

Scatter Plot, Correlation, and Regression on the TI-83/84

Summary: When you have a set of (x,y) data points and want to find the best equation to describe them, you are **performing a regression**. This page shows you how to determine the **strength of the association** between your two variables (correlation coefficient), and how to find the line of best fit (least squares regression line).

For an illustration of linear regression, we'll use the data given below in a table. The explanatory variable x is dial settings on a freezer, and the response variable y is temperature of the freezer.

Contents: [Step 0. Setup](#)
[Step 1. Make the Scatter Plot](#)
[Step 2. Perform the Regression](#)
[Step 3. Display the Regression Line](#)
[Step 4 \(optional\). Display the Residuals](#)

Step 0. Setup

Set floating point mode, if you haven't already.

[MODE] [▼] [ENTER]

Go to the home screen

[2nd MODE *makes* QUIT] [CLEAR]

Turn on diagnostics with the [DiagnosticOn] command.

[2nd 0 *makes* CATALOG] [x^{-1}]

Don't press the [ALPHA] key, because the CATALOG command has already put the calculator in alpha mode.

Scroll down to DiagnosticOn and press [ENTER] twice.

The calculator will remember these settings when you turn it off: next time you can start with Step 1.

Step 1. Make the Scatter Plot

Before you even run a regression, you should **first plot the points and see** whether they seem to lie along a **straight line**. If the distribution is obviously not a straight line, don't do a linear regression. (Some other form of regression might still be appropriate, but that is outside the scope of this course.)

Turn off other plots.

[Y=]

Cursor to each highlighted = sign or Plot number and press [ENTER] to deactivate.

Enter the numbers.

[STAT] [1] selects the list-edit screen.

Dial (x)	0	2	3	5	6
Temp, °F (y)	6	-1	-3	-10	-16

Cursor onto the label L1 at top of first column, then [CLEAR] [ENTER] erases the list. Enter the x values.

Cursor onto the label L2 at top of second column, then [CLEAR] [ENTER] erases the list. Enter the y values.

Set up the scatter plot.

[2nd Y= makes STAT PLOT] [1] [ENTER] turns Plot 1 on.

[▼] [ENTER] selects scatter plot.

[▼] [2nd 1 makes L1] ties list 1 to the x axis.

[▼] [2nd 2 makes L2] ties list 2 to the y axis.

Plot the points.

[ZOOM] [9] automatically adjusts the window frame to fit the data, but does not adjust the grid spacing.

(optional) [WINDOW], set Xscl=1 and Yscl=5, then [GRAPH] to redisplay it. (Appropriate values of Xscl and Yscl may be different for other problems. Pick the values that make the graph look best to you.)

Step 2. Perform the Regression

Set up to calculate statistics.

[STAT] [▶] [4] pastes LinReg (ax+b) to the home screen.

[2nd 1 makes L1] [,] [2nd 2 makes L2] defines L1 as x values and L2 as y values.

Set up to store regression equation.

[,] [VARS] [▶] [1] [1] pastes Y1 into the LinReg command.

Make it so!

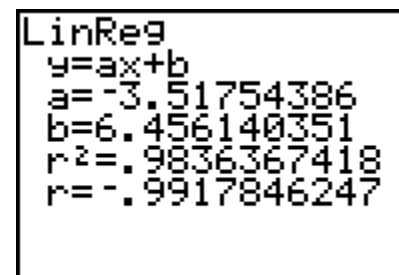
[ENTER] shows correlation and regression statistics *and* pastes the regression equation into Y1.

Write down a (slope), b (y intercept), r (correlation coefficient; r^* is our symbol). Round a and b to two more decimal places than your actual y values have; remember that final rounding should be done only at the end of calculations. Round r^* to two decimal places unless it's very close to ± 1 or to 0.

$$a = -3.52$$

$$b = 6.46$$

$$r^* = -0.992$$



```
LinReg
y=ax+b
a=-3.51754386
b=6.456140351
r^2=.9836367418
r=-.9917846247
```

R^2 is the **coefficient of determination**. The closer it is to 1, the better a predictor is the regression equation.

Another way to look at it is that in this case R^2 is about 98%, so 98% of the variation in y is associated with the variation in x.

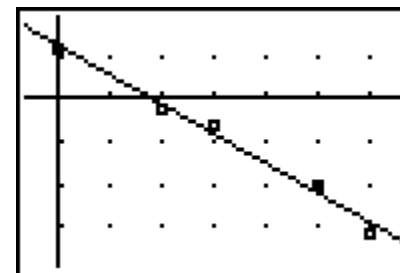
Statisticians say that R^2 tells you how much of the variation in y is “explained” by variation in x, but if you use that word remember that it means a numerical association, not necessarily a cause-and-effect explanation.

Only linear regression will have a correlation coefficient r, but any type of regression will have a coefficient of determination R^2 that tells you how well the regression equation predicts y from the independent variable(s). (The calculator uses r^2 , but most authors use R^2 .)

Step 3. Display the Regression Line

Show line with original data points.

[GRAPH]



Step 4 (optional). Display the Residuals

A plot of residuals can be helpful to show whether linear regression was the right choice.

If the residuals are more or less evenly distributed above and below the axis and show no particular trend, you were probably right to choose linear regression. But if there is a trend, you have probably forced a linear regression on non-linear data. If your data points looked like they fit a straight line but the residuals show a trend, it probably means that you took data along a small part of a curve.

The residuals are automatically calculated during the regression; all you have to do is plot them on the y axis against your existing x data.

Make the residuals visible in the statistics editor.

[STAT] [1] brings up the editor.

Cursor to the column heading of [L3] and press [2nd DEL makes INS] to open up a new list. You see the NAME= indicator at the bottom of the screen, with the blinking A to indicate alpha mode.

L2		L3	3
6		-----	
-1			
-3			
-10			
-16			

Name=A			

Press [2nd STAT makes LIST], then scroll to RESID and press [ENTER]. The list of residuals appears.

L2	RESID	L3	3
6	-.4561	-----	
-1	-.4211		
-3	1.0965		
-10	1.1316		
-16	-1.351		
-----	-----		
RESID = (-.4561403...			

Turn off other plots.

Press [Y=]. Cursor to the highlighted = sign next to Y1 and press [ENTER]. Cursor to PLOT1 and press [ENTER].

Set up the plot of residuals against the x data.

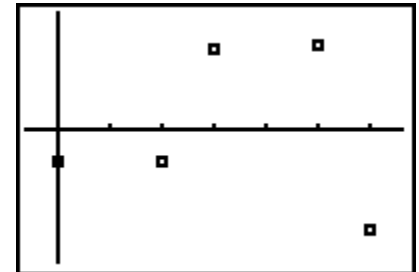
Set up Plot 2 for the residuals. Press [2nd Y= makes STAT PLOT] [▼] [ENTER] [ENTER] to turn on Plot 2. Press [▼] [ENTER] to select a scatter plot.

Plot1	Off	Plot3	
Type:	Scatter	Line	Box
Xlist:	L1		
Ylist:	RESID		
Mark:	+		

The x's are still in L1, so press [2nd 1 makes L1] [ENTER]. In this plot, the y's will be the residuals: press [2nd STAT makes LIST], cursor to RESID, and press [ENTER] [ENTER].

Display the plot.

[ZOOM] [9] displays the plot.



Don't worry about the magnitude of the residuals, because [ZOOM] [9] adjusts the vertical scale so that the points take up the full screen. What you want to look at is **whether there's a trend in the residuals**. Here there is no trend, so you conclude that a linear regression was the right choice, as opposed to regression against some curve.

(By the way, if you want to remove the residuals list from your statistics editor, just cursor to the column heading and press [DEL].)