Predictive Data Analytics

Assignment 3: The Perils of Regression

To complete assignment 3, you will need to follow the steps below. It is recommended that you read through each step before starting the assignment. I apologize in advance for the inevitable typos.

Step 1: Download the Excel File from Blackboard

On the course blackboard site, in the section called "Assignments", you will find a folder titled "Simple Multiple Regress". Inside that folder will be an Excel file titled "Data". Download that spreadsheet and open it.

Step 2: Simple Regression

The tab called "Simple Regression" contains two tables. The first is a set of data along with some regression predictions that depend on the values found in the table called "Your Model". Here is the breakdown of the data in the first table:

- **Obs.** is the observation number. It runs from 1 to 10.
- **X** is the set of values for an independent variable being used to predict the dependent variable.
- **Y** is the set of values for the dependent variable being predicted by the regression.
- **Y-hat** is the set of predicted values found by multiplying X by the coefficient for β_1 and then adding the value for β_2. This is what the regression line predicts for Y for a given value of X.
- **e** is the set of residuals for the regression. It is the difference between Y and Y-hat.
- **e^2** is the residuals squared. That simple.

Investigate the table of data and familiarize yourself with the equations used to calculate the regression predictions. Then start entering different values for β_1 and β_2 in the "Your Model" table. You'll notice things change when you do this. "Sum of Error" adds up the values of column F in the data table. "SSE" adds up the values in column G in the data table, and is otherwise known as the "Sum of Squares Error". Below this is the SSX or "Sum of Squares for X" which you will need to calculate the S.E. of the Estimate, or "Standard Error of the Estimate" below. And in the last row, you have to calculate the S.E. of β_1 or the "Standard Error of the Slope".

$$\hat{\beta}_1 = \frac{S_{xy}}{S_x^2} = \frac{cov(x,y)}{var(x)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$S_e\ (S.E.of\ the\ Estimate) = \sqrt{\frac{SSE}{n-2}}$$

$$S_{\hat{\beta}_1}(S.E.of\ \beta_1) = \frac{S_e}{\sqrt{SSX}}$$

You need to calculate values for each of the tan cells in the spreadsheet. For β_1 and β_2 you will need to write your own formulas which reference the data for X and Y in the data table. For SSX, remember that the SSX is the numerator of the variance of X, which you can find by multiplying the variance by n-1. For S.E. of the Estimate and the S.E. of β_1 you will again need to write your own formula which uses the SSE, SSX, and the count of observations.

Step 3: Upload your Excel File through Blackboard

Save your completed Excel file and upload it to Blackboard using the "Assignment 9 Submit" link found in the "Assignment 9" folder. In the submission form, you will need to attach your spreadsheet and click submit. Students who fail to submit their Excel file will have their quiz score reduced to a 0.

Step 4: You're Done!