

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-85 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-85 as two lists:

1. Hit **STAT** and choose **EDIT**. Name 'xlist Name' to be **L1**, and 'ylist Name' to be **L2**. Hit **ENTER**.
2. If there is already data in L1 and L2 choose **CLRxy** to remove all previously stored data in the lists.
3. Type one ordered pair (an input value and its output value) at a time, hitting **ENTER** between each value.
4. Hit **ENTER** after the last set of data values, then hit **EXIT** twice.

```
xlist Name=L1
ylist Name=L2
```

```
x=L1      y=L2
x1=
y1=1
```

```
x=L1      y=L2
↑x6=14
y6=1115
x7=
y7=1
```

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. The TI-85 is capable of doing this without the program, but the program chooses the best window, which is very important for making decisions about the true shape of the graph. Run the program named **STPLT**:

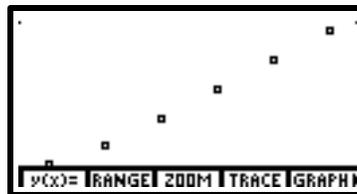
1. Before running the program, make sure all graphs are turned off. Hit **GRAPH** and choose '**y(x)=**'. Either clear out all equations, or turn them "off" by arrowing onto the equal sign and hitting **ENTER**. Hit **EXIT** twice.
2. To run the program, hit **PRGM** and choose **NAMES**, then **STPLT**.
3. 'STPLT' appears on the home screen. Hit **ENTER** to run the program.

```
y1=
```

```
NAMES EDIT
```

```
STPLT
```

4. The program creates a scatter plot *in the best window* and then quits. Hit **EXIT** to get back to the home screen.

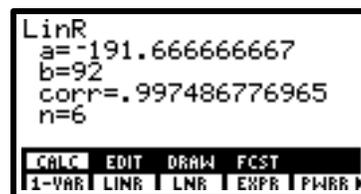


The data points appear to be fairly linear. (There is actually a slight curve to the scatter plot and a quadratic model will fit better.)

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 (note: the "y" must be lowercase):

1. Hit **STAT**, choose **CALC**.
Hit the **down arrow** twice.
(through the names L1 and L2 until the window and the menu on the bottom changes.)
2. Choose **LINR** from the new menu. (A quadratic would be P2REG) The TI-85 gives the y-intercept first, and then the slope: $y = a + bx$.
3. To store the regression equation into y1, hit **MORE** and choose **STREG**. "Name=" appears with a flashing cursor. To obtain a lowercase y, hit 2nd **ALPHA**, **Y**, then hit **ALPHA** twice, then "1". Hit **ENTER** to store the regression.
4. Hit **EXIT** twice to get back to the home screen.



a) The linear model for the data is:

$$y = 92x - 191.666666667 \text{ stores, } x \text{ years after 1980.}$$

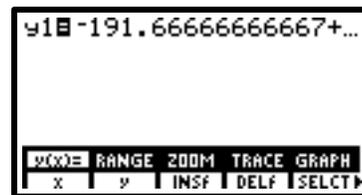
(It was rewritten to fit into the $y = mx + b$ form.)

- 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.

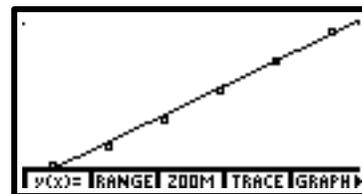
Viewing the regression equation in y1 and graphing

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, the program **STPLT** should be run again.



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

Predicting values using a model

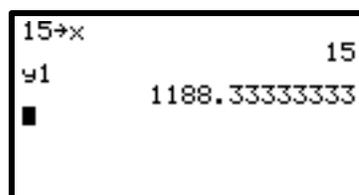
The model can be used to predict the number of stores in any year (close to the data values). The question was asked to predict the number of stores in 1995. Keeping in mind that the data was aligned to be the number of years after 1980, the desired input value (x value) to be used will be $x = 15$. 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" (not true function notation, but it is the name of the method). The equation was already stored into y1, and so the calculator will need to find y1 at x = 15:

1. From the home screen (if you are not there hit EXIT) store the value 15 into x by hitting "15" then STO► and "x". Hit ENTER.



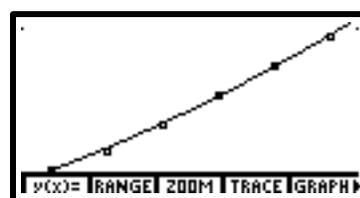
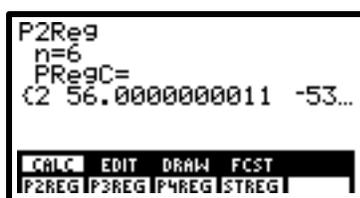
1. To evaluate the function at 15, type the function name by hitting 2nd ALPHA , Y and then "1". Hit ENTER. (A lowercase y is necessary.)



- 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:

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