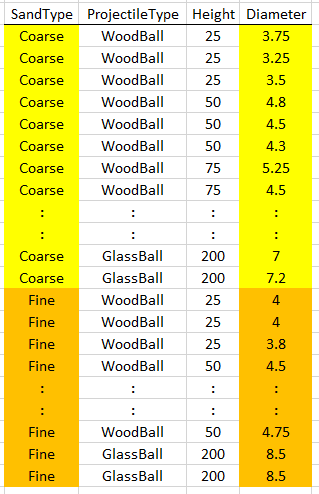
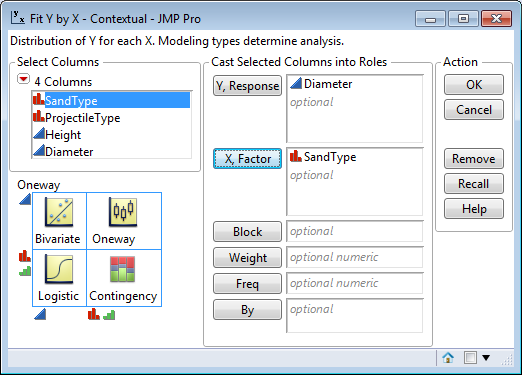
**Example 10.1** Consider the following study which investigated the effect of certain variables on the size of an impact crater.

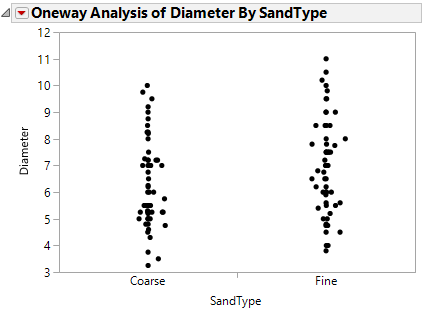
|  |  |
| --- | --- |
| Response Variable: Diameter of impact crater (cm)  Variables under investigation (i.e. independent variables):   * Height of which projectile was dropped * Project type (glass, steel, or wood ball) * Sand type (course, fine) | http://i.ytimg.com/vi/WZE25j_gMhU/0.jpg |

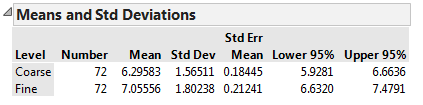


Looking at the effect of Sand Type.



A graphical display for making comparisons…





**Question**: From this data, do you think sand type has much of an impact on diameter of crater? Discuss…

Idea of a test for making comparisons between groups

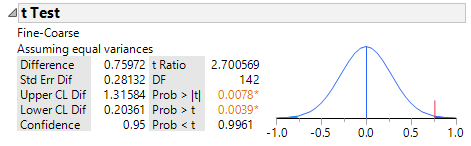


|  |  |
| --- | --- |
|  |  |

Sketch a set of data for which the Variation Between Groups is large compared to the Variation Within Groups

|  |  |
| --- | --- |
| **LARGE** | **VERY LARGE** |
|  |  |

Interpret the Test Statistic from this test.



**Assumptions**:

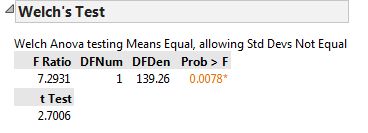
|  |  |
| --- | --- |
| j0213523[1] | 1. Groups must be independent of each other 2. Groups must have the same population variances 3. The observations from each group must follow a normal distribution |

