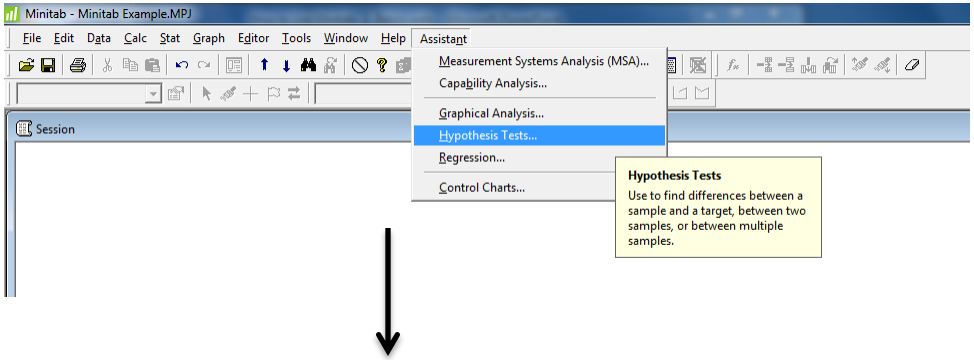
- Note: In this case, if you reject the null hypothesis, you ONLY know that at least one of the samples is different than the others. But you do not know which one it is.

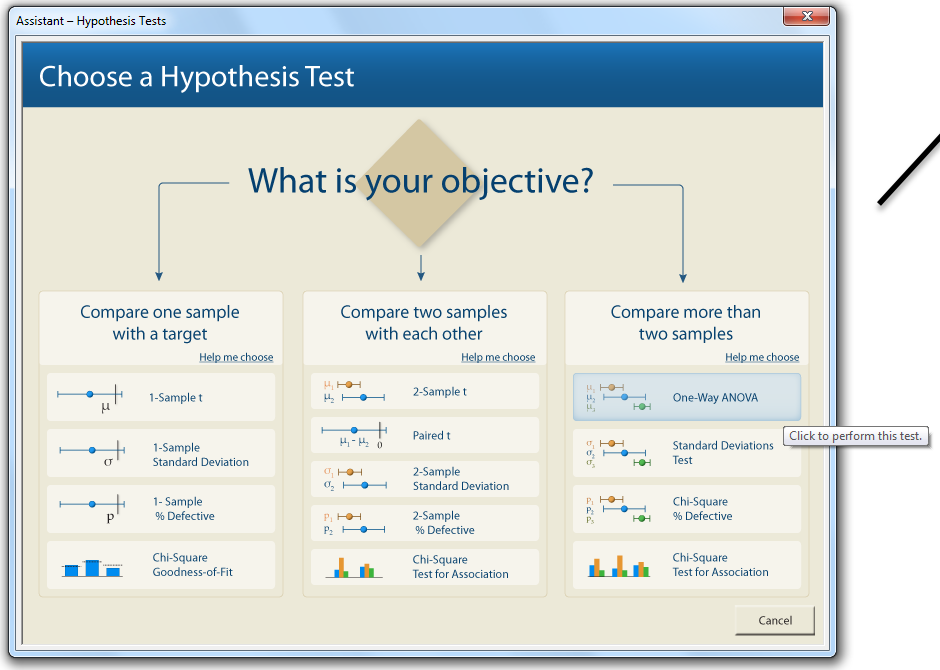
  Can use confidence intervals to estimate the difference between population means. Intervals that do not contain 0 indicate the means are significantly different.

Return to Contents

One-Way ANOVA / Hypothesis Tests

- Can use Minitab Assistant (v.16 only)





Select what type of test you want to do.

