EXPLORATIONS

CHAPTER 9

Statistics

In this chapter we will give an introduction to the statistical features of the TI-86. The tools available on the TI-86 will allow us to calculate statistical results, perform regression analyses, and graphically display statistical data.

Preliminaries

The baseball data below, taken during the second month of the 1996 season, will be used to illustrate the statistical capabilities of the TI-86.

Team	% of Games Won	Batting Average	Slugging %	Pitching ERA
Atlanta	66	0.277	0.450	2.75
San Diego	63	0.273	0.392	3.22
Montreal	57	0.275	0.426	3.95
Los Angeles	53	0.242	0.363	2.95
San Francisco	52	0.261	0.427	4.83
Colorado	51	0.283	0.460	5.90
Philadelphia	50	0.236	0.383	4.39
Florida	50	0.241	0.386	3.80
Houston	50	0.270	0.416	4.02
St. Louis	44	0.266	0.403	4.70
New York	42	0.265	0.406	4.40
Cincinnati	42	0.244	0.403	5.04
Chicago	41	0.247	0.396	4.60
Pittsburgh	38	0.252	0.400	4.69

Set the mode and graph format settings as shown in (2.0.1) and (2.0.2) of Chapter 2. Delete all functions in the ⟨**y(x)=**⟩ graph editor by using the ⟨**DELf**⟩ selection. In addition, since statistical data is stored in lists, it may be worthwhile to delete all possible existing lists from your calculator so that you can name the lists created in this chapter the same as we have named them. Recall that to delete lists from the calculator, press [2nd] [MEM] and select ⟨**DELET**⟩. Then select ⟨**LIST**⟩, and press [ENTER] as many times as is necessary to delete all possible existing lists from your calculator. As indicated in §25 of Chapter 1, the lists xStat, yStat, and fStat are system variables and cannot be deleted. However, they can be reduced to empty lists as we will illustrate in §6.

§1 – Entering and Viewing Lists

On the TI-86, data for a one-variable statistical analysis is divided into an x-list and an optional f-list that gives the frequency of each data point x. For a two-variable statistical analysis, data is divided into an x-list containing the independent variable sample points, a y-list containing the dependent variable sample points, and an optional f-list that gives the frequency of each data pair (x, y). Of course, the x- and y-lists are paired in the sense that the kth element of the y-list corresponds to the kth element of the x-list. In both the one and two variable cases, the default frequency for each data point is 1 on the TI-86. We will have no occasion to deviate from this default value, since if a data point has a frequency of say 3, then we merely list this data point 3 times (with a frequency of 1 each time).

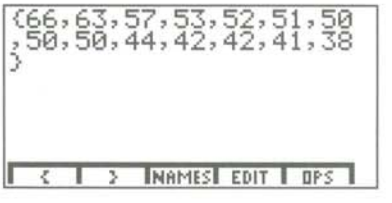
In order to analyze the baseball data, we will need to store four lists in the calculator: (1) the list of % of games won; (2) the list of batting averages; (3) the list of slugging percentages; and (4) the list of pitching ERAs. We will store these lists under the names WL, BA, SA, and ERA, respectively. Of course, we must store the data in sequential form; *i.e.*, the data for Atlanta must occupy the same list position for each of the four lists.

The most convenient way to enter lists on the TI-86 is to use the LIST editor. A second option is to enter a list directly from the home screen. The LIST editor can be accessed from either the LIST menu or the STAT menu. To access the LIST editor from the LIST menu, press [2nd] [LIST], and select ⟨**EDIT**⟩. To access the LIST editor from the STAT menu, press [2nd] [STAT], and select ⟨**EDIT**⟩.

1. We first illustrate the home screen method by entering the list WL from the home screen. From the home screen, type in

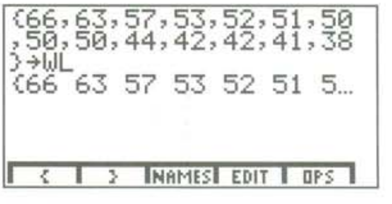
   {66, 63, 57, 53, 52, 51, 50, 50, 50, 44, 42, 42, 41, 38}

   as shown in (9.1.1). The characters **{** and **}** can be accessed by pressing [2nd] [LIST].



