

Lists and Statistics on the TI-83/4

This introduction assumes that you already have a basic familiarity with using TI-series calculators. You will be learning how to use the statistics functions of the calculator. Statistics involves collecting and analyzing data. We are often interested in determining whether one variable depends on another and in what way (e.g. linear, quadratic, exponential, etc.).

Population Growth

We will use the U. S. population for the years 1790-1890 given in the chart below to explore statistical analysis using the TI-83/84 calculator.

<i>Year</i>	<i>Population (millions)</i>	<i>Population Growth</i>
1790	3.929	1.379
1800	5.308	
1810	7.240	
1820	9.638	
1830	12.866	
1840	17.069	
1850	23.192	
1860	31.443	
1870	38.558	
1880	50.156	
1890	62.948	

It seems reasonable to consider that the larger a population gets, the more it will grow. In this investigation, we will be trying to answer the following questions.

Questions: Does the population growth in each 10-year period depend linearly on the size of the population? If, so what is the best line to describe the data

- First, complete the table above by calculating the growth in the population over each 10-year period. For example, from 1790 to 1800, it grew by $5.308 - 3.929 = 1.379$, which has been entered in the first row. **Note:** *The last row in the table above gives the population only so you can calculate the growth. There is nothing to enter into the black box. Also, when you enter data into the calculator later, you will not enter anything from that row.*

We will first introduce several of the calculator commands and menus that will be used in performing statistical analyses.

Lists and the Stat List Editor

The data for statistical analysis will be stored in lists. The TI-83/84 can store up to 20 lists in memory. To begin working with lists we will use the Stat List Editor.

- Press the [STAT] button located to the left of the arrow buttons.
You should see the following menu (left figure):

```

EDIT CALC TESTS
1:Edit...
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor

```

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

- Now select the Edit menu by pressing 1. You should see the table shown in the figure above (right). The list names L1, L2, and L3 are displayed on the top row and the number in the top right corner indicates the column of the currently highlighted list.
- Press the right arrow key to move through the six lists.

If there is already data entered in any of the columns or names other than L1-L6, you need to clear the lists and names from the editor*. Although you may not need to do this now, you will need to refer to these steps later.

- From the current editor screen, use the arrow keys move up to **highlight the actual list name**, (e.g. L2).
 - a) To delete the list elements only, press [CLEAR] followed by [ENTER].
 - b) To delete the entire list *and* its name, press [DEL].
- Use option **b)** to delete the list labeled L2 (even if it is empty). Notice that all other lists shift to the right by one column.

To restore the lists named L1-L6 to the first 6 columns,

- press the [STAT] button and select option 5: SetUpEditor. This will paste SetUpEditor into the home screen.
- Press [ENTER] and the word Done should appear.
- Return to the stat list editor by pressing [STAT], followed by option 1. Verify that the lists L1-L6 occupy the first 6 editor columns. Also make sure that there is no data stored in the lists.

To create a list of data,

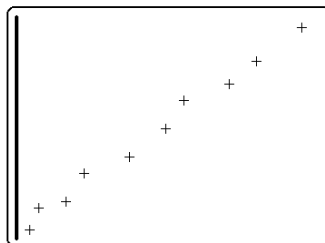
- use your arrow keys to highlight the space below L1. At the bottom of the screen you should see L1(1)=.
- Enter the value 3.929 and press [ENTER]. You will see the cursor move down a row in the same column and the display changes to L1(2) indicating the row or position of the data. Enter 5.308 in this position. Continue entering the *population* data from the second column of the table into list L1. [**Remember:** Do not enter 62.948 from the row for 1890 since it does not have any growth data.]
- Use the right arrow button to move to the second list L2 and enter the *population growth* data that you calculated previously. [**Remember:** Do not enter anything in L2 from the row for 1890 since it does not have any growth data.] Look over your data lists to be certain that you have entered the correct values. If any of the values are wrong, you can simply change them by using your arrow buttons to highlight the incorrect value, typing the correct value and then pressing [ENTER].

* The lists and names are only removed from the stat list editor, they are still retained in memory so that you may refer to them later.

Plotting

Let's first plot the data and see if we can find a relationship between the two sets of points.

