

LINEAR REGRESSION

(Finding a best-fit line)

Q: According to the *Toys R Us 1995 annual report*, the number of stores between the years 1984 and 1994 is shown in the following table:

YEAR	1984	1986	1988	1990	1992	1994
# of STORES	213	338	522	712	918	1115

- Find a linear model that best fits the data.
- Describe what the slope of the line represents.
- Use the model to predict the number of stores in 1995.

This handout will cover how to align the data, create lists using the given data, view a scatter plot, perform a linear regression, store the regression equation, graph it, and use the equation to predict values on the TI-86 graphing calculator.

Aligning the data

Before typing any of the data into the calculator, the data must first be aligned. This is a process of making the data values smaller, and therefore easier for the calculator to manipulate. It will also give an equation with smaller coefficients.

When using data that has an input value that represents a year, ALWAYS align the data to represent the number of YEARS (not decades or centuries) after a given year. It is easiest to use a multiple of ten as the starting year. Align this data to be **the number of years after 1980**. The new values will be as follows:

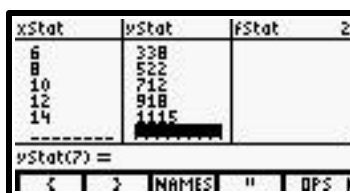
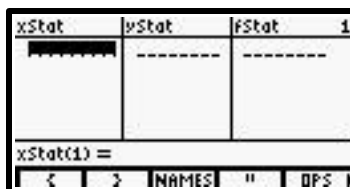
# of YEARS after 1980	4	6	8	10	12	14
# of STORES	213	338	522	712	918	1115

If the data had been aligned to be the number of years after 1900, the input values would have been 84, 86, 88, 90, 92, and 94. It is very important that a new definition be written for the input values, based on how the data was aligned. This will become more apparent when trying to interpret the slope of the model, or use the model to make a prediction.

Creating lists

The first step on the calculator is to put the data into the TI-86 as two lists:

1. Hit 2nd **STAT** and choose **EDIT**
2. If there is already data in xStat arrow up to the name 'xStat' and hit **CLEAR** then **ENTER**, to remove all previously stored data in that list.
3. Type all input values into xStat. Hit **ENTER** between each value.
4. Use the right arrow key to go over to yStat. Clear it, if necessary. Type all the output values into yStat, hitting **ENTER** between each value. (Make sure both lists are the same length.)



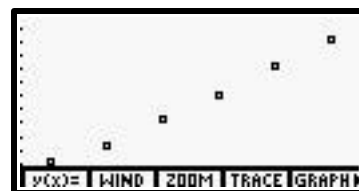
Creating a scatter plot

The next step is to create a graph of the data, or scatter plot. This step is important when trying to decide which type of graph (equation) will best fit the data.

1. Hit 2nd **STAT** and choose **PLOT**, then **PLOT1**.
2. Turn Plot1 on by hitting **ENTER** on the word 'On'.
3. The type should be a scatter plot. Make sure 'Xlist Name' is xStat and 'Ylist Name' is yStat.
4. Before looking at the scatter plot, make sure all graphs are turned off. Hit **GRAPH**, and choose "y(x)=" and either clear all equations out, or turn them "off" by arrowing onto the equal sign and hitting **ENTER**. Hit **EXIT** once.



5. To look at the data points in the best window possible, let the calculator choose the window by hitting **ZOOM**, **MORE**, then **ZDATA**.

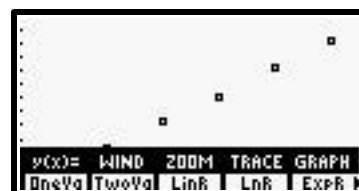


The data points appear to be fairly linear. (There is actually a slight curve to the scatter plot and a quadratic model will fit better.)

