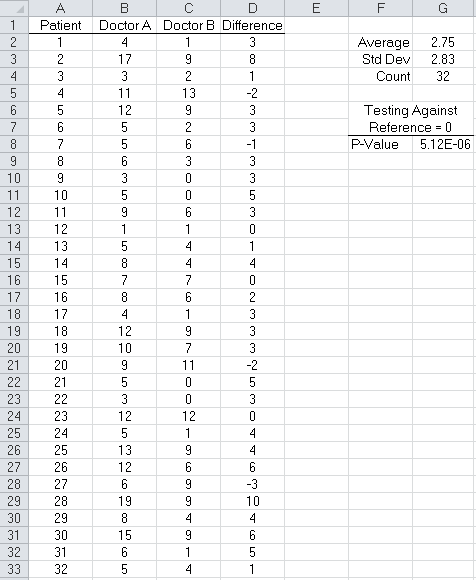
**Chapter 7: Relationships Amongst Numerical Variables**

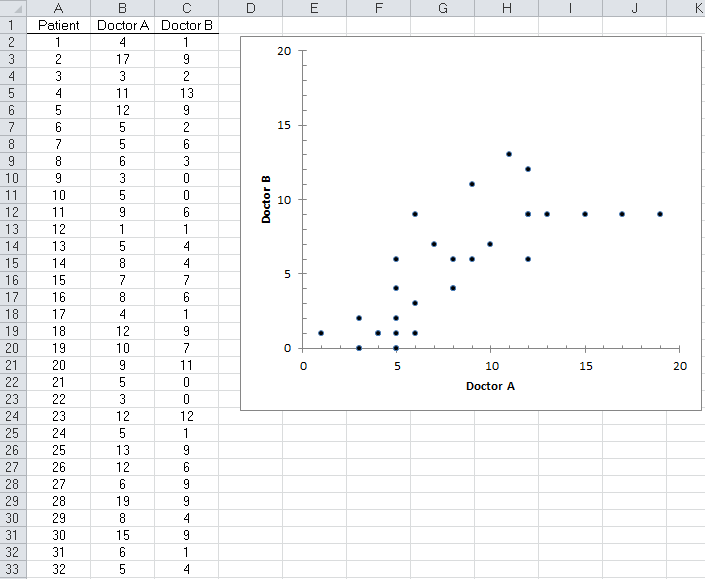
**Section 7.1: Descriptive Methods**

**Example 7.1** Consider the degree of clinical agreement among two different physicians on their assessment of generalized lymphadenopathy. The data from this study is given here along with summary statistics and the outcomes from the paired t-test (see Section 6.1).



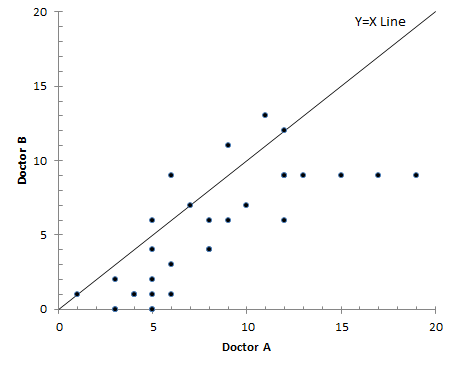
|  |  |
| --- | --- |
| Some concerns with looking at the differences |  |

Looking at a scatterplot, or as Excel calls it an XY Scatter graph of Doctor A vs. Doctor B



What can we say about the differences between these two doctors? Discuss…

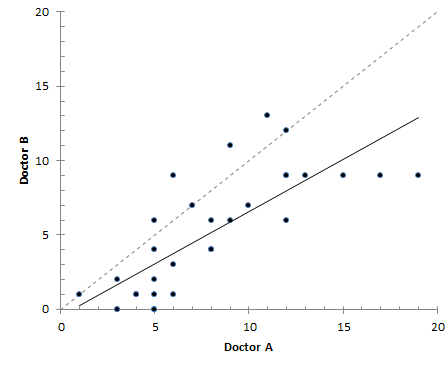
On this plot, we want to make sure the scales on the two axes match as a 1:1 relationship is expected. Also, we can/should add a Y=X reference line to the plot.