- Press the [Y=] button to display the function editor. If any of the Y-variables already have functions defined, use the arrow buttons and [CLEAR] to clear them. Also, turn off any active plots. Recall that a plot is on if the plot and number (e.g. Plot1) are highlighted at the top of the screen. To turn off the plot, place your cursor over the highlighted plot name and press [Enter]. Similarly for the Y-equations, they are on if the equal sign is highlighted. You can also toggle between off and on by placing the cursor on the equal sign and repeatedly pressing [ENTER].
- Now press [2nd] [STAT PLOT] (located above the [Y=] button). A new screen will appear with the heading STAT PLOTS.
- Press [1] to select Plot1. You are now in the stat plot editor for Plot1. Your cursor should be over the word ON, press [ENTER] to turn on Plot1. To the right of the word Type, you should see 6 small plots with the top, left one highlighted. This first figure indicates a scatter plot in which each data point is plotted as a coordinate pair (x,y). The second figure indicates an xy-line plot in which small line segments joint the data points in consecutive order in the list. The third figure is for a histogram plot. We will only be concerned with the scatter plot option and the bottom row of plot options are beyond the scope of this project. If the scatter plot is not already selected, use the arrow buttons to select this option.
- Now move your cursor down to the Xlist option. Although your calculator has probably already entered the default value of L1, we will enter it explicitly so you can see how to refer to the pre-defined list names L1-L6. You should notice that L1-L6 are located in yellow on your calculator above the number buttons 1-6.
- Press [CLEAR] to remove L1 from Xlist.
- Now press [2nd] [L1] to place list L1 as the Xlist. Anytime you need to refer to lists labeled L1-L6 you can use this method. Enter L2 for the Ylist if it is not already there.
- You also have the option of changing the symbol used to plot each data point. Select the plus '+' mark.
- Press [ZOOM] and scroll down to find and select option 9:ZoomStat. You should now see a graph similar to the one below, which plots *population growth* (y-axis) and *population size* (x-axis).



Although all of the points do not lie on a single straight line, you should be able to see that the general trend looks linear.

- Using [Y=] menu, plot the line given by $y = 0.3x - 0.15$ on the same set of axes, to see that it fits the data approximately.

- Now plot the line given by $y = 0.25x + 1.0$ to see that it also seems to fit the data approximately.

Since more than one line seems to be a good approximation, we would like to know which line best fits the data points. We will use statistical analysis, namely linear regression, to determine the best-fit line.

- Before going on, clear the lines from Y1 and Y2.

Statistical Analysis

In order to determine a line which best fits the data, we will review a few terms that will help us measure how good the line fits the data.

The first measure is called the **residual** which measures the difference between the actual (experimental) y -value and the predicted y -value given by the line $y_L = mx + b$, i.e. $\text{residual} = y - y_L$. You can think of the residual measuring how well the line matched the data at each point.

The second measure is called the **residual sum of squares** SS_{res} which takes the residual at each data point, squares it, and then adds them all together. This SS_{res} essentially gives you one number that measures how well the line fits *all* the data points. So what we are looking for is a line that gives us the smallest possible value of SS_{res} .

Let's compute SS_{res} for the two lines considered above.

Creating new lists and attaching formulas

- Go to the stat list editor by pressing [STAT] [ENTER] where you should still see the values for x and y in L1 and L2.
- Use the arrow buttons to place your cursor over L3 in the top row.
- First we will insert a new list in column 3 and name it YL indicating the y -values for the first line. Press [2nd] [INS] will create a new, unnamed list in this column and the Name= prompt should appear at the bottom of the screen with the alpha-lock on.
- Type the letters YL from the keyboard and press [ENTER]. You will now see that the name YL appears as the list name of the 3rd column.
- Use your right arrow key to move over to the 4th column. You will see that the list named L3 now occupies this column since all lists were shifted right one column by inserting a new list. Insert a new list named RES for the residuals
- Repeat the process to insert another list named RES2, indicating the residual squared res^2 .

It is now possible to enter each of the values by hand, but it will be easier if we learn to attach formulas to a list name. This way, if we change any data values, the calculator will automatically recalculate the residuals from the new data.

- Return your cursor to be over the *list name* YL. The bottom prompt should be YL =. We want the list YL to **compute** the y -value from the line $y_L = 0.3x - 0.15$ for each value of x found in L1.
- Formulas are enclosed in quotation marks, so to attach the formula, enter [ALPHA] ["] 0.3*[2nd] [L1] - 0.15 [ALPHA] ["] [ENTER]. You should now see a list of values appearing under list YL starting with the value 1.0287. You will also see a small lock symbol to the right of the list name YL indicating that a formula is attached.

- Move your cursor over the list name **RES** to attach the formula for the residual $y - y_L$, using **L2** to refer to y and **YL** to refer to y_L . Press **[ALPHA]** **["]** **[2nd]** **[L2]** **[-]**. To finish entering this formula, you need to refer to the list named **YL**.
 - o To retrieve this list name, go to the list menu by pressing **[2nd]** **[LIST]** (found above the **[STAT]** button). The three submenu options listed across the top of the screen are **NAMES**, **OPS**, and **MATH**. You should currently be in the submenu **NAMES**, which lists all list names available in alphabetical order.
 - o Scroll down to highlight **YL** and press **[ENTER]**. Your screen should return to the stat list editor and **LYL** should now be pasted in your formula.
- Close the formula by pressing **[ALPHA]** **["]**. It is important that the small **L** designating a list appears before **YL**, otherwise your calculator will not understand your formula.
- Press **[ENTER]** and the calculator will calculate the residuals, with the first one as **.3503**. Now enter the correct formula for **RES2** to compute the residuals squared. The first value should be **.12271**.

