JMP® Introductory Lab Activities

Activity 5: Least Squares Regression

Data Set: Big Class.jmp

Summary

In this lab you will be creating scatter plots and finding the least-squares regression line for bivariate data. You will also group the data and create a scatter plot with least-squares regression lines for each group.

At the end of the activity you will write a report containing your graphs, commentary and interpretations of your results (required output and discussion is in italics).

The Big Class Data

Open the file **Big Class.jmp** (go to **Help > Sample Data**, then click **See an Alphabetical List of all Sample Data Files**).

The data table contains five columns and 40 rows. This data represents 40 students who were randomly selected from the 300 students who are enrolled in a martial arts class. Their teacher has to buy new uniforms for an upcoming competition and has taken measurements to help her. For each student, four variables are shown: age in years, sex, height in inches and weight in pounds.

You are going to examine the relationship between **height** and **weight** for this class of students. If these two variables are related to one another, a scatter plot should show a pattern of some sort.

First, decide which variable you think makes the most sense for the role of the explanatory variable. *Briefly explain why you made this choice.*

Next, try to anticipate what you think the pattern of the scatter plot might be. *Briefly describe the expected pattern*.

Exploring the Relationship Between Weight and Height

In JMP, go to **Analyze > Fit Y by X**. This opens a dialog box where you specify which variables to analyze. Select the variable you chose as the response variable for the **Y, Response** variable. Next, select the variable you chose as the explanatory variable for the **X, Factor** variable, and click **OK**.

Describe the pattern you see in the resulting scatter plot. Is this close to what you expected?

To produce a regression line, click on the **red triangle** and select **Fit Line**. You should now see the scatter plot with the least-squares regression line drawn on it. You will also see an equation for the line, along with other output. At this point you do not need to be concerned with the Analysis of Variance table. Click on the gray triangle next to the table to close it.

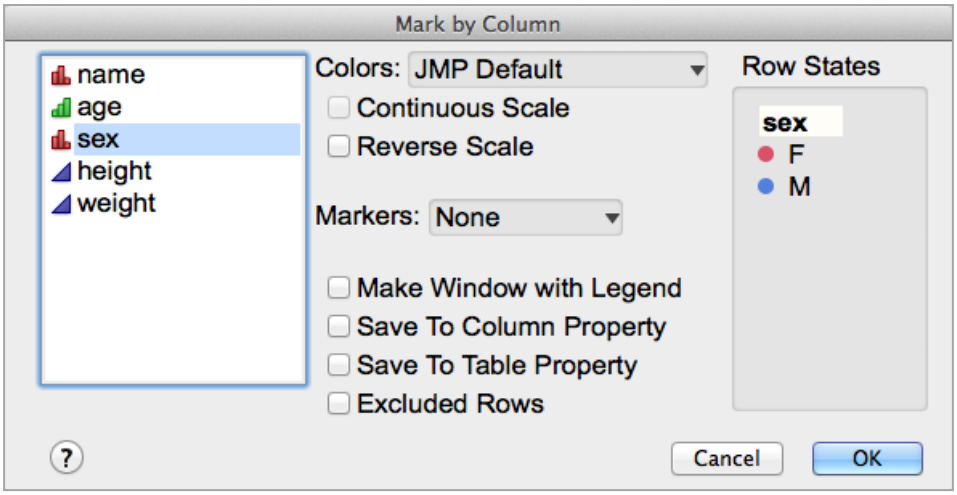
A copy of the scatter plot with the least-squares regression line should be included in your report. Include the equation of the line as well.

Include an interpretation of the slope for this line in the context of this problem. Also determine whether or not the y-intercept has any meaning in the context of this problem.

Exploring Differences Between Males and Females

Because both male and female students are represented by the data, you can do a separate analysis on each sex to get a sense of whether there is a difference in the distribution of weight and height by sex.

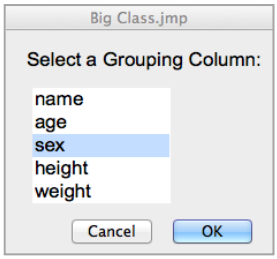
In order to distinguish between the two groups on your scatter plot, you can use a different color (and marker) for each sex. **Right-click** over the scatter plot, and select **Row Legend**. Then select **Sex** and click **OK** to get red and blue markers and a legend.



Click on “M” and “F” in the legend to select points.

What differences can you observe between these two groups?

JMP can also calculate individual lines for each of these two groups. Click on the top red triangle and select **Group By**. Select **sex** as the grouping column, and click **OK**.



Then click on the top red triangle again and select **Fit Line**. You should see two lines, one for each group.

Copy this scatter plot with the two regression lines to your lab report document. Also include the equations for both lines.

Do these lines indicate a difference between these two groups? In some detail, describe the differences that the lines indicate.

Include an interpretation of the slope (in context) for each line.

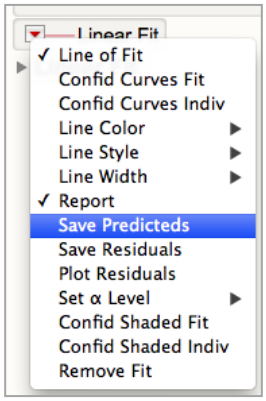
Making Predictions

This sample of students only includes 40 students. One common use of regression is to make predictions; in this case, our predictions would be about students who don't appear in the data (below).

Student Name	Age	Sex	Height	Weight
Shirley	17	F	67	122
Roberto	16	M	70	130
Charmaine	12	F	60	105
Huy	14	M	66	112

These predictions can be calculated using the equation for the line and a calculator. Or the predictions can be made directly in JMP.

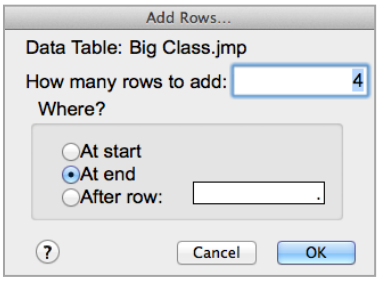
To use JMP to make predictions from the original line, click on the **red triangle** next to **Linear Fit** (below the graph in Fit Y by X), and select **Save Predicteds**.



This creates a new column in your Big Class data table containing the predicted heights or weights for all students. Not only are the predicted values stored, but the formula creating the data is also

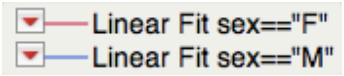
stored. To see this formula, right-click on this new column in the data table, and select **Formula**.

To find the predicted values for each of the four students, select **Rows > Add Rows**, and add four rows at the end of the data table. Or simply start typing at the bottom of the data table to add a new row.



Enter the values for the four students. After entering all of the values for a student, the predicted value will appear in the Predicted column for that student. (Note: Don't enter the value of the response variable you are trying to predict – it is not needed to make a prediction).

Return to the Fit Y by X output window, and save the prediction formulas for the line for **Females** and the line for **Males**.



The data table has two new columns – one for each line. Each column contains a second prediction for each of the four students. Change the name of the first new column to Predicted Female and the second column to Predicted Male.

Which predictions were generally more accurate (the predictions made using the first line, or the predictions made using the line for females or males)? (Be sure to use the correct predicted column, according to the sex of the student.)

Write Your Report

To summarize, your report should include the following components:

- *Your reasoning behind your choice of explanatory and response variables.*
- *Your prediction for the scatter plot pattern(s).*
- *The ungrouped scatter plot with least-squares regression line and the equation of the least-squares regression line.*
- *Analysis of how the scatter plot relates to your expectations.*
- *An interpretation of the slope in the context of this problem and a determination of the meaning (or non-meaning) of the y-intercept in the context of this problem.*

- *The scatter plot showing the two groups with the regression lines and the equations of the least-squares regression lines.*
- *An interpretation of the slope (in context) for each of the lines.*
- *Your discussion of the difference between males and females as shown on the scatter plot.*
- *The predictions you made for the height or weight of the four students above, along with your analysis of the accuracy of the prediction.*
- *A summary of the differences between the groups that the lines convey. Include detail here, and include your personal speculation about why these differences might exist.*