

cor·re·la·tion

/ˌkôrəˈlāSH(ə)n/

noun

a mutual relationship or connection between two or more things.

"research showed a clear **correlation between** recession and levels of property crime"

synonyms: connection, association, link, tie-in, tie-up, relation, relationship, interrelationship, interdependence, interaction, interconnection; [More](#)

- STATISTICS
interdependence of variable quantities.
- STATISTICS
a quantity measuring the extent of interdependence of variable quantities.

We will study correlation by examining scatterplots.

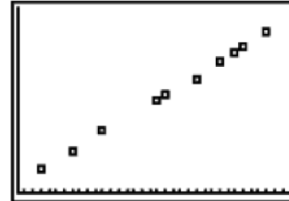
Media in the United States

Year	Daily Newspaper Circulation (millions)	Number of Households With Television (millions)
1950	54	4
1960	59	46
1970	62	59
1980	62	76
1990	62	92
2000*	55	101

Scatter Plots

A **scatter plot** is a graph used to determine whether there is a relationship between paired data.

In many real-life situations, scatter plots follow patterns that are approximately linear. If y tends to increase as x increases, then the paired data are said to be a **positive correlation**. If y tends to decrease as x increases, the paired data are said to be a **negative correlation**. If the points show no linear pattern, the paired data are said to have **relatively no correlation**.

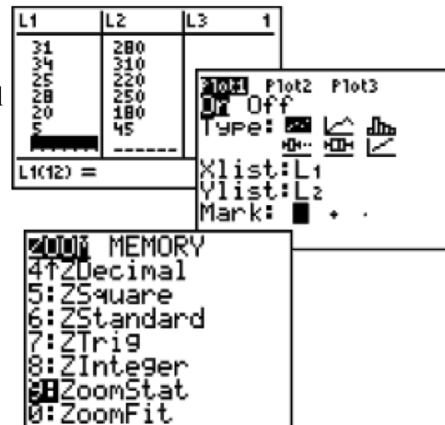


To set up a scatter plot:

Clear (or deactivate) any entries in Y= before you begin.

1. Enter the X data values in **L1**. Enter the Y data values in **L2**, being careful that each X data value and its matching Y data value are entered on the same horizontal line.

(See [Basic Commands](#) for entering data.)



2. Activate the scatter plot. Press **2nd STATPLOT** and choose **#1 PLOT 1**. You will see the screen at the right. Be sure the plot is **ON**, the scatter plot icon is highlighted, and that the list of the X data values are next to **Xlist**, and the list of the Y data values are next to **Ylist**. Choose any of the three marks.

3. To see the scatter plot, press **ZOOM** and **#9 ZoomStat**. Hitting **TRACE** and right arrow will move along the data points.

4. To turn the scatter plot off, when you are finished with this problem:

Method 1: Go to the **Y=** screen. Arrow up onto the **PLOT** highlighted at the top of the screen.

Press **ENTER** to turn it off.

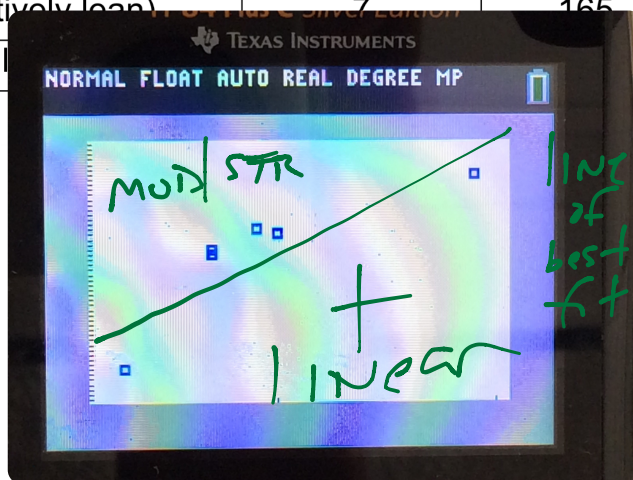
Method 2: Go to **STAT PLOT** (above **Y=**). Choose your **PLOT** location. Arrow to **OFF**.

Press **ENTER** to turn it off.

1. Use the given data to make a scatter plot.

Calories and Fat Per Portion of Meat & Fish

	Fat (grams)	Calories
Fish sticks (breaded)	3	50
Shrimp (fried)	9	190
Tuna (canned in oil)	7	170
Ground beef (broiled)	10	185
Roast beef (relatively lean)	7	165
Ham (light cure, 1		



To do regression

STAT

CALC

4: LinReg (ax + b)

Xlist:

Ylist:

FreqList:

Store RegEQ:

Calculate:

4: LinReg (ax + b)

Xlist: L1 (2nd 1 – L1)

Ylist: L2 (2nd 2 – L2)

FreqList: Keep blank

Store RegEQ: Y1 (ALPHA F4 Y1)

Calculate: Press Enter

Linear Regression

A linear regression is also known as the "line of best fit".

Side note: Although commonly used when dealing with "sets" of data, the linear regression can also be used to simply find the equation of the line between two points.

Example: Find the equation of the line passing through (-1, 1) and (-4, 7).

Entering the information as shown in the screens below, we arrive at the equation of the line:

L1	L2	L3	2
-1	1		
-4	7		
L2(3) =			

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

LinReg(ax+b) L1,
L2:

LinReg
y=ax+b
a=-2
b=1
r ² =1
r=-1

The equation is $y = -2x - 1$.

The correlation coefficient is -1 since both points are "on" the line and the line slopes negatively.

Linear Regression Model Example

Let's examine an example of the linear regression as it pertains to a "set" of data.

Data: Is there a relationship between Math SAT scores and the number of hours spent studying for the test? A study was conducted involving 20 students as they prepared for and took the Math section of the SAT Examination.



Hours Spent Studying	Math SAT Score
4	390
9	580
10	650
14	730
4	410
7	530
12	600
22	790
1	350
3	400
8	590
11	640
5	450
6	520
10	690
11	690
16	770
13	700
13	730
10	640

- Task:**
- Determine a linear regression model equation to represent this data.
 - Graph the new equation.
 - Decide whether the new equation is a "good fit" to represent this data.
 - Interpolate data: If a student studied for 15 hours, based upon this study, what would be the expected Math SAT score?
 - Interpolate data: If a student obtained a Math SAT score of 720, based upon this study, how many hours did the student most likely spend studying?
 - Extrapolate data: If a student spent 100 hours studying, what would be the expected Math SAT score?