The last thing that we need to compute is SS_{res} .

- Exit the stat list editor and go to the home screen by pressing **[2ND]** **[QUIT]**.
- Enter the list menu **[2ND]** **[LIST]** and choose the submenu **MATH**. Many of the options are self-explanatory, such as **min** and **max**, which will return the min and max values of the list entered as the argument. We would like to sum the list **RES2**.
- Choose option **5:sum(** and press **[ENTER]**. You should see **sum(** pasted to the home screen.
- Now go to the list names and choose **RES2** to paste to the screen.
- Press **[ENTER]** to compute the sum of this list. Enter the value for SS_{res} in the appropriate box in the table below (keep 4 decimal places).

Line	SS_{res}
$y = 0.30x - 0.15$	
$y = 0.25x + 1.0$	

Let's compute the SS_{res} for the second line $y = 0.25x + 1.0$.

- Go back to the stat list editor and place your cursor over the list name **YL** then press enter. Your cursor should now be in the formula.
- Edit this formula for the new line $y = 0.25x + 1.0$ using **[DEL]** and **[INS]** as needed. Be sure that you keep all necessary parts of the formula and quotation marks. Notice that the lists for **RES** and **RES2** change automatically because their formulas depend on these new values of **YL**.
- Compute SS_{res} for these values and enter it in the table above. [Hint: Using the **[ENTRY]** button saves time.]
- Make an educated guess for another pair of values m and b for the slope and intercept and write it in the table above.
- Repeat the process to compute SS_{res} for these new values and enter it in the table above.

Which of the three lines above is the best approximation (so far) to fitting the actual data? _____

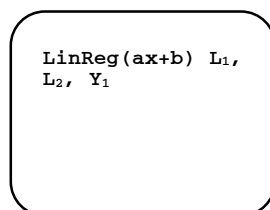
Linear regression

Although the calculator can easily calculate these values, we still have infinitely many combinations of m and b that we could try to find the lowest value of SS_{res} . Fortunately, there are formulas based on Calculus that will give us the values of m and b that will result in the smallest possible SS_{res} . Even more fortunate is that these formulas are already programmed into your calculator so that you don't have to do much work.

- Go to the stat list editor and remove the lists YL, RES, and RES2 by pressing [DEL] when the cursor is over the list name. You should see that L3 is again restored to column 3.
- Press [STAT] to enter the stat menu and choose the [CALC] submenu.

Starting with option 4 you will see several options ending in REG for regression (e.g. LinReg(ax+b), QuadReg, etc.). These options will find the values of coefficients that give the best type of fit for the desired function type. In our current example, we have already supposed that the data tends to follow a line, so

- Choose option 4:LinReg(ax+b), which will paste that option to the home screen.
- Enter [2nd],[L1][,][2nd][L2][,] and now press [VARS], choose the Y-VARS submenu, and select option 1:Y1 to paste Y1 to the home screen. It should look similar to:



LinReg(ax+b) L1,
L2, Y1

This command tells the calculator to use L1 for the x-values, L2 for the y-values and will store the resulting best-fit line in the variable Y1.

- Press [ENTER] and the calculator will give the optimal values for the slope a (same as m above) and the intercept b . Enter the values for a and b in the table below. (Note: We do not yet have SS_{res} .)

Slope a (m)	Intercept b	SS_{res}

- Press the [Y=] button to see that the formula for this line was automatically entered into the equation for Y1. (Clear any other lines, if necessary)
- Press [GRAPH] and you will see this line drawn in the plot along with the data points.

Finally, let's calculate the associated minimal SS_{res} for these values of slope and intercept. Whenever you perform a regression, the calculator automatically stores the residual vector in the list named RESID.

- Place this list on the stat list editor, by returning to the stat list editor, placing your cursor over L3 and pressing [2ND] [INS]. Now use the List names menu to find and place the RESID list in this column. You should see a list of values appearing starting with the value -0.5195 .
- Now we can compute SS_{res} by pressing [2ND] [QUIT] and then going to the LIST menu and selecting the MATH submenu. Choose again option 5:sum(, but this time give the argument LRESID^2 . This will square each term in RESID before summing them together.
- Enter this minimal value for SS_{res} in the table above. You should notice that it is, in fact, smaller than any of the SS_{res} values computed previously.

Another measure of how well the line fits the data set is to look at the r - and r^2 - coefficients. To view these values, you must turn on the diagnostics.

- Press [2ND] [CATALOG] and select the DiagnosticOn option to paste it to the home screen.
- Press [ENTER] and Done should appear.
- Now repeat the linear regression (Using the [ENTRY] button saves time.) and enter the values for r and r^2 in the table below. The closer $|r|$ and r^2 are to 1, the better the fit.

r	r^2

How well do you think the line describes the data? _____