**Questions**:

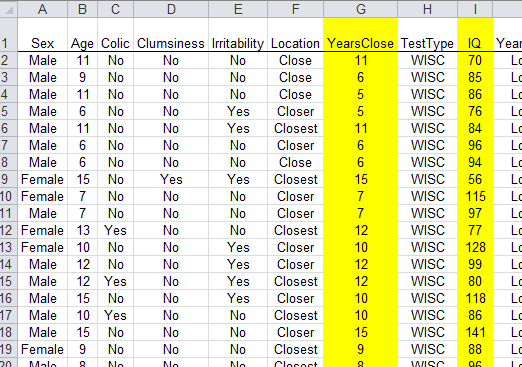
1. What does the Y=X reference line mean?
2. How many points fall exactly on the line? What does this mean?
3. Most of the points fall below the Y=X reference line. What does this tell you about the amount of agreement between these two doctors?

Adding the trend line (from the data) to this plot…

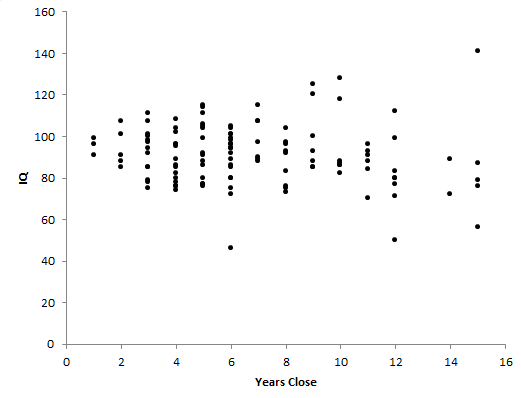


1. Notice that the trend line (from the data) and Y=X reference line start out about in the same spot. What does this tell us about the amount of agreement between these two doctors?
2. The trend line appears to be more flat than the Y=X reference line. What does this tell us about the amount of agreement between these two doctors?

**Example 7.2** Consider the El Paso lead study. Of interest here is the relationship between YearsClose and IQ. That is, does a living near the lead pollution source have an impact on IQ?



Here is a scatterplot showing the relationship between YearsClose and IQ.

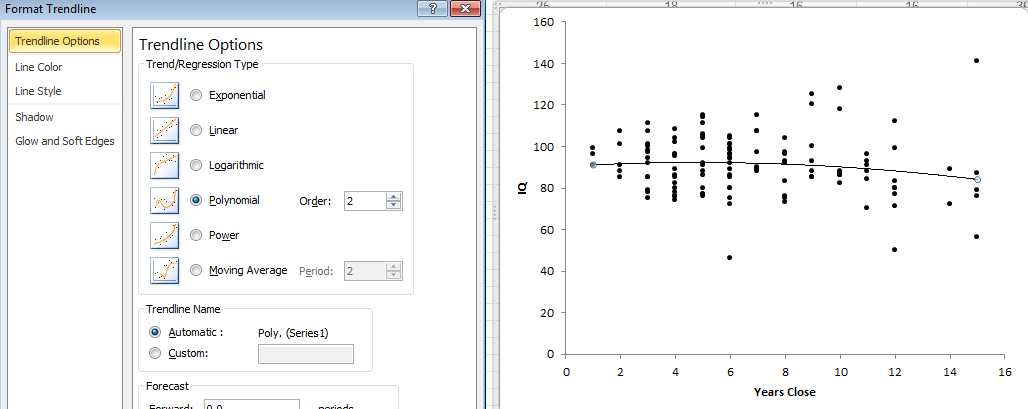


1. Discuss the general trends in this plot.
2. What impact does YearsClose have on IQ?

Adding a trendline in Excel, right click on dots and select Add Trendline…

|  |  |
| --- | --- |
| Adding a trend line in Excel |  |

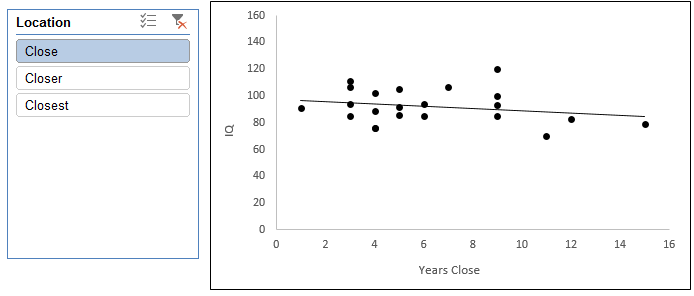
Maybe a trend curve would be better than a trendline…

