

Getting your TI-83 to do your Statistics IV

Muhammad Zafrullah

6. Scatter plots and regression line

When data comes to us in the form of ordered pairs (x, y) , where x and y are variables (qualitative or quantitative) we would like to know what kind of information (with reference to the data) can be gleaned from one variable about the other variable. When both x and y are quantitative we can use TI 83 for reasonably sized sets of data.

As you know a scatter plot of two variable data can be of tremendous help in deciding if the association is linear, weak or strong, positive or negative. For this let us take up a set of two variable data and do the scatter plot for it. The data comes from Exercise 2.10 of your book. To fix ideas the X variable will represent the explanatory variable and the Y variable the response variable. Here the X variable is the speed of a car in km/hour and the Y variable is the consumption of fuel in liters/100km

Data:

X	Y	X	Y
10	21	90	7.57
20	13	100	8.27
30	10	110	9.03
40	8	120	9.87
50	7	130	10.79
60	5.9	140	11.77
70	6.3	150	12.83
80	6.95		

Turn your calculator on, get to the Edit screen and clear at least two lines, say L1 and L2. Punch the X-values in L1 and Y-values in L2. Make sure that both L1 and L2 are of the

L1	L2	L3	Z
100	8.27		
110	9.03		
120	9.87		
130	10.79		
140	11.77		
150	12.83		
-----	-----		
L2(16) =			

same length. After punching in the data they should look like:

Clear your “Y=” screen. Get to STAT PLOT screen ([2nd], [Y=]). Turn the plot that suits you best on, but do not forget to turn the other ones off. I would stick to **Plot 1**, which is on, on my calculator. Take the cursor to “Type”. The cursor is blinking on the first picture in front of “Type”. Fix it there by hitting ENTER. The Plot 1 screen would now look like:

Change the marker to something visible, if it is not already quite visible. I would change it to “+”:

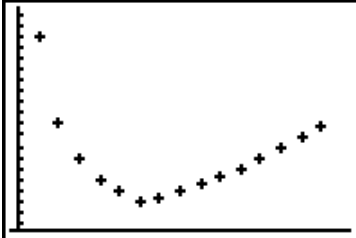
Plot1	Plot2	Plot3
On	Off	Off
Type:		
Xlist: L1		
Ylist: L2		
Mark:		

```

21041 Plot2 Plot3
Off Off
Type: [ ] [ ] [ ]
      [ ] [ ] [ ]
Xlist:L1
Ylist:L2
Mark: [ ] [ ] [ ]

```

Now do ZOOM, [9] to see:



This is the scatter plot of the data. You see that there is quite a strong association between X and Y variables but it is not even close to a linear association. (If you are economy oriented you would say that it would be cheapest to drive at 80 km/hour, because the fuel consumption rates are higher if the speed is significantly different from 80km/hr).

For our next example I have chosen an old favorite of mine and to do with cars. This data is taken from N. Weiss' Elementary Statistics". The idea to explore here is that the price of a used car depends on its age. Eleven cars are picked at random to get an idea of the relationship between the age of the car and its price.

Age (yrs)	Price
5	8500
4	10300
6	7000
5	8200
5	8900
5	9800
6	6600
6	9500
2	16900
7	7000
7	4800

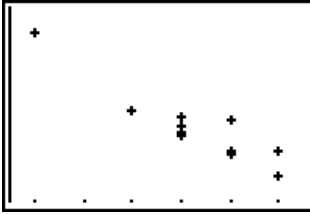
Let us make a scatter plot. I would prefer to stick to Plot 1 and so I will punch in the data in L1 and L2; the explanatory variable in L1 and the response variable in L2. (I trust that you remember that you should first delete the existing data in L1 and L2 before punching in this data.) My Plot 1 screen looks like:

```

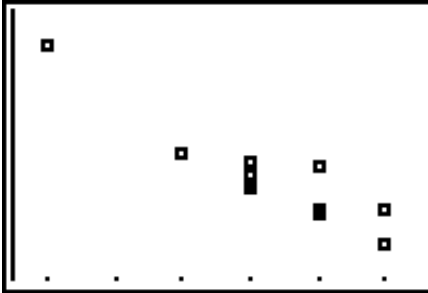
21041 Plot2 Plot3
Off Off
Type: [ ] [ ] [ ]
      [ ] [ ] [ ]
Xlist:L1
Ylist:L2
Mark: [ ] [ ] [ ]

```

Now do ZOOM, [9] to see:



The marker looks stupid, so let us change the marker to a square brick. So that we can see what is going on. On changing the marker the picture is:



The points in the scatter plot seem to be bunched about a line with a negative slope. So there is a strong negative linear association.