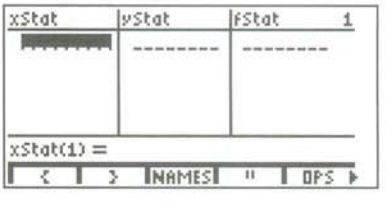
(9.1.1)

2. Then press [STO▶], type in WL, and press [ENTER] to store this list under the name WL. See (9.1.2).



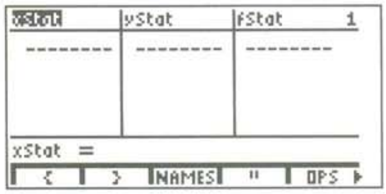
(9.1.2)

3. To view the list WL in the LIST editor, press [2nd] [LIST], or [2nd] [STAT], and select ⟨**EDIT**⟩ to obtain a display similar to (9.1.3). *Your display may be different depending on your previous use of the TI-86.*

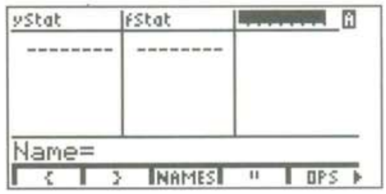


(9.1.3)

4. Use the [▲] key to move the block cursor to the top line as in (9.1.4), and then use the [▶] key to move the block cursor until it is in an unnamed column as in (9.1.5).



(9.1.4)



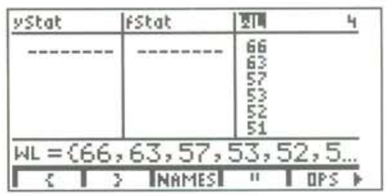
(9.1.5)

5. Press [ENTER] to obtain (9.1.6) in which the blinking [A] cursor is after Name=.

(9.1.6)

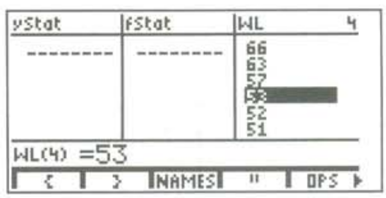
6. Then select ⟨**WL**⟩ to obtain (9.1.7), and press [ENTER] to obtain (9.1.8).

(9.1.7)



(9.1.8)

7. Use the up and down arrow keys to view the elements of the list as illustrated in (9.1.9), where we see that the 4th element of the list WL is 53.

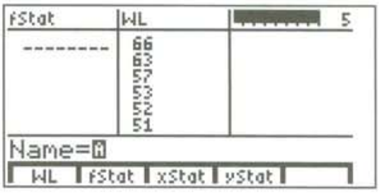


(9.1.9)

§2 – The LIST Editor Entry Method

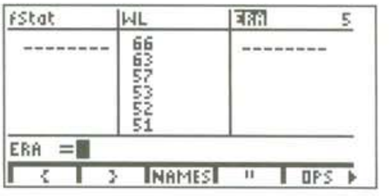
In this step we use the LIST editor to enter the list ERA.

1. With the display as in (9.1.9), use the [▲] key to obtain (9.1.8) again, press [▶] to move to an unnamed column, and press [ENTER] to obtain (9.2.1).



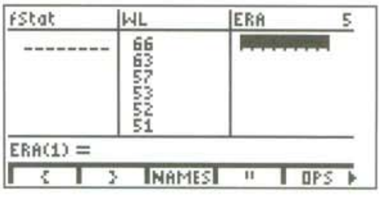
(9.2.1)

2. Type in ERA at the Name= prompt, and press [ENTER] [ENTER] to obtain (9.2.2).



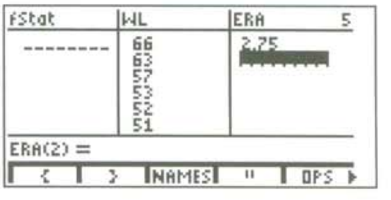
(9.2.2)

3. At this point, we could enter the ERA list much as we entered the WL list from the home screen in §1 by using the { and } characters. However, it is much more convenient to proceed as follows. With the display as in (9.2.2), press [ENTER] one more time to obtain (9.2.3).