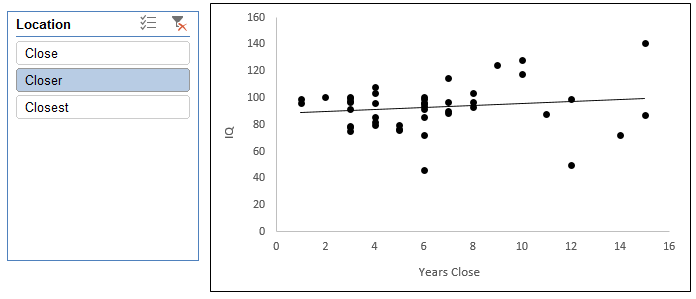


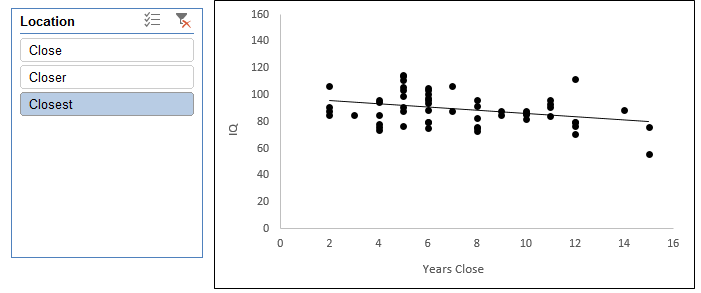
Using a Slicer to understand trends across groups…

|  |  |  |
| --- | --- | --- |
| Select Insert > Slicer | Select the variable(s) for the slicer | The Slicer window… |

Looking at the relationship between Years Close and IQ for each Location







Getting summaries with PivotTables – note that Location has been placed in the FILTERS box….

|  |  |
| --- | --- |
| Setup for the PivotTable… | Location = (All), i.e. Close, Closer, and Closest |

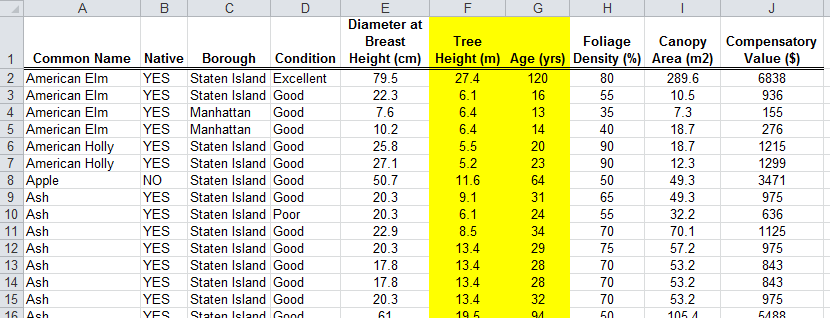
To apply the filter, simply make a selection from the Location drop-down box….

|  |  |  |
| --- | --- | --- |
| Location = Close | Location = Closer | Location = Closest |

1. Do the patterns we observe in the above table agree with saw in the scatterplot? Discuss.

**Section 7.2: Modeling Y using One X variable**

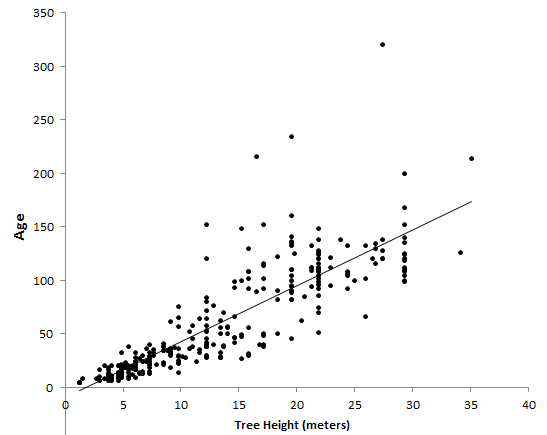
Consider again the NYC Tree dataset that we have discussed previously in class.



Of interest here is understanding the relationship between a Tree’s height and it’s age. Typically the age of a tree is determined by counting the growth rings; however, this requires cutting the tree down.

|  |  |  |
| --- | --- | --- |
| http://karenswhimsy.com/public-domain-images/tree-clipart/tree-clipart-4.jpg | < --? Relationship ? -- > | http://www.microscopy-uk.org.uk/mag/imgjan02/PJRing2.jpg |

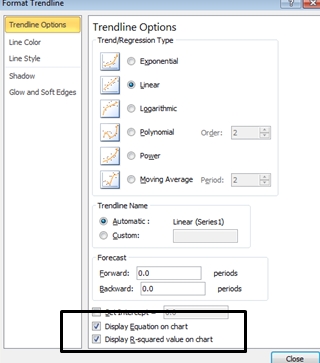
Consider a scatterplot showing the relationship between Tree Height and Age of tree.