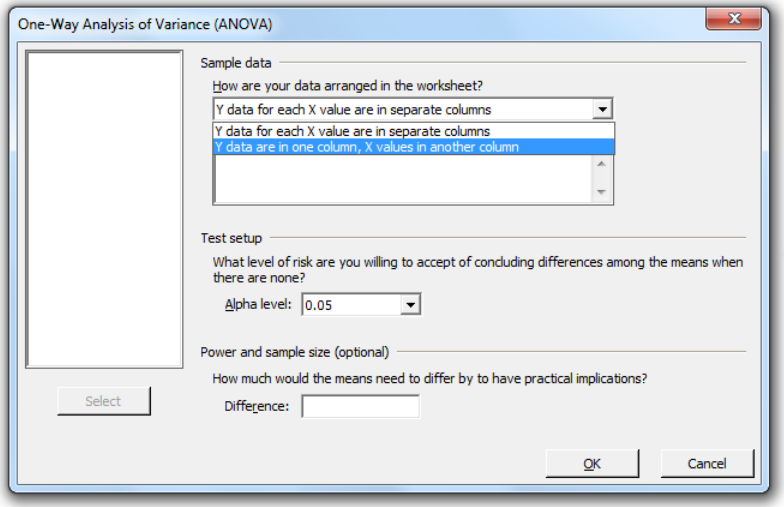
Here, we will proceed with "One-Way ANOVA" in the 3rd column.

Remember, this is for comparing means of more than 2 samples.

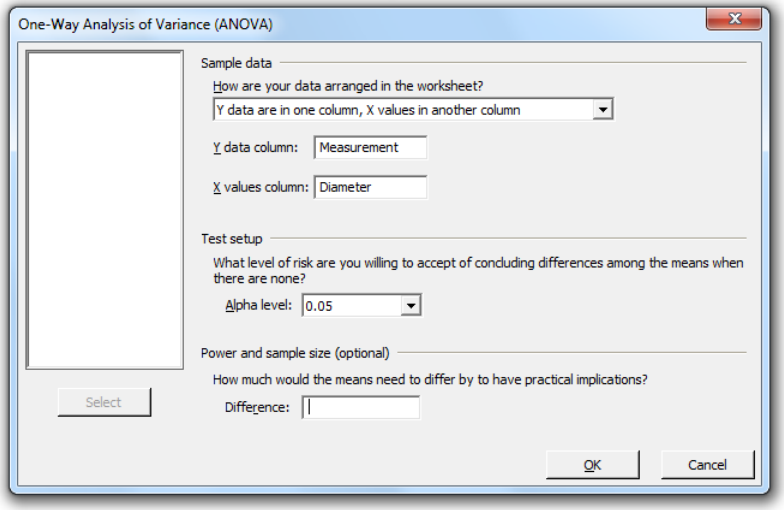
If you are comparing only two samples, use the 2-Sample t test. Data from your samples should be in separate columns.

Return to Contents

One-Way ANOVA / Hypothesis Tests



Choose how your data is set up



Several windows will now pop up with your results, as well as some interpretation.

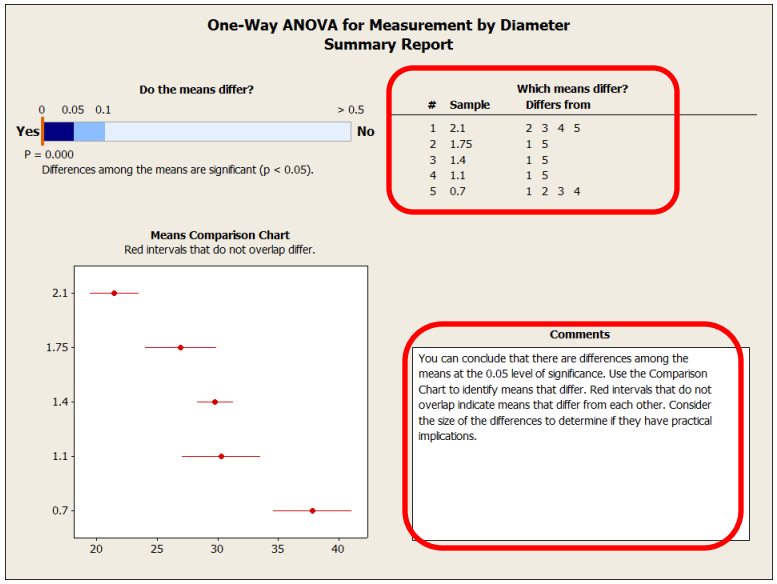
* This process in Minitab only works for 12 samples or less. *