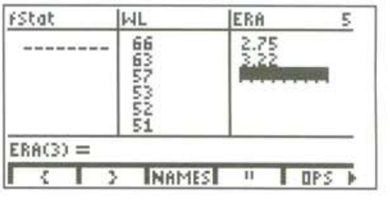
(9.2.3)

4. Type in 2.75 and press [ENTER] to obtain (9.2.4).



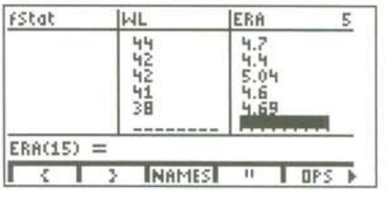
(9.2.4)

5. Type in 3.22 and press [ENTER] to obtain (9.2.5).



(9.2.5)

6. Continue in this fashion until all elements in the ERA list have been entered as is indicated in (9.2.6). *We emphasize again that we need to store the lists WL, BA, SA, and ERA in sequential form so that we will be able to carry out pair-wise comparisons of the four lists in question.* The two remaining lists, BA and SA, will be entered in §8 using the LIST editor.

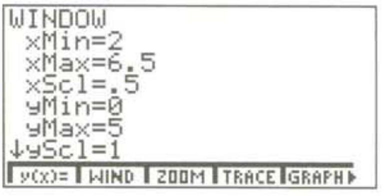


(9.2.6)

§3 – Plotting a Histogram Using STAT PLOT

In the next few sections, we will do a one-variable statistical investigation of the ERA data. We start by using the STAT PLOT feature of the TI-86 to plot a histogram of the ERA data. The first step in setting up a histogram plot is to define an appropriate viewing window and to either delete or deselect all functions in the ⟨**y(x)=**⟩ graphing editor. At the beginning of this chapter, we deleted the functions in the ⟨**y(x)=**⟩ graphing editor.

1. Thus, with the display as in (9.2.6), press [GRAPH] and select ⟨**WIND**⟩ to open up the viewing window. The ERA's range from 2.75 to 5.90. Thus, it is reasonable to take xMin = 2 and xMax = 6.5. The xScl value determines the width of the bars in the histogram. We will take xScl = 0.5. The height of each bar gives the number of x-values within the bar width. An x-value at a bar boundary is counted in the bar to the right. We take yMin = 0, yMax = 5, and yScl = 1. Enter these viewing window values as shown in (9.3.1).



(9.3.1)

Note: *Later we will see how the* ⟨**ZDATA**⟩ *selection within the ZOOM submenu can be used to have the TI-86 automatically select an appropriate viewing window for the ERA data.*

2. On the TI-86 up to three stat plots can be defined. The TI-86 refers to them as Plot1, Plot2, and Plot3. We will set-up a histogram plot for the ERA data in Plot1. With the display as in (9.3.1), press [2nd] [STAT], and select ⟨**PLOT**⟩ to obtain (9.3.2).

(9.3.2)

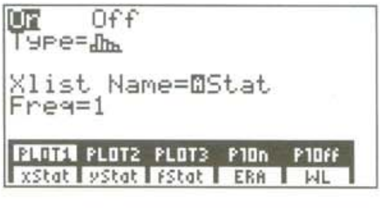
3. Then select ⟨**PLOT1**⟩ to obtain (9.3.3) in which the blinking block cursor is on On.