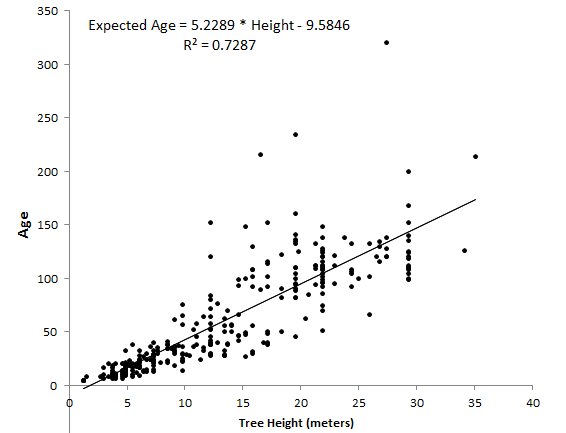
Questions:

1. What is the general trend in this plot?
2. The data points on the lower end are closer to the line than the points on the upper end. In the context of this problem, what does this mean?

In Excel, we can obtain the exact equation for this line. Typically a R2 value is reported as well. Select each of these on the bottom of the Trendline window.



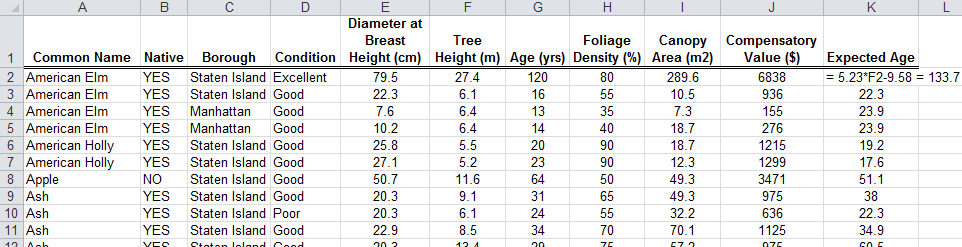
The scatterplot with the equation and R2 value.



Questions:

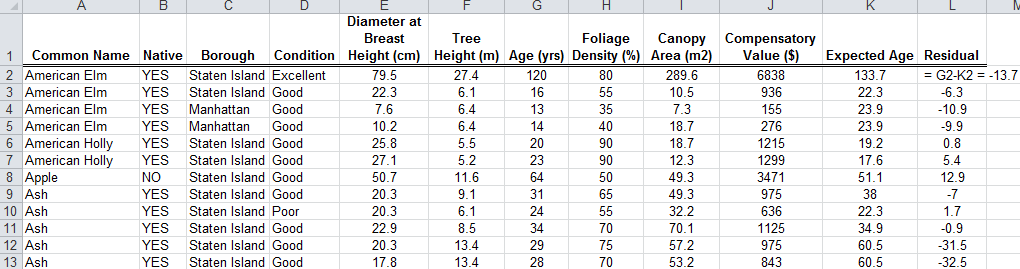
1. What does the value of -9.5 mean in the context of this problem?
2. What does the value of 5.23 mean in the context of this problem?

Use this equation to obtain the expected age for the trees in our dataset.

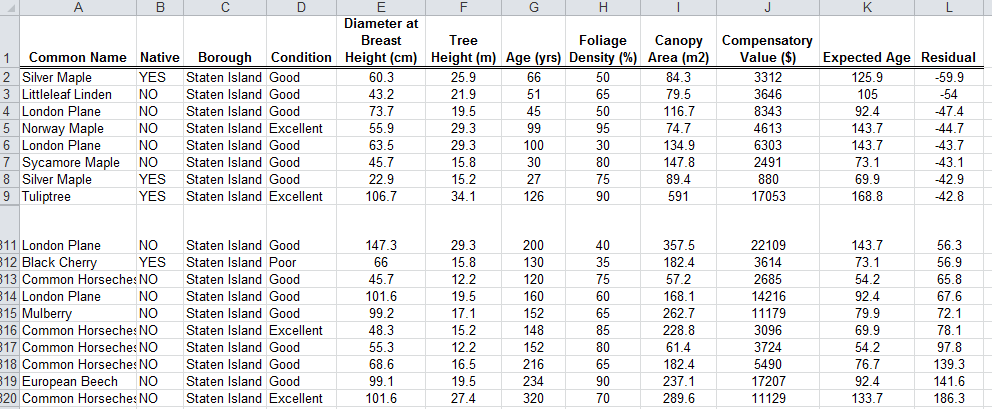


* Identify trees for which the expected (i.e. predicted) age closely matches the actual age
* Identify trees for which the expected (i.e. predicted) age does not match the actual age very closely