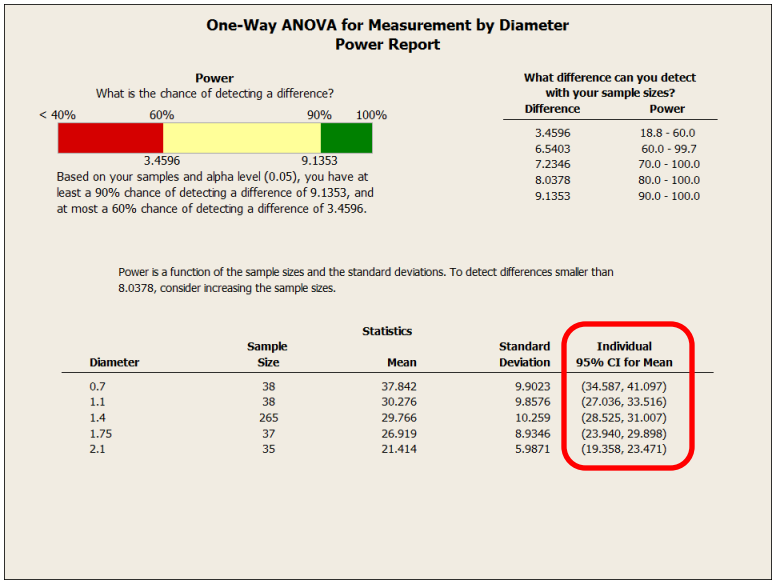
Use Tukey method if you have more than 12