The above statement raises several questions: Where is that line these points are bunched about? What is that line called? Is there a number that determines the strength of relationship? Your instructor will have more satisfying answers. For now that line is called the line of regression, there is a number r , between -1 and 1 (inclusive) that determines the strength of relationship. This number r is called correlation. If r is closer to -1 or 1 the linear association is strong and the calculator can only calculate and draw that line, that we will now.

This is what we do: Keeping the data in L1 and L2 and the Stat Plot and Y= screen intact press STAT and take the cursor to CALC by pressing the right arrow once to see:

```

EDIT [CALC] TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7↓QuartReg
  
```

Move the cursor down to "LinReg(ax+b)":

```

EDIT [CALC] TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7↓QuartReg
  
```

Press ENTER to see:

Punch in L1, L2 (or the lines in which your data is):

```

b=-3.659586682
r2=.9883313819
r=.9941485714

SortA(L1
                               Done
LinReg(ax+b) L1,
L2

```

Once you press ENTER your calculator screen would show one of the following two screens:

I	II
<pre> LinReg y=ax+b a=-2026.126126 b=19546.84685 </pre>	<pre> LinReg y=ax+b a=-2026.126126 b=19546.84685 r²=.8533733464 r=-.9237820881 </pre>

If your calculator shows the second screen (II) you are good to go skip the following procedure. (If your calculator shows the first screen **you need turn diagnostic on** For that you do the following:

Press [2nd], [0]. This takes you into the catalogue screen:

```

CATALOG A
▸abs(
and
angle(
ANOVA(
Ans
argument(
AxesOff

```

(The A in the right hand corner tells you that your calculator is in ALPHA mode.) Press D (in green) to see:

```

CATALOG A
▸dbd(
▸Dec
Degree
DelVar
DependAsk
DependAuto
det(

```

Press the down arrow key until you get to diagnostic on. This is how it looks:

```

CATALOG A
Degree
DelVar
DependAsk
DependAuto
det(
DiagnosticOff
▸DiagnosticOn

```

Press ENTER twice:

```

a=-2026.126126
b=19546.84685

DiagnosticOn
                               Done

```

Now do the STAT, CALC, “LinReg” routine to get the second screen.):

```
LinReg
y=ax+b
a=-2026.126126
b=19546.84685
r2=.8533733464
r=-.9237820881
```

This tells you that the equation of the regression line is approximately:
 $Y = -2026X + 19546.8$ and the correlation is approximately $r = -.9238$. There is another quantity there called the **coefficient of determination** and it is equal to the square of correlation.

All that remains is to draw the line of regression. For that you proceed as follows:
 Press “[Y=]”. If there is any justice in the world this screen should be clear. Now press the [VARS] key. Choose “5:Statistics” to see:

```
Σ EQ TEST PTS
1:
2:
3:
4:
5:
6:
7:↓
```

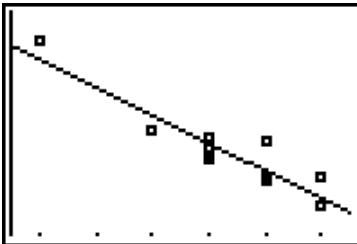
Press the right arrow twice to get to:

```
XV Σ EQ TEST PTS
1: RegEQ
2:
3:
4:
5:
6:
7:↓
```

Press ENTER to see the “[Y=]” screen with an equation from hell:

```
Plot1 Plot2 Plot3
\Y1=2026.126126
1262X+19546.8468
46847
\Y2=
\Y3=
\Y4=
\Y5=
```

Do not be frightened. This is the equation whose approximation you have seen above.
 Now all you have to do is Press ZOOM and [9]:



Voila! Here’s the line.