Getting the residual values in Excel…



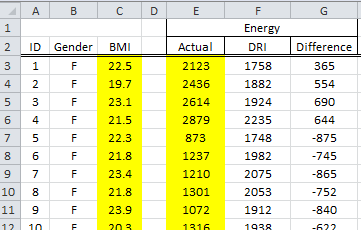
List of smallest and largest residuals



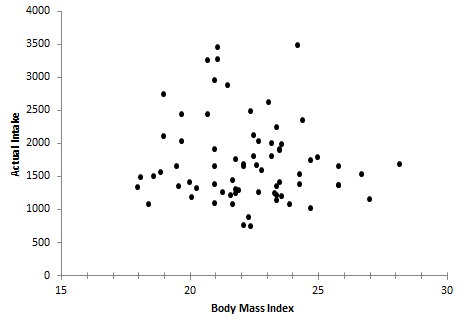
**Section 7.3: Modeling Y using Multiple X variables**

**Example 7.3** For this example, consider the Vitamin Intake dataset. This investigation was done here at Winona State with students and was centered on the actual intake of vitamins versus their daily recommended intake (DRI). The DRI values may between different people.

Research Question: What is the impact of body mass index, i.e. BMI, on the Energy Intake, i.e. Calories?



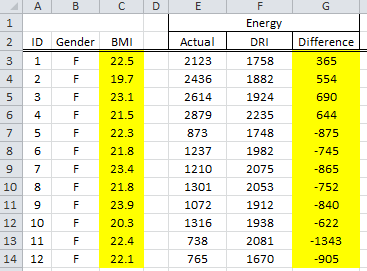
Scatterplot of relationship between BMI and Actual Energy Intake.



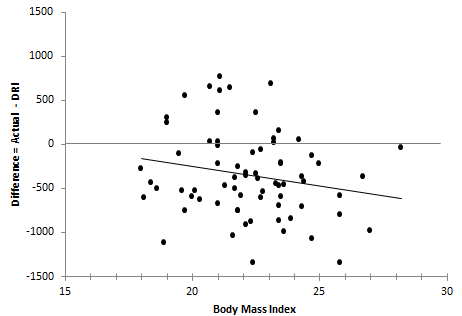
Discuss:

Taking into consideration the DRI values would require using the Difference column instead of the actual intake.

Difference = Actual - DRI



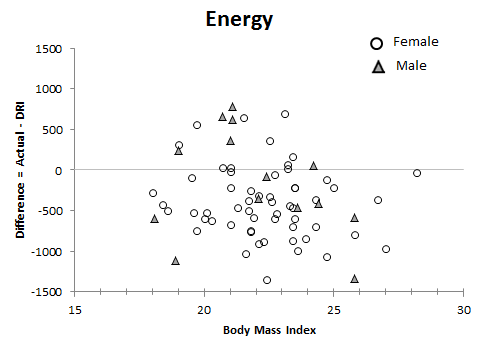
Scatterplot with trend line added.



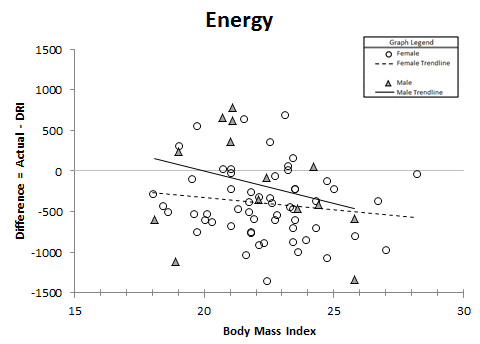
Discuss:

Research Question: What is the impact of BMI on the Energy Intake? Do differences exist between Males and Females?

Question: How can we add the effect of Gender into our investigation?



Adding trend lines to help identify the trends.



Discuss:

What about the other vitamins/minerals?



Discuss: