

L1	L2	L3
180	24	
190	21	
200	16	
210	13	
220	11	
230	9	
-----	-----	
L2(9) =		

Now, to actually plot the data, press **2nd** **STAT PLOT** (actually the **Y=** key) and then **ENTER** to select the first item. Make sure that On is selected (it should be highlighted). For Type: you want to select the first (leftmost) option. Next you want set Xlist to L1. Do this by scrolling down to Xlist and pressing **2nd** **L1** (the **1** key). In the same manner, set Ylist to L2. Lastly, select the leftmost Mark.

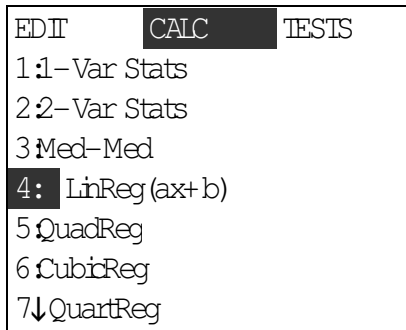
Press **Y=** to return to the function definition screen. On this screen you should see Plot1 highlighted. Turn off any functions that are entered here (move the cursor over the = and press **ENTER**) For example: $\text{Y1} = 2x+1$ is on while $\text{Y1} = 2x+1$ is off. Before actually graphing the data, we should stop and think about a good window to view it in. If we just graph it in the standard window ($-10 < x < 10$, $-10 < y < 10$) we won't be able to see any of the points, so we need to pick a better window. I suggest trying the following window: (press **WINDOW**)

WINDOW
Xm in= 150
Xm ax= 240
Xsc= 10
Ym in= 0
Ym ax= 35
Ysc= 5
Xres= 1

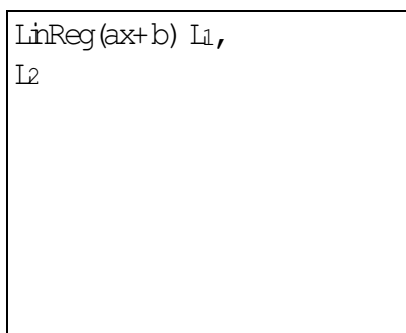
Think about why this is a good window for this example. Finally, press **GRAPH**. You should now see a scatter plot similar to the one found on page 42 of the textbook.

So now you want a regression line:

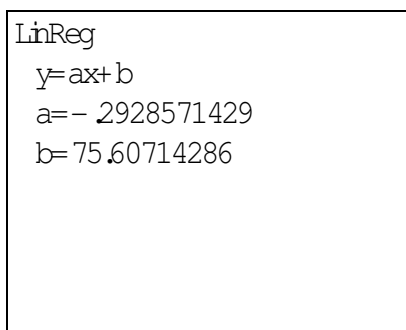
This is actually the easy part. Press **STAT** and arrow right to the CALC menu, then scroll down to LinReg(ax+b) :



Press **ENTER**. You should have a clear screen that says only $\text{LinReg}(ax+b)$. Since our data is in lists 1 and 2, press **2nd** **L1** **,** **2nd** **L2** :



Press **ENTER**, and we have our linear regression line:



Keep in mind that a is the slope (what we have called m) and b is the vertical-intercept, which thankfully, we have also called b . So, writing this line in the form, $y = mx + b$, we get that the equation of the regression line is $v = 75.607 - 0.293T$.

What about the correlation coefficient?

First, we need to know what a correlation coefficient is. When a calculator or computer computes a regression line, it also gives the *correlation coefficient*. This is a number that is always between -1 and 1. It measures the degree to which the data is linear and whether the data is increasing or decreasing. The closer to 1 or -1 the correlation coefficient is, the more linear the data. If the correlation coefficient is negative, the data

is decreasing and if the coefficient is positive, the data is increasing. If the correlation coefficient is close to zero, the data is either completely scattered or there may be a relationship between the data that is not linear. It is possible that the relationship between the variables is sinusoidal (looks like a sine or cosine function). If this were true, the correlation coefficient would be close to zero telling us that the relationship between the variables is not linear not that there is no relationship between the variables.

Also it is important to understand the difference between relationship, correlation and causation. As mentioned above, there may be a relationship between two variables that is sinusoidal but the correlation is 0. Remember, all the correlation coefficient does is measure how linear the relation is between the variables. It is also true that a high correlation does not necessarily imply causation. The classic example (taken from *Statistics*, 2ed, by David Freeman, et al) is that there is a high correlation between a child's reading level and shoe size. However, this does not mean that a child with unusually large feet would be a better reader. Obviously, our reading skills and our foot size increase as we grow older.

So how do we find this correlation coefficient? This is actually a very simple issue to address. All we need to do to get the correlation coefficient is turn on the correct mode. Press **2nd** **CATALOG** (**0**), **D** (**x^{-1}**) and scroll down to **DiagnosticOn** , press **ENTER** twice, and repeat the linear regression process above. This time you should get:

```

LinReg
y= ax+ b
a= - .2928571429
b= 75.60714286
r2= .9841920375
r= - .9920645329

```

Note that they gave us an **r** and an **r²** value. Both are useful in practice, but for now we are only concerned with the **r** value.

Assignment:

You are given the following data which gives a student's IQ as a function of the number of hours of television that person watches each week:

No. of hours of TV viewed	IQ of the individual
10	110
12	105
20	105
5	125

30	95
17	102
2	140
8	130
12	110
15	100
9	120
0	145
1	150
15	110
6	120

1. Make a scatter plot of the data
2. Use your calculator to find the least squares regression line (linear regression) and the correlation coefficient.
3. Give interpretations of the slope and vertical intercept on the line.
4. Explain what the correlation coefficient tells you about the relationship between the number of hours a person views in a week and that person's IQ.
5. Are there any outliers in this data? An outlier is a data point that does not fit well with the other data points. Explain any outliers.