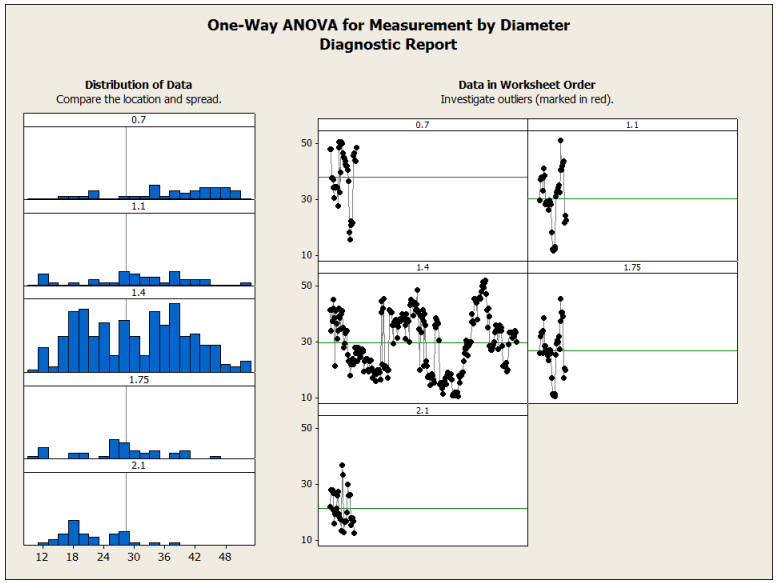
Return to Contents

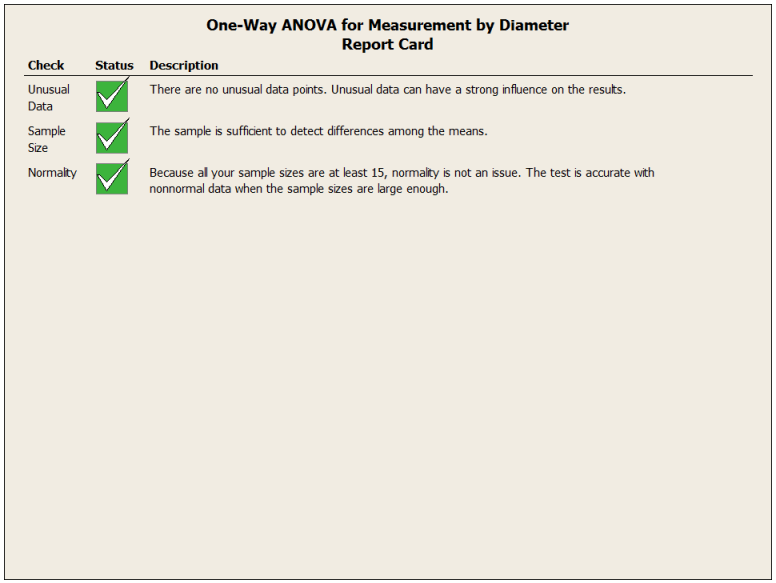
Assistant Output

Also provides confidence intervals for the means









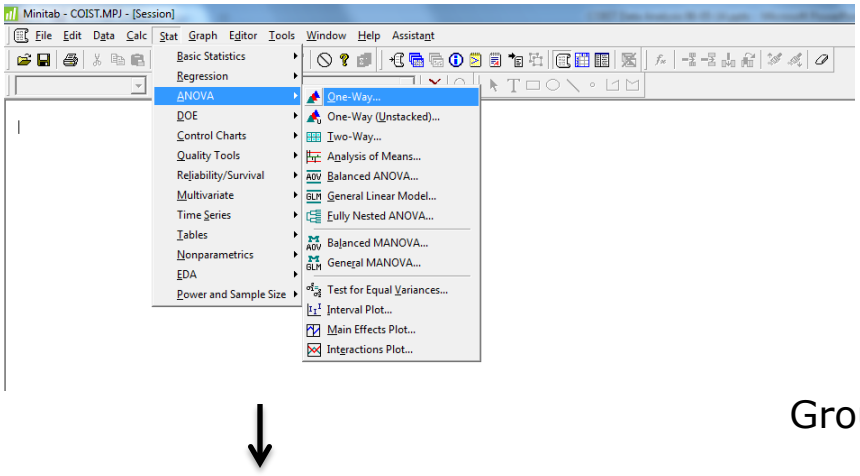
Return to Contents

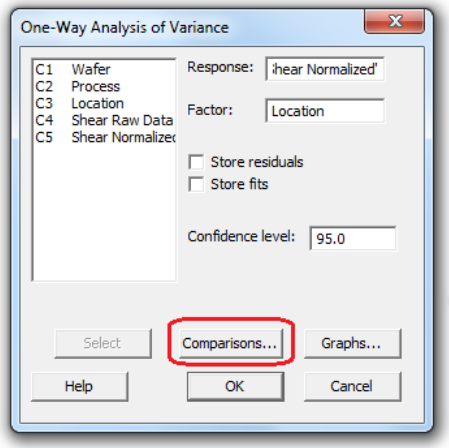
Tukey Comparisons

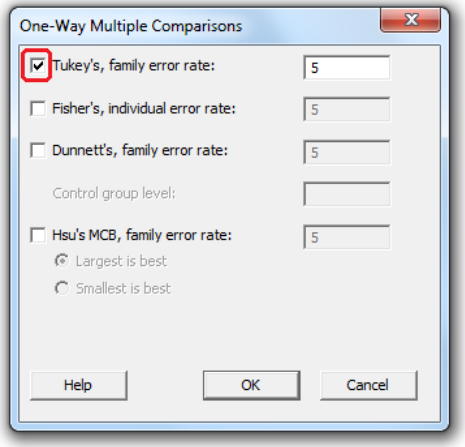
- Are means of samples/groups significantly different?
- Use Tukey's Method
  - Pair-wise comparisons of means
  - Calculates confidence intervals for Mean 1 – Mean 2, for all pairs.
  - If the confidence interval contains 0, no significant difference b/w the two means.
  - Then, using these intervals, can "group" means. Means in same group are not significantly different.
  - What's special about Tukey's method? Considers all comparisons simultaneously. Confidence levels of the individual intervals are adjusted in order to reduce error.
  - Use if you have more than 12 "groups" whose means you want to compare.