1. Checking Independence Assumption:
2. Checking the Equal Variance Assumption:

|  |  |
| --- | --- |
| Checking this assumption in JMP | JMP Output: |



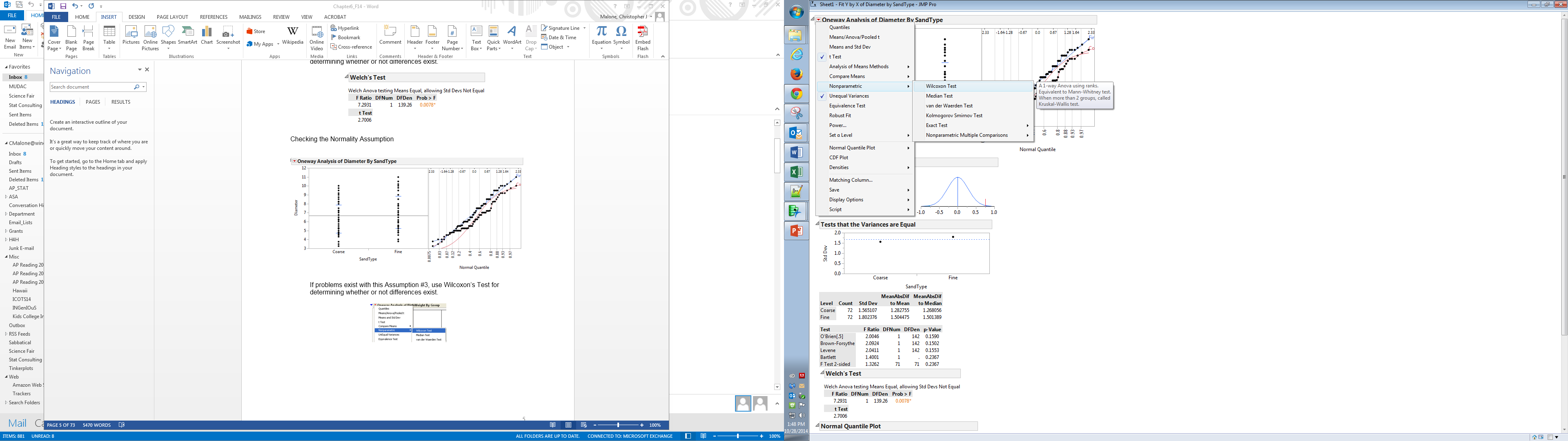
If problems exist with this Assumption #2, use Welch’s Test for determining whether or not differences exist.

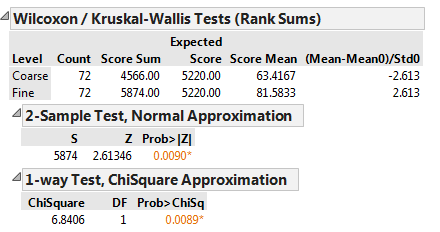


1. Checking the Normality Assumption

|  |  |
| --- | --- |
|  |  |

If problems exist with this Assumption #3, use Wilcoxon’s Test for determining whether or not differences exist.

