MATH 124 – Draw residuals plots using the Data Analysis Toolpak
## bolstad_math124@bmbolstad.com
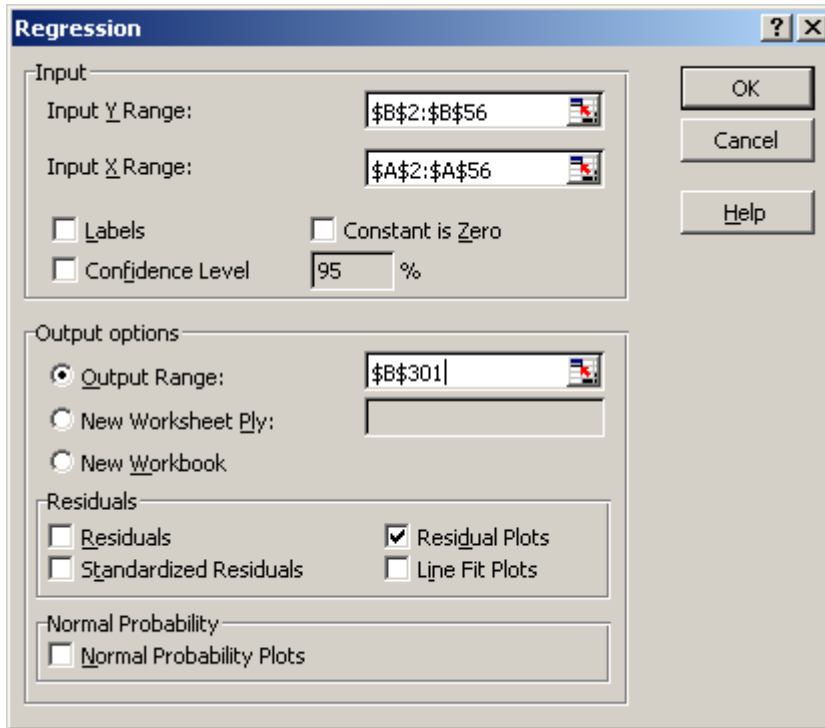## http://math124sfsu.bmbolstad.com

This document explains how to use Data Analysis toolpak to fit a simple linear regression, return the fitted values and residuals and get residuals plots. It is expected that you have already read the "Transform" and the "Find Regression Coefficients using the Data Analysis Toolpak" documents before you reached this point (if not now would be a good time to do so). We will again concentrate on the first dataset in the transform data file. A completely worked spreadsheet for all 4 datasets (and the various models) is available on the website.

**Using the data analysis toolpak**

1. It was explained in the Histogram document how to install the Analysis Toolpak. If you do not have it installed do so now.

2. As before to fit the regression model using the Analysis Toolpak. Go to the Tools menu and select the Data Analysis Option. Then choose "Regression" from the list and click ok.
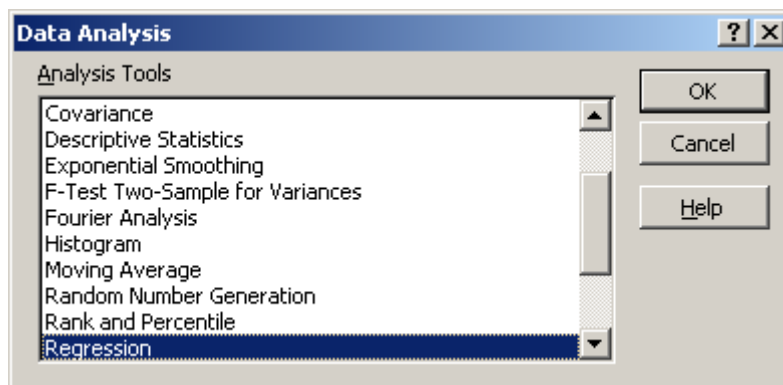

3. Fill in the fields in the dialog box in a similar manner to the picture below. In particular the Input Y Range should be the set of cells containing the response variable (B2:B56), the Input X range should be the set of cells containing the explanatory variable (we will use the untransformed X values in this example) and we will place the output back into the original spreadsheet at any empty location (B301). Make sure you have the Output Range radio button clicked. Also now make sure that the "Residuals Plot" option is checked. Click ok to finish.

Regression ? ☒

Input
Input Y Range: `$B$2:$B$56` 🔺
Input X Range: `$A$2:$A$56` 🔺

☐ Labels   ☐ Constant is Zero
☐ Confidence Level `95` %

OK
Cancel
Help

Output options
◉ Output Range: `$B$301` 🔺
○ New Worksheet Ply:
○ New Workbook

Residuals
☐ Residuals   ☑ Residual Plots
☐ Standardized Residuals ☐ Line Fit Plots

Normal Probability
☐ Normal Probability Plots

4. You should now find output that looks something like you saw before with a Summary Output table, but two new things have a appeared. A table containing predicted (aka fitted) y values with residual values and a plot of the residuals against the explanatory variable. The plot should look something like this

X Variable 1 Residual Plot
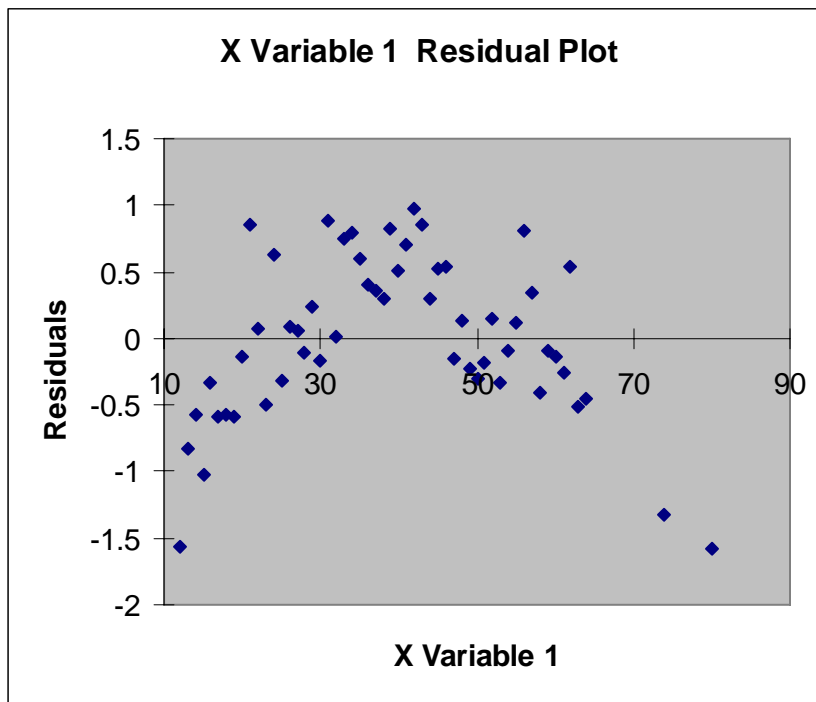
5. Before we interpret this plot we should more appropriately label it (Excel by default does not give the best labeling). Specifically, we should change "X variable 1" to "X1" (our explanatory variable) and adjust the title accordingly. Doing so should give you a residuals plot that looks something like this.

Residuals vs X1 Plot
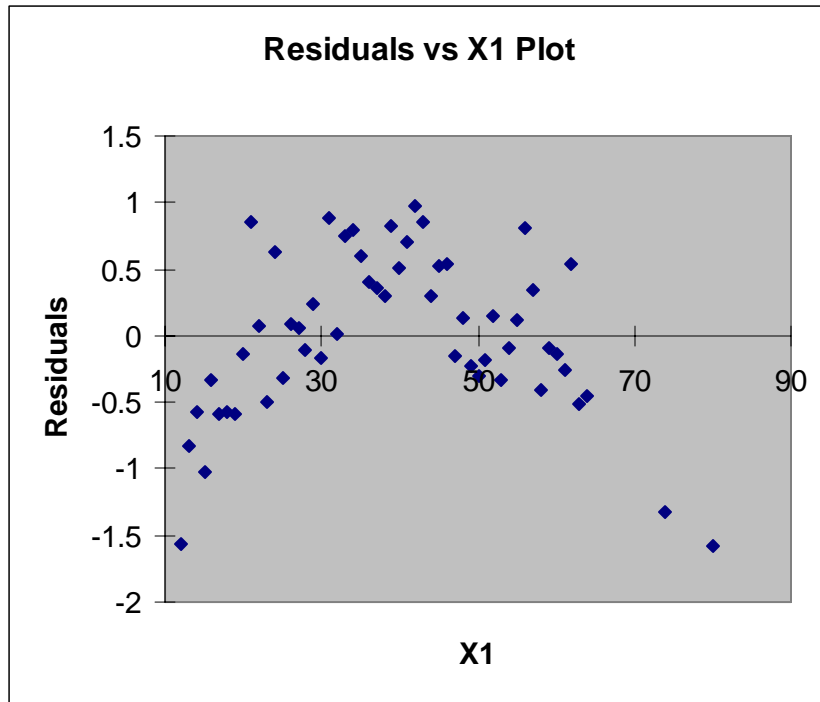
6. Next notice that there is a clear non-linear curve pattern apparent on this plot. This indicates that the regression line did not completely model the relationship between X1 and Y1. As we have seen in the other documents a log transformation on the X1 variable gives a better linear regression fit. In a few moments we will look at the Residuals plot for that case. Before we do that though we could create a plot of the residuals versus the predicted Y1 values (this is another standard residuals plot). Luckily for us we already have the residuals (in D325:D379) and predicted Y1 values (in C325:C379) stored for us in the previously produced table. Creating a scatterplot of Residuals vs Fitted Y1 values using the standard methods should give you a plot that looks like this

Residuals vs Fitted Y1
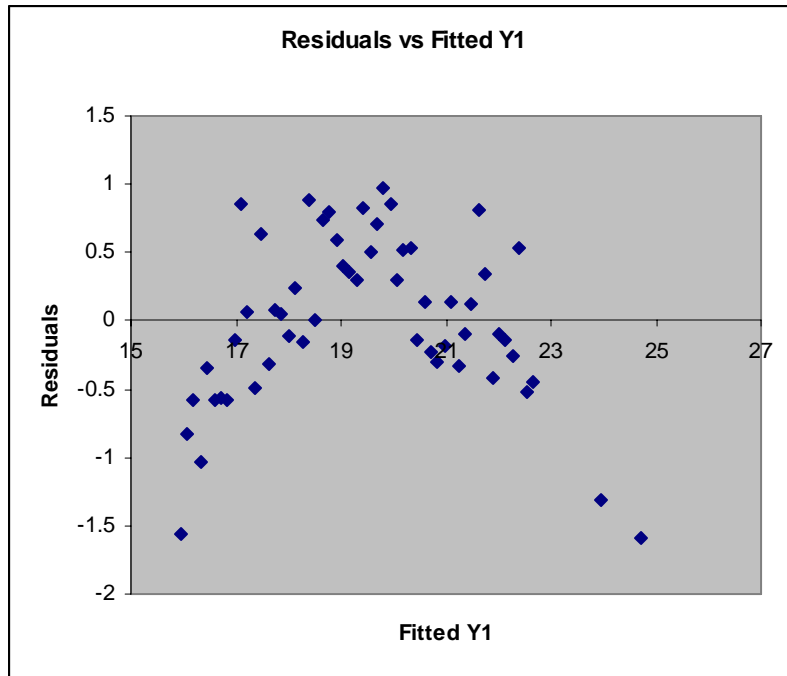
7. Again notice the curve is visible in the plot. In general if either (or both of) the plots of Residuals versus Explanatory variable or Residuals versus Fitted response variable show a pattern then the linear regression line did not fit the observed data well.

8. Next repeat the steps above, but using the log(X1) values as the explanatory variable and putting your output into B383. You should see that the output Residuals plot (after fixing up labeling) looks like this:

Residuals vs log(X1) Plot
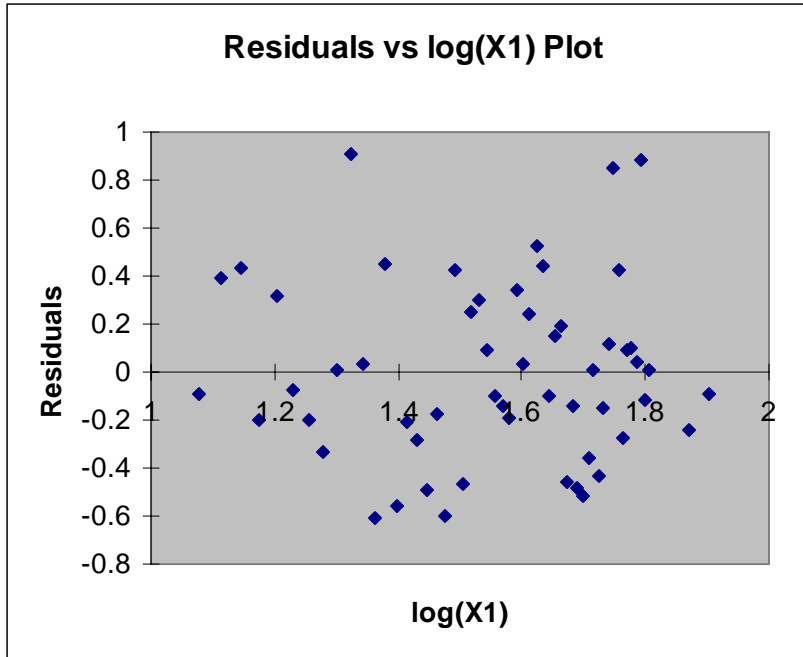
9. Now the scatter of the residuals seems to be even and with out apparent pattern. This means that the linear regression was appropriate for the transformed data.

10. The worked excel spreadsheet has additional Residuals plots for the other datasets.