

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-82 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-82 as two lists:

1. Hit **STAT** and choose **1:Edit...**
2. If there is already data in L1 arrow up to the name 'L1' and hit **CLEAR** then **ENTER**, to remove all previously stored data in that list.
3. Type all input values into L1. Hit **ENTER** between each value.
4. Use the right arrow key to go over to L2. Clear it, if necessary. Type all the output values into L2, hitting **ENTER** between each value. (Make sure both lists are the same length.)

```

STAT CALC
1:Edit...
2:SortA(
3:SortD(
4:ClrList
  
```

L1	L2	L3
-----	-----	-----
L1(1)=		

L1	L2	L3
4	213	-----
6	338	-----
8	522	-----
10	712	-----
12	918	-----
14	1115	-----
-----	-----	-----
L2(7)=		

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 PLOT** and choose **1:Plot1....**
2. Turn Plot1 on by hitting **ENTER** on the word 'On'.
3. Select the first type of plot (scatter plot). Make sure 'Xlist:' is L1 and 'Ylist:' is L2. (L1 - L6 are above the numbers 1 - 6.)
4. Before looking at the scatter plot, make sure all graphs are turned off. Hit "**Y=**" and either clear all equations out, or turn them "off" by arrowing onto the equal sign and hitting **ENTER**. (This keeps the equation stored without graphing it).

```

STAT PLOTS
1:Plot1...
  Off L1 L2
2:Plot2...
  Off L1 L2
3:Plot3...
  Off L1 L2
4:PlotsOff
  
```

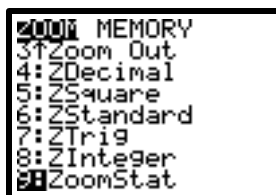
```

Plot1
Off
Type: [ ] [ ] [ ] [ ]
Xlist: [ ] L1 L2 L3 L4 L5 L6
Ylist: [ ] L1 [ ] L3 L4 L5 L6
Mark: [ ] +
  
```

```

Y1=[ ]
Y2=
Y3=
Y4=
Y5=
Y6=
Y7=
Y8=
  
```

5. To look at the data points in the best window possible, let the calculator choose the window by hitting **ZOOM**, then **9:ZoomStat**.



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.)

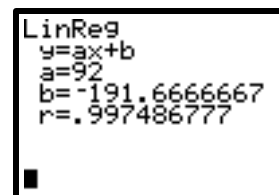
Performing a linear regression

Next, the calculator must find the equation of the line that best fits the data:

1. Hit **STAT**, arrow right to **CALC** and select **5:LinReg(ax + b)**.
2. The statement 'LinReg(ax + b)' should appear on the home screen, with a flashing cursor. Hit **ENTER**. If the data was stored in lists other than L1 and L2, the names of the lists would have to be typed, separated by a comma. i.e. LinReg (ax + b)L3,L4



The calculator will show a value for 'r'; this value will be ignored in this course.



- a) The linear model for the data is:

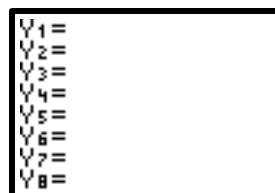
$$y = 92x - 191.6666667 \text{ stores, } x \text{ years after 1980.}$$

- 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.

Storing the regression equation into 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 first step is to have the calculator put the regression equation into Y1, rather than type it in manually. (Sometimes the regression equation is rather lengthy and cumbersome to type in.)

1. Hit 'Y='. Clear out Y1, if there is an equation, by hitting **CLEAR**. The cursor **MUST** be placed here before retrieving the regression equation (instead of having the cursor on the home screen).
2. Get the regression equation by hitting **VAR**s then **5:Statistics...**
3. Arrow right to **EQ** and choose **7:RegEQ**. This places the most recent regression equation into Y1.
4. To view the line, along with the scatter plot, hit **GRAPH**.



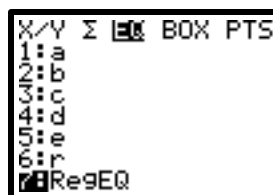
```

Y1=
Y2=
Y3=
Y4=
Y5=
Y6=
Y7=
Y8=
  
```



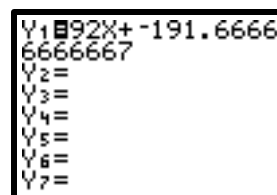
```

VARS
1:Window...
2:Zoom...
3:GDB...
4:Picture..
5:Statistics...
6:Table...
  
```



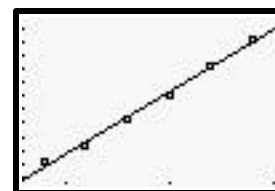
```

X/Y Σ [ ] BOX PTS
1:a
2:b
3:c
4:d
5:e
6:r
7:RegEQ
  
```



```

Y1=92X+ -191.6666
66666667
Y2=
Y3=
Y4=
Y5=
Y6=
Y7=
  
```



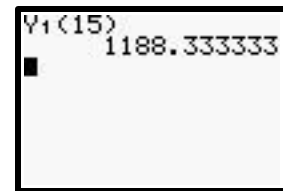
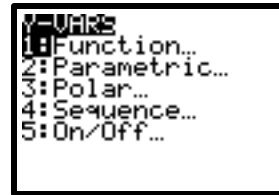
*It is now more apparent that the scatter plot has curvature.

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):

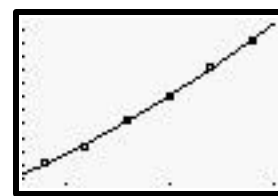
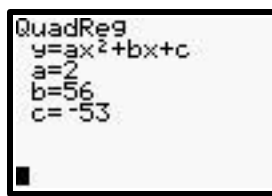
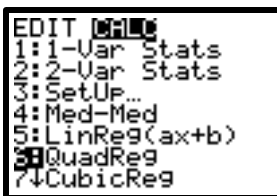
1. From the home screen (if you are not there, hit 2nd QUIT), get the variable name 'Y1' by hitting 2nd Y-VARS and choose 1:Function....
2. Choose 'Y1', unless the regression equation was stored into a different location.
3. 'Y1' appears on the home screen with a flashing cursor after it. Finish the 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:

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