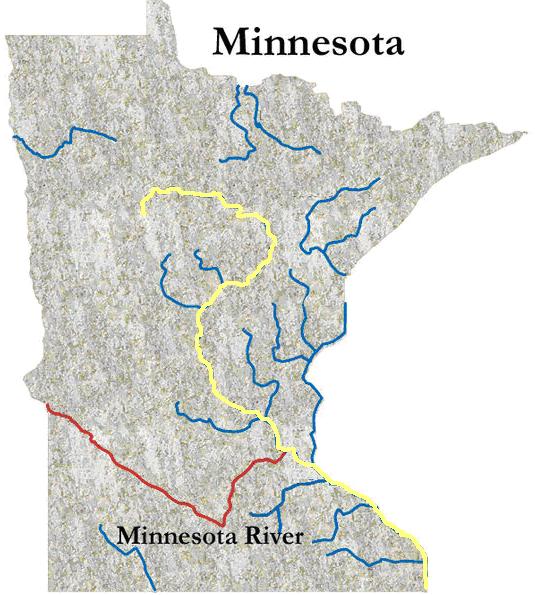




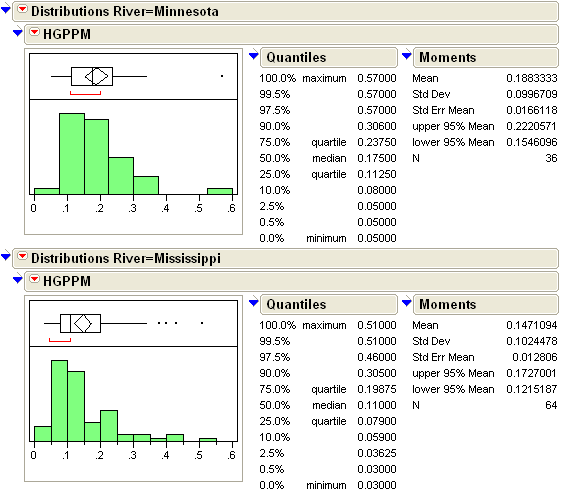
|  |  |
| --- | --- |
| Step 0 | **Research Question:** Is there statistical evidence to suggest that a difference exists in average crater diameter across these two sand types? |
| Step 1 | Obtain null and alternative hypothesis.  HO : The Average Diameter is the SAME HA : The Average Diameter is DIFFERENT |
| Step 2 | Use a 5% an error rate which implies conclusions will be made with 95% confidence. |
| Step 3 | Completing the test in JMP…          Ratio (i.e. Test Statistic) : \_\_\_\_\_\_\_\_\_\_\_\_\_ |

|  |  |
| --- | --- |
| Step 4 | Determine p-value and make the statistical decision.  The Decision Rule: If the p-value is less than the error rate, then the data is said to support the alternative hypothesis.  P-Value: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Statistical Decision: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Step 5 | Conclusion – Writing a final statement in laymen’s terms |
| Step 6 | Compute the appropriate 95% confidence interval for this problem    Construct this interval on the line below.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Interpret the meaning of this 95% confidence interval. |

**Example 10.2** Consider again the MN Walleyes dataset. This dataset contains information on Mercury Levels and PCB Levels of Walleyes from two Minnesota Rivers – Minnesota and Mississippi.

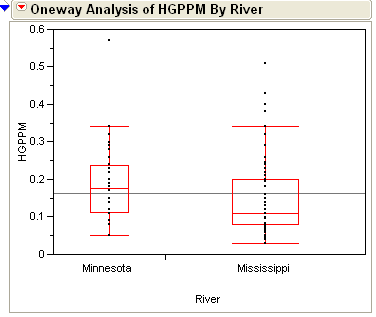


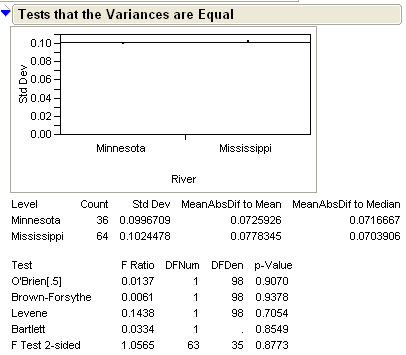
Comparing the summary statistics for the Mercury Levels across the two rivers



Checking Independence Assumption:

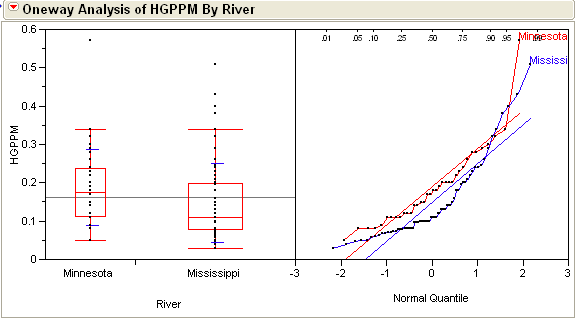
Checking the Equal Variance Assumption:

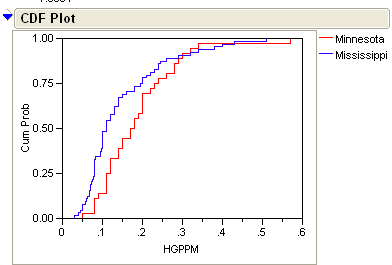






Checking the Normality Assumption



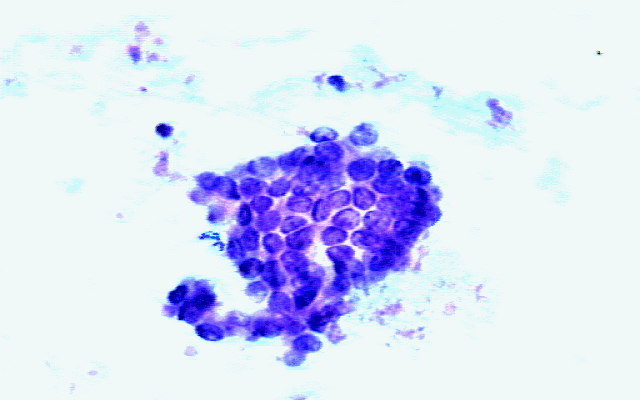


**Question**: What is the appropriate testing procedure?

|  |  |
| --- | --- |
| What does a “Normal” CDF plot look like? | |
| Step 0 | | **Research Question:** Is there statistical evidence to suggest that a difference exists in average Mercury Levels between Minnesota and Mississippi River? | |
| Step 1 | | Obtain null and alternative hypothesis.  HO : The Average Mercury Level is the SAME HA : The Average Mercury Level is DIFFERENT | |
| Step 2 | | Use a 5% an error rate which implies conclusions will be made with 95% confidence. | |
| Step 3 | | Completing the test in JMP…      Test Statistic : \_\_\_\_\_\_\_\_\_\_\_\_\_ | |
| Step 4 | | Determine p-value and make the statistical decision.  The Decision Rule: If the p-value is less than the error rate, then the data is said to support the alternative hypothesis.  P-Value: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Statistical Decision: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | |
| Step 5 | | Conclusion – Writing a final statement in laymen’s terms | |
| Step 6 | | Compute the appropriate 95% confidence interval for this problem   |  | | --- | | In Minitab |   In JMP      Construct this interval on the line below.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Interpret the meaning of this 95% confidence interval | |

**Example 10.3** Consider the BreastDiag.JMP data set. These data come from a study of breast tumors conducted at the University of Wisconsin-Madison.  The goal was determine if malignancy of a tumor could be established by using shape characteristics of cells obtained via fine needle aspiration (FNA) and digitized scanning of the cells.  The sample of tumor cells were examined under an electron microscope and a variety of cell shape characteristics were measured.    
  
The variables in the data file are:

* ID - patient identification number (not used)
* Diagnosis determined by biopsy - B = benign or M = malignant
* Radius = radius (mean of distances from center to points on the perimeter
* Texture texture (standard deviation of gray-scale values)
* Smoothness = smoothness (local variation in radius lengths)
* Compactness = compactness (perimeter^2 / area - 1.0)
* Concavity = concavity (severity of concave portions of the contour)
* Concavepts = concave points (number of concave portions of the contour)
* Symmetry = symmetry (measure of symmetry of the cell nucleus)
* FracDim = fractal dimension ("coastline approximation" - 1)



Medical literature citations:

W.H. Wolberg, W.N. Street, and O.L. Mangasarian.

Machine learning techniques to diagnose breast

cancer from fine-needle aspirates.

Cancer Letters 77 (1994) 163-171.

W.H. Wolberg, W.N. Street, and O.L. Mangasarian.

Image analysis and machine learning applied to breast cancer

diagnosis and prognosis.

Analytical and Quantitative Cytology and Histology, Vol. 17

No. 2, pages 77-87, April 1995.

W.H. Wolberg, W.N. Street, D.M. Heisey, and O.L. Mangasarian.

Computerized breast cancer diagnosis and prognosis from fine

needle aspirates.

Archives of Surgery 1995;130:511-516.

W.H. Wolberg, W.N. Street, D.M. Heisey, and O.L. Mangasarian.

Computer-derived nuclear features distinguish malignant from

benign breast cytology.

