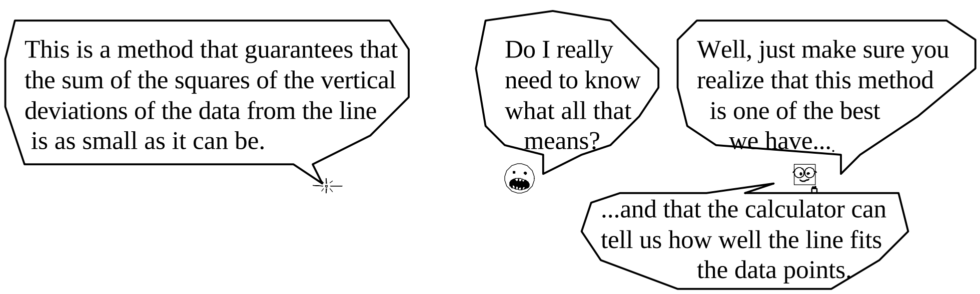
The Least Squares Line

[Instructor's note: This lesson requires that each student use a calculator or computer with linear regression capability.]

This lesson contains information that will help you to produce a linear function that describes how one measurement changes as a related measurement changes. The procedure is called the method of Least Squares. There are many situations where it can be used. Here are some examples from the publication *Critical Indicators Measuring Spartanburg County*:

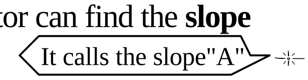
* The percent of Upstate high school graduates going to college increased at a near constant rate from 1983 through 1990. If that trend continues, what percent can we expect next year? in five years? (This question is important to college administrators for projecting enrollment.)
* From 1987 through 1990, aggravated assault cases in Spartanburg County rose at a near constant rate from 55 to 120 per 10,000 residents. If this condition is not changed, how many assaults per 10,000 are expected each year for the next five years. (This question is important to county administrators for anticipating staff needs in law enforcement.)
* The number of registered and licensed nurses per 100,000 residents in the Upstate increased at a near constant rate from about 475 in 1980 to about 700 in 1990. If you are a student nurse, how many do you project will be registered/licensed when you graduate?



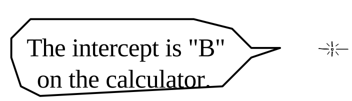
The procedure described here is for use with TI-82/83 calculators. [If you are using another calculator or computer, read your instructions on how to get the least squares line.]
The data given here represents accumulated profit (y in $100) per hour (x) at a concession stand.

x	**2**	**3**	**4**
y	**0**	**.5**	**1**

Here's how the calculator can find the **slope**



and the **y-intercept**.



Input the data into your TI-82/3. To do this, hit STAT ENTER and clear out columns L1 and L2. Put your x-values in L1 and your y-values in L2. To plot the three pairs of points, press 2ND STATPLOT ENTER and turn on the plot. Choose the scatter plot. Make sure xLIST is L1 and yLIST is L2. Next press WINDOW and change xmin to 0, xmax to 10 ymin to 0 and ymax to 5. Press graph and show your graph on the next page:

Now to get the linear regression equation expressing y in terms of x, press STAT, choose the CALC menu and select LinReg(ax+b). It will be number 5 on the TI-82 or number 4 on the TI-83. Once the command is on the home screen, press 2^{ND} 1 (L1) [,] 2^{ND} 2 (L2) to ensure that your calculator is looking at the correct lists. Your screen should look like this:

LinReg (ax+b) L1,L2

Then press ENTER. Record the slope, intercept and correlation coefficient here:

a =___________ b =___________ r =___________

(If your TI-83 does not give you the correlation coefficient (r) , press 2^{ND} 0 (CATALOG) and select DiagnosticOn. Repeat the calculation.)

Write the equation of the line here: y = ____

Now suppose I want to predict my profit after 4.5 hours. I can substitute 4.5 into the equation for x and compute the corresponding value for y. Calculating by hand is easy, but your calculator can also be used. One way is as follows:

Press Y= and clear any function you may find in Y1. Press VARS 5 ▸ ▸ to get to the EQuation menu. Select ReqEQ which is 7 on the TI-82 and 1 on the TI-83. The most recently computed regression equation should appear in the function list as Y1. It can now be graphed, a table of values can be computed, etc. Our purpose, though, is to evaluate the function at a specific value of x using the Y-VARS feature found above the VARS key on the TI-82 and as a menu under VARS on the TI-83. Starting from the home screen, select Y-VARS and press ENTER twice to copy Y1 to the screen. Press (4.5) so that the screen contains Y1(4.5)

Then press enter. Write your answer here: When x=4.5 hours, y= _____ hundred dollars

To estimate how long I have to stay in the stand in order to earn $1000 (corresponding to y=10), I can press 2^{ND} TBLSET and set TblMin (or TblStart) to 0 and ▲Tbl to 1. Then pressing 2^{ND} TABLE gives a list of ordered pairs I can scroll to find the number of hours (x) corresponding to y=10. Record your answer:

It takes about ________ hours to make $1000.

Finally, to graph the linear regression equation you stored in the function list, simply press GRAPH. The line should graph through the points graphed earlier. Add it to your graph above.

The points in this example fit exactly on the line. So if you put, say, 3 in for x, you'll get exactly 0.5 , as in the table of data. If the points don't fit a line exactly, the same calculator entry procedure is used, but the line you get cannot fit all the points exactly. But it will be the "best" fit in the Least Squares sense.

<u>MicrowaveActivity:</u> The time that it takes a glass of water to boil in a microwave oven is a near-linear function of the amount of water in the glass. This data shows 10 tests of various amounts of water. Use your calculator to find the linear function that predicts boiling time (Y) from amount of water (X).

	(X) Ounces of water	(Y) Boiling time (sec.)	
1	1	42	
2	2	55	
3	3	59	
4	4	85	Slope:________
5	4	95	
6	6	120	
7	6	130	Intercept:________
8	6.5	152	
9	7	152	
10	8	158	Equation of the Line:______________________

Interpret the meaning of the slope (in English and in context): ______________________________

____

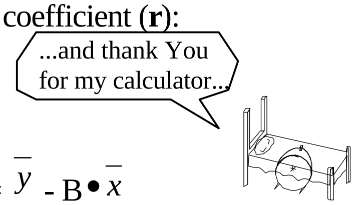
Interpret the meaning of the intercept. Why do you suppose it isn't zero? ____________________

____

If you place five ounce of water in this microwave, how long will it take to boil? _____________

If a glass of water took 200 seconds to boil, how much water was in the glass? ______________

If we don't have a statistical calculator we can use the following formulas for the slope (B), the intercept (A) and the correlation coefficient (**r**):



$$B = \frac{\sum[(x-\bar{x})(y-\bar{y})]}{\sum(x-\bar{x})^2}, \quad A = \bar{y} - B \bullet \bar{x}$$

and $$\mathbf{r} = \frac{\sum[(x-\bar{x})(y-\bar{y})]}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} .$$