

### Computer Project 1: Linear Regression

**Goals:** The first goal of this project is to introduce students to modeling data using linear regression. Students will use Excel to fit a linear model to data, and use the model to answer questions about the real world situation. Students will discuss how well and to what extent the model describes the situation. The second goal of this project is to give students hands-on experience using a spreadsheet (Microsoft Excel). Students will plot data, fit a trendline to the data, and use the “fill” tool to do a calculation repeatedly for a list of data.

I. The length of the femur versus the height of the individual.

- Launch Microsoft Excel.
- The boxes are called Cells. Cells are named by their column and row. For example the top left cell is A1.

This is Problem 1, page 207 (Femur Length and Height). Anthropologists use a linear model that relates the length of a human’s femur to his/her height. This model allows anthropologists to estimate the height of an individual when only a partial skeleton (including the femur) is present. In this problem we find the model by analyzing the data on femur length and height for the eight males given in the table.

Femur length (cm)	Height (cm)
50.1	178.5
48.3	173.6
45.2	164.8
44.7	163.7
44.5	168.3
42.7	165.0
39.5	155.4
38.0	155.8

- Click on the cell A1. A conspicuous border appears around the active cell.
- Type “Femur length (cm),” and push enter. The active cell becomes A2.
- Type 50.1 into cell A2, and push enter. Continue in the fashion to enter in the femur length data. Then click on cell B1, and type in the height data, beginning with its title.
- Click on the File in the menu bar at the top of the screen. Click on the Save As option. Save the sheet by giving it a name.
- To the data (cells A2 to B9), click on the cell A2, and holding the button down, move the cursor to cell B9, releasing the button.
- Click on Insert in the menu bar at the top of the screen. Click Chart... A dialog box called “Chart Wizard – Step 1 of 4 – Chart Type” will appear. Click on x-y scatter plot. Click Next.
- The title of the dialog box changes to Step 2 of 4 – Chart Source Data. We wont use these options at this time. Click Next.
- Step 3 of 4 – Chart Options. Select the Titles tab, if it is not already selected. Type a title for your chart. For the Value (X) axis, type Femur Length (cm). For the Value

(Y) axis, type Height (cm). Click the tab at the top of the dialog box that says Gridlines. Check Major gridlines under Value (X) axis and Value (Y) axis. Click Next.

- Step 4 of 4 – Chart Location. This dialog box allows you to choose whether the chart will appear in a new window, or as a part of the worksheet containing the data. Mark the “As an object in” radio button, if it is not already marked. Click Finish.
  - The relationship between the x and y values on this graph appear linear. You will now add a trendline, which is a line that is as close as possible to all of the data. Click on the chart to make it active (black squares will appear around its perimeter). From the menu bar, click the Chart menu and then select Add Trendline... A dialog box will appear.
  - The Linear Trend/Regression type is marked. Click on the Options tab. Click in the box beside Display equation on chart. Click OK.
  - A line of best fit should now appear, along with the equation of the line. You can click and drag the equation of the line to a convenient place on the chart.
  - From the menu bar, choose File and Save.
  - Click in cell A1. From the menu bar, choose File and then “Print preview.” You will see an image showing how the page will print. Click close. You will see dotted lines showing the page breaks in your document. Drag and drop the chart so that it fits entirely in the first page. You may need to adjust the chart’s size. Do this by clicking on the chart to activate it, and then clicking and dragging one of the corners of the chart, to adjust its size. Click in cell A1, and do Print preview again. If all is well, click print. From the menu bar, choose File and Save.
  - Click on the chart to activate it. Do Print preview. If you wanted to print only the chart, and not the numerical data, you could do so in this way. Click Close.
1. Write the formula for your trendline as a function  $f(x)$  that yields the approximate height of an individual with femur length  $x$ . In your homework, write a sentence giving your formula and saying what the function yields.

If the formula were perfect, then when you plugged each femur length into the function, you would get the precise height of the person.

2. Calculate  $f(44.7)$ . Write your answer in a sentence that includes what  $f(44.7)$  represents verbally. Compare your answer to what the data says the height of the man with a femur length of 44.7 cm really is.

If the formula were perfect, the line would pass directly through all of the data points. It is impossible for the trendline to pass through all of the points, since they are not precisely in a line. Instead, the trendline is chosen to be as close as possible to all points simultaneously. Even though, when you plug in a femur length, you do not get the precise height of the individual, you do get an approximation.

3. Calculate  $f(58)$ . What does  $f(58)$  represent verbally? Make sure your verbal description includes the word approximate.

4. Anthropologists find a femur that is 42 cms long. Use your function to approximate the individual's height.

II. Life Expectancy. This is problem 9 on page 208. The average life expectancy in the United States has been rising steadily over the past few decades, as shown in the table.

Year	Life Expectancy
1920	54.1
1930	59.7
1940	62.9
1950	68.2
1960	69.7
1970	70.8
1980	73.7
1990	75.4
2000	76.9

- Launch Microsoft Excel. If Excel is already open, from the File menu, select File and then New..., and select "Blank Workbook."
  - Enter the table above, including the column titles, into your worksheet. You will need to widen the second column to make space for the words Life Expectancy. Put your mouse over the line in the grey area between the letters B and C in the column headings. The cursor should change shape. Click and drag to widen the column.
  - Save your work so far.
  - Insert a scatter plot of this data and fit a trendline, as you did for Problem I.
  - As before, your line will not pass precisely through your data points. You will use the spread sheet to calculate the values that the model gives for each of the years listed above. Save your work so far.
1. Write the formula for your trendline as a function  $f(x)$ . Write your answer in a sentence that explains what the function yields and what  $x$  stands for. For example, write, "The function  $f(x)$ =your formula yields the ....., where  $x$  is..." Remember that your model only approximates life expectancy, so the words "approximate" or "estimated" or something similar should appear in your answer.
  2. Calculate  $f(2010)$ . What does  $f(2010)$  represent verbally?
  3. What is the slope of your line? The slope of a line tells us how much the  $y$ -value increases or decreases if the  $x$ -value is increased by 1. Write a sentence explaining what the slope of this model represents verbally, by rewording this sentence by replacing  $x$ -value and  $y$ -value by year and approximate life expectancy respectively.
- We will use the spread sheet to calculate the values that the model produces at each year in our data. Enter the word "Model" in the cell C1.
  - Above you identified the slope and intercept of your line. In cell C2 type  
    =slope\*  
and then click on cell A2 (do not hit enter yet). It should say (in the white strip below the tool bar)

=slope\*A2

then type + and your intercept. In this case your intercept is negative, so you'll be adding a negative number (like  $7+(-3)$ ); you need only type the minus sign (like  $7-3$ ). It should say

=slope\*A2+intercept

and then press enter. (I wrote +intercept above because the intercept is negative (the minus sign is part of the intercept). You should see the numerical value that the model yields for the year in cell A2.

- Click on cell C2 to activate it. Put your cursor on the bottom right hand side of the cell. Your cursor should change into a plus shape. Hold the mouse so that the cursor shows a plus shape, press the left mouse button, and holding it down, drag down to cell C10. Excel should put the formula into each of the cells you selected. This is called "filling down." You will do this again with another formula in a moment.
- Click on cell C3 and look at the formula in the white strip below the tool bars. Notice that Excel change the A2 that you had in cell C2, so that you have A3 in cell C3, A4 in cell C4, etc... One of the main reasons to use a spread sheet is if you want to apply the same formula to a variety of data. I use Excel to calculate my grades, when I apply the same formula to all of my students.

When we construct a model, we hope that it will give a reasonably accurate estimate of our situation. Since we construct the model from data, it may only be reasonable to use the model when then number we want to input (in this case, the year) is near the data on which we based the model.

- Above, you calculated the predicted life expectancy in the year 2010. Now you will do this every 10 years starting at 2010. Click in cell A11, and type =, then click cell A10, then type +10. Hit enter. 2010 should appear in cell A11.
- Do the "Fill down" operation described above to apply this formula to cells A12 through A30. Click on A14. It should say =A13+10. We have just made a list where each entry is obtained by adding 10 to the previous entry.
- Now click on cell C10. Fill down to C30. The value that appears in C11 should be the answer you got for  $f(2010)$ .

4. What does the model tell us the estimated life expectancy will be in the year 2100? Does this seem reasonable to you?

There is (obviously) no way to know what the life expectancy will be in the year 2100. Since the data looks like it lies on a line, it is reasonable to use a linear model to predict future life expectancies. It is also reasonable, and to be correct, necessary, to question the validity of the prediction when applying it to years far beyond our data set.

5. Write a sentence or two explaining what the numbers in cells C11 through C20 represent verbally and how they were obtained algebraically. For the verbal part, use your answer to number 2 above as a model. For the algebraic part, explain what you would have had to do if you were calculating these numbers by hand.

- Type “No data” in cell B11 and press enter. Click on B11 again, and fill down to B30.
- Save your work.
- Click in cell A1. Go to the file menu and select Print preview. At the top of the page, click Setup.... In the dialog box that appears, click Landscape. Click OK. Click close. You will see dotted lines showing you where the page breaks are. Adjust the size of the chart so that everything fits on one page. Click in cell A1, and do Print preview again. If all is well, click print.
- Save your work (so that your print settings and page layout will be saved).

III. Olympic Swimming records. Problem 11 on page 210 of the College Algebra text includes tables that show the gold medal times in the women’s and men’s 100-m freestyle Olympic swimming events.

- Type titles for the columns in the first row. Type the years in column A, the Men’s times in column B and the women’s times in column C. (For the tie in the men’s in 1984, only type the time once.)
  - Select all of the data (cells A2 through C22). From the menu bar, select Insert and then Chart....
  - Step 1 of 4 in the Chart Wizard dialog box that appears. Click on x-y scatter plot.
  - Step 2 of 4. Click on the tab “Series.” While Series 1 is highlighted, type, “Men’s” into the box labeled Name. Then click on Series 2 to highlight it, and type, “Women’s” into the box labeled Name.
  - Steps 3 and 4 of 4. Do the same as you did in part I to finish creating the graph.
  - From the tool bar, click Chart and then Add trendline.... In the Type tab of the Add trendline dialog box, you have the option of selecting the Men’s or Women’s data. First select Men’s. Then click on the Options tab, and check the box in front of “Display equation in chart.” Click OK.
  - Repeat the above for the Women’s data.
  - Print your work (formatted nicely) and include it with your homework.
1. Write the functions  $m(x)$  and  $w(x)$  that yield the approximate gold medal times in the 100-m freestyle Olympic swimming event in year  $x$  for men and women respectively.
  2. At this time the men are faster. Explain how you can tell that this is true by looking at the graph. Find the year in which the model predicts that the women and men will take the same time to swim the 100-m race. (How? Find the point at which the lines intersect using the formulas.)
  3. Calculate  $m(2100)$ ? What does  $m(2100)$  represent verbally? Does it make sense to apply the model to predict the Olympic gold medal times in the year 2100? Write a sentence explaining why or why not.

IV. Demand for candy bars. Do problem 14 on page 211. When you make your scatter plot, put the prices in the first column and the number of people who would buy the candy bar for each price in the second column. Print your graph and include it with your homework. Answer the questions in parts (c) and (d) using sentences.