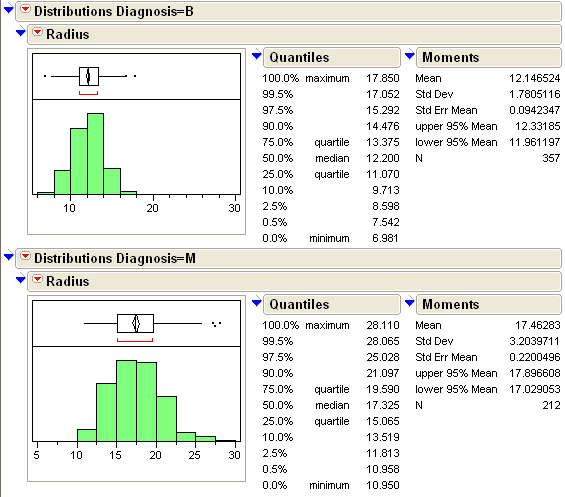
Human Pathology, 26:792--796, 1995.

See also:

<http://www.cs.wisc.edu/~olvi/uwmp/mpml.html>  
 <http://www.cs.wisc.edu/~olvi/uwmp/cancer.html>

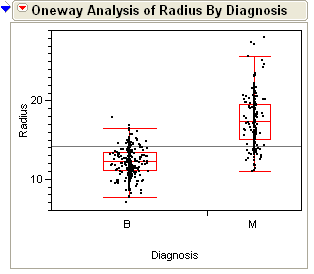
**Question-of-Interest:** Is there evidence to suggest that differences exist in cell radii of Malignant vs. Benign Breast tumors?

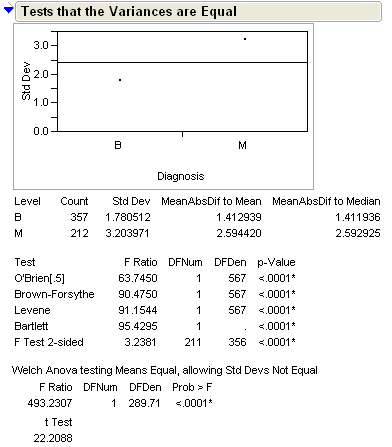
Comparing the summary statistics



Checking Independence Assumption:

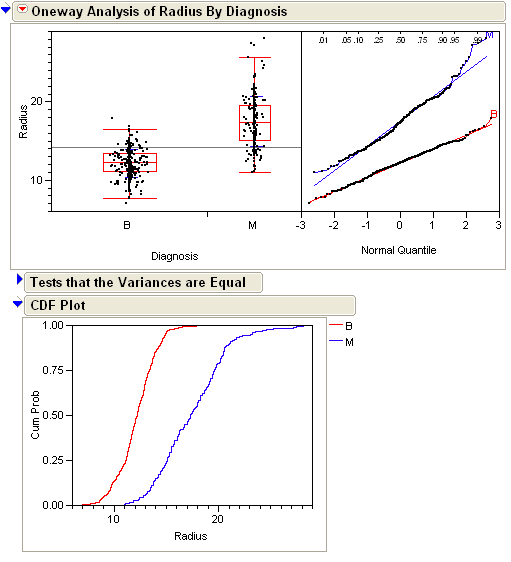
Checking the Equal Variance Assumption:







Checking the Normality Assumption



|  |  |
| --- | --- |
| Step 0 | **Research Question:** Is there statistical evidence to suggest that a differences exist in average cell radii of Malignant vs. Benign Breast tumors? |
| Step 1 | Obtain null and alternative hypothesis.  HO : The Average Cell Radii is the SAME HA : The Average Cell Radii is DIFFERENT |
| Step 2 | Use a 5% an error rate which implies conclusions will be made with 95% confidence. |

|  |  |
| --- | --- |
| Step 3 | Completing the test in JMP…      Test Statistic : \_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Step 4 | Determine p-value and make the statistical decision.  The Decision Rule: If the p-value is less than the error rate, then the data is said to support the alternative hypothesis.  P-Value: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Statistical Decision: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Step 5 | Conclusion – Writing a final statement in laymen’s terms |
| Step 6 | Compute the appropriate 95% confidence interval for this problem    Construct this interval on the line below.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Interpret the meaning of this 95% confidence interval. |