Return to Contents

MINITAB- TUKEY







Output

Groupings based on Tukey

Minitab - COIST.MPJ - [Session]

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Results for: COIST-019 thru COIST-143

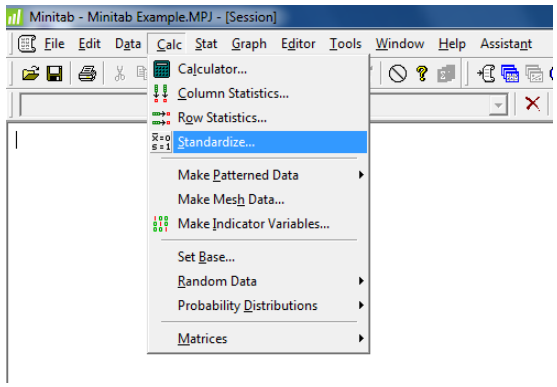
One-way ANOVA: Shear Normalized versus Location

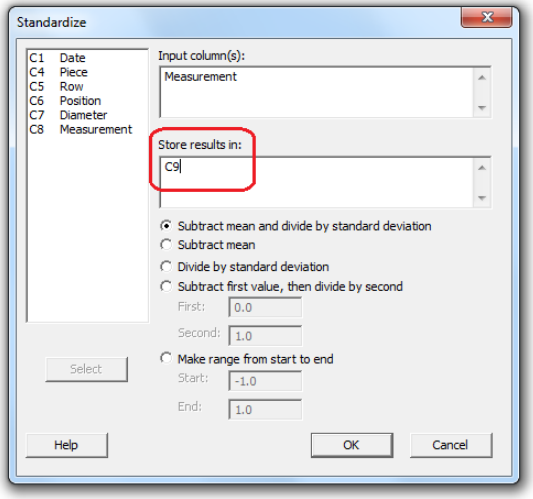
```
Source      DF       SS     MS     F      P
Location    18   268.75  14.93  6.62  0.000
Error     1117  2521.15   2.26
Total     1135  2789.90

S = 1.502   R-Sq = 9.63%   R-Sq(adj) = 8.18%

                                Individual 95% CIs For Mean Based on
                                Pooled StDev
Level    N    Mean  StDev  ----+---------+---------+---------+-----
1A       8  12.227  0.989                      (----*----)
1G       8   8.225  0.839  (----*----)
2A     216  10.699  1.716                (*)
2B      38  11.341  1.209                  (--*-)
2D       8   9.244  0.225       (----*----)
2G      15  11.195  0.727                (---*---)
4C       8  12.816  0.426                       (----*----)
4D     253  10.641  1.305                (*)
5D      24  11.038  0.644                (--*--)
5E      16  11.275  0.673                  (--*---)
6C      16  10.058  0.759             (--*---)
6E     232  11.116  1.554                  (*)
6F       8  12.049  0.757                  (----*----)
7F       8  12.678  0.747                     (----*-----)
7G      23  11.291  0.563                (--*---)
7H       8  12.866  0.815                       (----*-----)
8A      24  11.202  1.652                (--*--)
8F      16  11.697  1.118                  (--*---)
8G     207  11.081  1.843                 (*)
                                ----+---------+---------+---------+-----
                                  8.0      10.0      12.0      14.0

Pooled StDev = 1.502

Grouping Information Using Tukey Method

Location    N    Mean  Grouping
7H          8  12.866  A
4C          8  12.816  A
7F          8  12.678  A
1A          8  12.227  A B
6F          8  12.049  A B
8F         16  11.697  A B
2B         38  11.341  A B
7G         23  11.291  A B C
5E         16  11.275  A B C
8A         24  11.202  A B C
2G         15  11.195  A B C
6E        232  11.116  A B C
8G        207  11.081  A B C
5D         24  11.038  A B C
2A        216  10.699    B C
4D        253  10.641    B C
6C         16  10.058    B C D
2D          8   9.244      C D
1G          8   8.225        D

Means that do not share a letter are significantly different.
```


Return to Contents

Standardizing Data

- Sometimes, standardizing data is useful.
  - Can be easier to work with the data
  - Standardizing: For each data point, subtract off the mean, divide by standard deviation





Store your results in the next empty column in your dataset

Return to Contents

References

- Khan, R. M. (2013). *Problem solving and data analysis using minitab: A clear and easy guide to six sigma methodology* (1st ed.). West Sussex, United Kingdom: Wiley.
- http://en.wikipedia.org/wiki/One-way_analysis_of_variance
- http://en.wikipedia.org/wiki/Confidence_interval
- http://en.wikipedia.org/wiki/Statistical_hypothesis_testing
- http://www.itl.nist.gov/div898/handbook/pri/pri.htm

Return to Contents