 Creating Residual Plots in Excel 2016

You should have already created a scatterplot using Funny and Happy for the males in the Excel tutorial file following the directions in the “Scatterplot and Correlation” tutorial and fit a linear model to the data following the direction in the “Linear and Exponential Models” tutorial.

After you have created a scatterplot and linear model for your data, you can create Residual Plots to see if your data meets the assumptions for a linear model.

1. Copy and paste your data (Funny and Happy for males) into a new sheet. The Data Analysis Toolpak will give you an error if you try to conduct the analysis with any missing data. You must first remove any cases which are missing values for either variable. To do this, sort your data by your first variable. Any missing values will be found at the bottom of your screen. Delete the entire row of data for any cases missing data on the first variable. Then repeat this process sorting by the second variable.

2. Open the Data Analysis Toolpak from the Data tab. Select “Regression” from the menu.

3. In the ‘Input Y Range’ select the data select the data for your dependent variable. Be sure to include the column header and then all of the data. NOTE: in the ‘Regression’ function of the Toolpak you cannot include the entire column of data. If you select the entire column (ex. $A:$A in the box) you will get an error.

4. Repeat this for the ‘Input X Range’ with the independent (predictor) variable.

5. Check the “Labels” box if the first row of the data you selected includes the column title.

6. Check the “Residual Plots” box. This plots the residuals against the independent variable (X).
7. Inspect the residual plot for any clear patterns which may indicate a linear model is not appropriate for your data.

Note: You should see no pattern in your residual plot; the scatterplot should look entirely random. If you see a clear curvature to your points, this indicates your data may not be linear, and a linear model is not appropriate. If you see that the spread of the points is not consistent throughout, this may indicate you have violated the equal spread or the Does the plot thicken? condition. Also, look to see if there are any outliers present in the data as they will be easier to identify here.

Examples of common data problems: