JMP INTRO® Lab Activities

Lab Activity – Least-Squares Regression

Data Set: Big Class

In this lab you will be creating scatterplots and finding the least-squares regression line for a bivariate set of data from an existing file. You will also group the data and create a scatterplot 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. Elements to be included in your report will appear in *italics type* throughout the activity.

Start your computer and open JMP INTRO. Open the data file named **Big Class** from the Sample Data folder. When you open the file, you should see data for 40 students, presented in a spreadsheet-like format. 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 scatterplot 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 scatterplot might be.

Briefly describe the expected pattern.

In JMP INTRO, open **Analyze** on the menu bar and select **Fit Y by X**. This opens a dialogue box that gives you the opportunity to specify which variables to analyze. Select the variable you chose as the response variable for the **Y,Response** variable. Then select the variable you chose as the explanatory variable for the **X,Factor** variable. Then click **OK**. Once you have drawn the scatterplot, the regression line is produced by first clicking the red triangle next to the heading **Bivariate Fit of *response variable* By *explanatory variable***. From the pop-up menu, select **Fit Line**. You should now see the scatterplot with the least-squares regression line drawn on it. You will also see several analyses of the data. At this point, you do not need to be concerned with the **Analysis of Variance** information and can close the analysis by clicking on the blue triangle to the left of the **Analysis of Variance** title.

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

Include a summary of how the scatterplot either confirmed or disputed your expectations of the relationship between weight and height for these students.

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.

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 or not there is a difference in the distribution of weight and height by sex.

In order to distinguish between the two groups on your scatterplot, you should use a different marker for each sex. You can specify a different color as well.

First, we are going to group the sexes together in the data table. To do this, open **Tables** on the menu bar and select **Sort**. This opens a dialogue box to specify which variable will be used for sorting. You are going to sort **By** the variable **sex**. Think of a name for the new table that will be created, and enter the title in the box by **Output Table**. Then click the option of **Sort**. You can also rename this data table by opening **Window** on the menu bar and selecting **Set Title**. Renaming this table may eliminate some of the confusion inherent in having so many windows open.

To select all of the female students, click on the area to the left of the first female student and drag down to the final female student, or click on the first female student, hold down the shift button, and click on the final female student. (You should have 18 rows highlighted.) Then choose **Rows ® Colors** to assign the color for female students and **Rows ® Markers** to assign the marker you'd like for female students. Repeat this process for the male students.

Make a new scatterplot, and you should be able to see male and female students are now represented with the colors and markers you chose.

What differences can you observe between these two groups?

JMP INTRO can also calculate individual lines for each of these two groups. Once again, click on the red triangle next to the heading **Bivariate Fit of *response variable* By *explanatory variable***. This time, choose the **Group By . . .** option, select **sex** and **OK** before opening this pop-up menu again and selecting **Fit Line**. You should see two lines, one for each group.

Copy this scatterplot 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.

This sample of students only includes 40 students. One common use of regression is to make predictions, in this case about students who don't appear in the data. Here are four additional students whose weights and heights are given.

Using the first fitted line you drew, make a prediction for each of the four students, using your explanatory variable to predict the response variable.

These predictions can be calculated with a calculator, or the calculations can be performed with JMP INTRO. To use JMP INTRO to make predictions from the original line, choose the **Big Class – Fit Y By X** window. Using the red triangle menu for **Linear Fit**, **Save Predicteds**. This creates a new column in your **Big Class** data table containing the predicted heights or weights for all students. Not only is the data stored, but the formula creating the data is also stored. To find the predicted values for each of these four students, on the menu bar, choose **Rows ® Add Rows** to add four rows. Enter the value of the expected value for each student in one of the new rows, one student per row, and the predicted value will appear in the **Predicted** column.

Using the pair of lines you drew in the second part of the lab, make a second prediction for each of the students, again using your explanatory variable to predict the response variable.

Similarly, these predictions can be calculated with a calculator, or the calculations can be performed with JMP INTRO. If you use JMP INTRO, you will have two columns of predicted values, one for each line. Be sure to use the correct predicted column, according to the sex of the student.

Which predictions were generally more accurate?

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

To summarize, your report should include the following components:

- Your reasoning behind your choice of explanatory and response variables;
- Your prediction for the scatterplot pattern(s);

- The ungrouped scatterplot with least-squares regression line and the equation of the least-squares regression line;
- Analysis of how the scatterplot relates to your expectations;
- An interpretation of the slope in the context of this problem and a determination of the meaning (or nonmeaning) of the y-intercept in the context of this problem;
- The scatterplot 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 scatterplot;
- 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.

JMP INTRO® Lab Activities
Teacher Notes

Lab Activity – Least-Squares Regression

Objectives:

- Gain familiarity with some of the basic operational tools of JMP INTRO.
- Use JMP INTRO to display a scatterplot, least-squares regression line, and a scatterplot with least-squares regression line for grouped data.
- Create a word processing document, incorporating graphs from JMP INTRO.

Time Required: 45 minutes

Materials:

- Least-Squares Regression student activity directions
- **Big Class** data set

Prerequisites:

- Students should have basic knowledge of how to use JMP INTRO.
- Students should have experience producing scatterplots by hand and/or with graphing calculators.
- Students should have experience calculating least-squares regression lines by hand and/or with graphing calculators.
- Students should have experience analyzing scatterplots and least-squares regression lines for both grouped and ungrouped data.

JMP INTRO® Notes:

- The **Big Class** data file is included in the Sample Data folder as part of JMP INTRO.
- JMP INTRO includes an extensive online help system. It contains a table of contents and/or can be used to search for a specific topic. In addition, JMP INTRO has context sensitive help. You can access it by selecting the help tool (?) from the tools toolbar and clicking inside a data table or report. JMP INTRO opens help specific to the clicked-on item.
- To calculate the least-squares regression line, students are asked to click the red triangle next to the heading **Bivariate Fit of *response variable* By *explanatory***

variable. Note that the actual variable names that will appear in the JMP INTRO window have not been written here. This was so that students would be forced to choose an explanatory and a response variable.

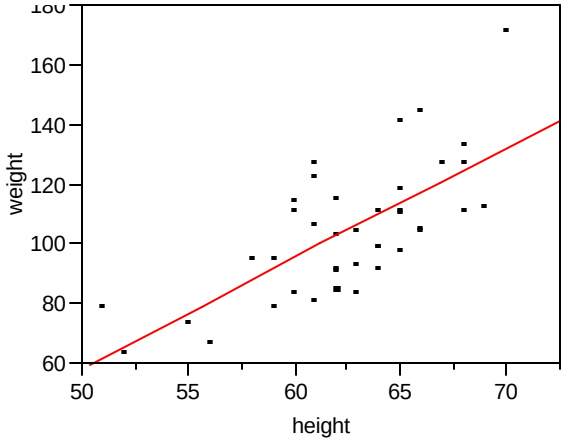
- The **Analysis of Variance** computations would be useful for concepts generally developed late, if at all, in an introductory statistics course.
- As students work through the activity, they need to follow directions carefully. Anticipating steps in the activity will generally result in incorrect displays.
- Students who choose **weight** as the explanatory variable will get the following regression equation for the entire data set: height = 48.321692 + 0.1355077 weight. For the grouped data, students who choose **weight** as the explanatory variable will get the following regression equation for females: height = 49.440931 + 0.1134085 weight and the following regression equation for males: height = 48.5498 + 0.1417979 weight.

JMP INTRO® Lab Activities
Answer Key

The following are suggested answers for Least-Squares Regression.

Height is the explanatory variable. A student's height tends to have an effect upon their weight. A positive association should be seen on the scatterplot, since one would expect that as a student's height increases, their weight also increases.

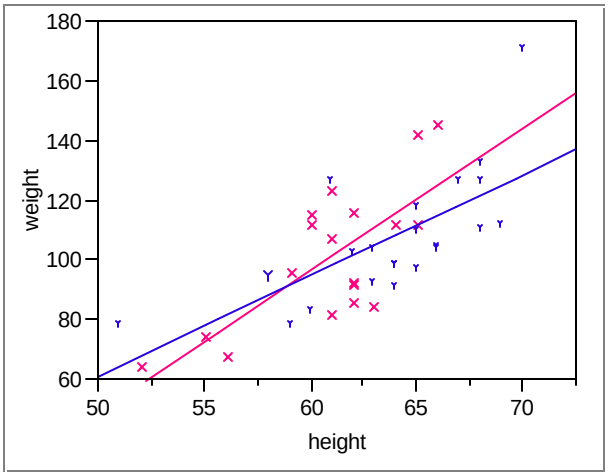
Scatterplot of ungrouped data with least squares regression line and equation of the least square regression line:



weight = -127.1452 + 3.7113549 height

The scatterplot shows a positive association between height and weight.
For every one inch increase in height, the weight of the student tends to increase 3.71 pounds. In this particular problem, an interpretation of the y-intercept does not make sense for several reasons. First, a height of zero inches in children is senseless, and even if it were possible to have a human with a height of zero inches, a weight of –127.15 pounds is impossible.

Scatterplots of grouped data with least squares regression lines and equations of the least squares regression lines:



Linear Fit sex=="F"
weight = -189.7725 + 4.7745491 height
Linear Fit sex=="M"
weight = -109.0145 + 3.400653 height

For females, every one inch increase in height results in an approximate 4.77 pound increase in weight. For males, every one inch increase in height results in an approximate 3.4 pound increase in weight.

In general, the female students seem to be shorter than the male students. For the shorter heights, males tended to weigh more than females; whereas, for the taller heights, females tended to weigh more than males.

Student Predictions:

Student Name	Height	Actual Weight	Predicted Weight Original Line	Predicted Weight Grouped by Sex
Shirley	67	122	121.516	130.122
Roberto	70	130	132.650	129.031
Charmaine	60	105	95.536	96.700
Huy	66	112	117.804	115.429

For three of the four students, our predictions were slightly more accurate when using the least squares regression equation created by sex of student; however, for Shirley, the prediction from the original least squares regression line was considerably more accurate.

For every inch increase in height, the female students tend to put on more weight.