Performing a linear regression and storing the equation in y1

Next, the calculator must find the equation of the line that best fits the data, and store it into y1:

1. Hit 2nd **STAT**, choose **CALC**, Then **LinR**.
2. The statement 'LinR' should appear on the home screen, with a flashing cursor.
3. To store the regression equation into y1, hit 2nd **CATLG-VARS**, choose **REAL**, and arrow down to the lower-case "y" hit **ENTER** and then hit "1".
4. Hit **ENTER** to run the regression AND store it into y1 simultaneously.



```
LinReg
y=a+bx
a=-191.66667
b=92
corr=.997486777
n=6
```

If the data was stored in lists other than xStat and yStat, the names of the lists would have to be typed, along with the storage location y1, each separated by a comma:
LinR L1,L2,y1

- a) The linear model for the data is:

$$y = 92x - 191.66667 \text{ stores, } x \text{ years after 1980.}$$

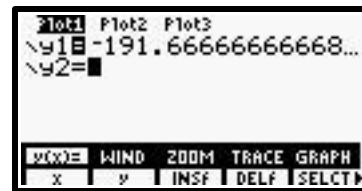
(It was rewritten to fit into the $y = mx + b$ form.)

- b) The slope of the model is 92. It is interpreted as the increase in the number of stores per year. (Slope represents a rate of change.) The fact that x is defined as the number of years after 1980 does not effect the interpretation of the slope.

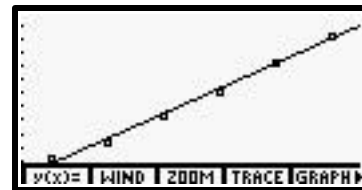
Viewing the regression equation in y1 and graphing

Graphing the regression equation on the same screen as the scatter plot will help decide if the graph is a good fit to the data. The calculator has already stored the regression equation into $y1$, the next step is to view the graph.

1. The equation has already been stored into $y1$. To view it, hit **GRAPH**, then " $y(x)=$ ". To view the rest of the equation scroll right with the arrow key.



2. To view the line, along with the scatter plot, hit **EXIT**, then choose **GRAPH**.



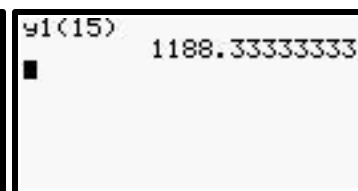
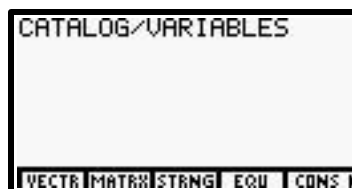
*It is now more apparent that the scatter plot has curvature.

Predicting values using a model

The model can be used to predict the number of stores in any year (close to the data values). The question was asked to predict the number of stores in 1995. Keeping in mind that the data was aligned to be the number of years after 1980, the desired input value (x value) to be used will be $x = 15$. If the data were aligned to be the number of years after 1900, the regression equation would have been different, and the desired input value would have been $x = 95$. Therefore, it is VERY important to write a clear definition for the alignment of the data before performing a regression.

There are several ways to use the calculator to find the output value, given an input value (from the first handout: *Evaluating Expressions*). The quickest method is to use function notation. The equation was already stored into y1, and so the calculator will need to find y1(15):

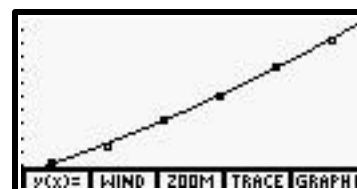
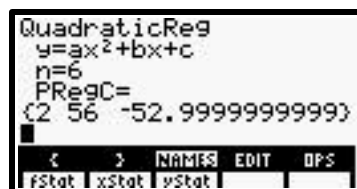
1. From the home screen (if you are not there, hit 2nd QUIT), get the variable name 'y1' by hitting 2nd CATLG-VARS, then hit MORE, and choose EQU.
2. Arrow down to 'y1', unless the regression equation was stored into a different location, and hit ENTER.
3. 'y1' appears on the home screen with a flashing cursor after it. Finish the function notation by typing "(15)". Hit ENTER to evaluate.



c) According to the model, there were approximately 1188 stores in 1995.

Running other regressions on the data

From the scatter plot it is apparent the data has curvature. Since there is only one concavity, a quadratic model may be appropriate.



The quadratic graph appears to fit the data better than the linear graph. The first two handouts (*Evaluating Expressions*, and *Solving Equations*) used this quadratic model for their applications:

$$y = 2x^2 + 56x - 53 \text{ stores, where } x \text{ is the number of years after 1980.}$$