Lab Assignment 5: Simple Linear Regression

Fitting a Regression Line.
Download the class data from the course website, and open it with JMP-IN. Let’s explore our class survey data.

1. Analyze the distributions of all the variables. Are any of them normally distributed? If yes, which ones? Which results do you find the most surprising?

2. Look at the correlations between all of the continuous variables (see lab 4 for directions).
   - Choose the variables that have strongest positive correlations? Repeat for negative correlations.
   - Which pair(s) of variables appear to have the weakest correlations?
   - What are the correlation values for the strongest positive, strongest negative and weakest correlations?

3. Plot Exercise (x) and GPA (y).
   - Describe the shape of the scatter plot.
   - Is the relationship between Exercise and GPA postively or negatively correlated?
   - Add the 95% density ellipse to the scatter plot. Are there any outlying points? If so, what is their gender? To remove the ellipse, select the red arrow next to fit below the scatter plot and choose Remove Fit.
   - Let’s fit a regression line to our data. Click on the red arrow next to Bivariate Fit and choose Fit Line. What is the regression equation? How are these values interpreted?
   - Do students with higher GPA’s tend to exercise more or less than those with lower GPA’s?
   - Does this mean extra exercise causes low GPA’s?
   - What is the rms error?
   - How well does the line fit the data?
   - Select the red arrow next to Linear fit and choose Plot Residuals. Are the residuals homoscedastic? Is there a pattern to the residuals?

4. To fit the regression lines by class year, first remove the previous fitted line by selecting the red triangle by Linear fit and choosing Remove Fit. Then, select the red triangle beside Bivariate Fit and choose Group By and then select Class. Now, go back to Bivariate Fit and select Fit Line, this will calculate the regression lines for the each class separately. Remove the fit for grad students.
   - Is the relationship the same for these three groups? If not, explain how they differ.
   - Which group has the smallest slope? Largest?
   - Are the correlations between the groups positively or negatively correlated?
• How do the $R^2$ and RMS error values differ between the groups?
• For which group is Exercise a better predictor of GPA?

**Eyeing the Least Squares Line.**
The principle of least squares can be easily seen with one of the sample scripts included in the Sample Scripts folder. Select **File → Open**, from the Open Dialogue box, change the **Files of Type** drop down to list ”JSL Scripts (*.JSL)”, then select the subdirectory **JMP IN Scripts**, and choose the script **demoLeastSquares.jsl**. To execute the script, select **Edit → Run Script**.

You will see a scatter-plot with two small rectangles on it. These two rectangular handles are draggable, and are used in this case to move the line to a position that you think best summarizes the data. Press the **Your Residuals** button. Use the handles to move the line around until you think the residuals (in blue) are as small as they can be. (Be sure to click on the handles. If you miss, the program will create a new data point. If this happens click on **Delete Last Point** to recreate original dataset.) Press the **Your Square** button. You will see each residual expanded into a square with the height of the square equal to the value of the residual. The area of these squares represent the quantity $(\text{predicted} - \text{actual})^2$. Again try to minimize the total area covered by the blue squares. Press the **LS Line** and check to see if you correctly determined the least squares line. Press the **LS Residuals** and **LS Squares** as a further check. How close was your line to the actual least squares line? Close the least squares demonstration window, and then close the script window.