(9.3.3)

4. Press [ENTER] to select On and then [▼] to obtain (9.3.4) in which the blinking block cursor is after Type=.

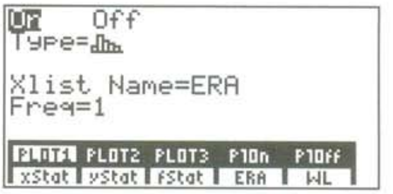
(9.3.4)

5. Select ⟨**HIST**⟩ and press [▼] to obtain (9.3.5).



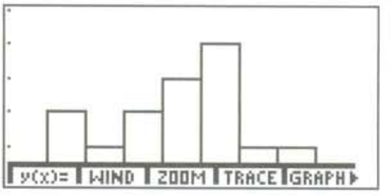
(9.3.5)

6. Then select ⟨**ERA**⟩, and press [ENTER] to save ERA as the new Xlist Name. The result is shown in (9.3.6) with the blinking cursor on the 1 in Freq=1. Since we want the default frequency of each ERA to be 1, we do not change this parameter. Plot1 is now set up to graph the histogram of the ERA data in the viewing window [2, 6.5, 0.5] × [0, 5, 1].



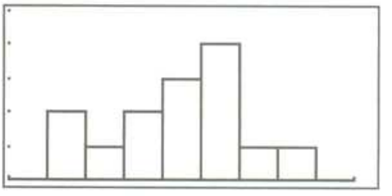
(9.3.6)

7. With the display as in (9.3.6), press [GRAPH] and select ⟨**GRAPH**⟩ to obtain (9.3.7).



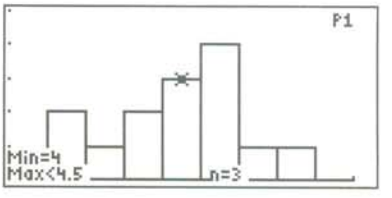
(9.3.7)

8. Press [CLEAR] to obtain (9.3.8) in which the bottom menu line has been removed.



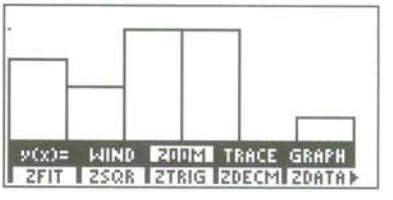
(9.3.8)

9. Press [EXIT] to obtain (9.3.7) again, select ⟨**TRACE**⟩, and use the [◄] and [►] arrow keys to trace along the histogram as is illustrated in (9.3.9). From (9.3.9), we deduce that there are three ERA's in the interval [4, 4.5).



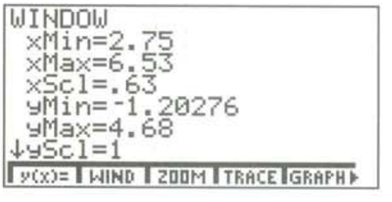
(9.3.9)

10. As mentioned earlier in this section, we can use ⟨**ZDATA**⟩ to have the TI-86 automatically select an appropriate viewing window. To illustrate, with the bottom menu line as in (9.3.7), select ⟨**ZOOM**⟩, press [MORE], and select ⟨**ZDATA**⟩ to obtain (9.3.10).



(9.3.10)

11. Then select ⟨**WIND**⟩ to obtain (9.3.11), the dimensions of the new viewing window. In most cases, we will choose our own customized viewing window rather than using the one forced on us by choosing ⟨**ZDATA**⟩.



(9.3.11)

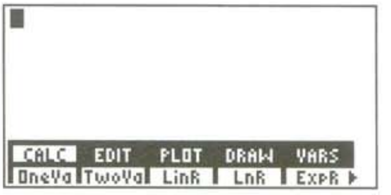
§4 – Using OneVar to Compute One-Variable Summary Statistics

The **OneVar** command is used to compute the summary statistics for one-variable data. This command has the syntax

OneVar *x-list*, *f-list*.

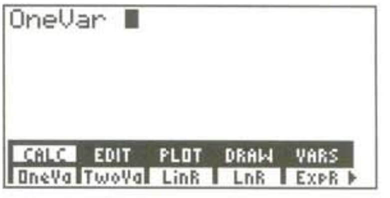
The *f*-list is optional. If none is given, then the TI-86 assumes that each data point in x-list has a frequency of 1. If neither an x-list nor an *f*-list is specified then the TI-86 assumes by default that x-list is xStat and that the frequency of each data point is 1.

1. To compute the summary statistics for the ERA data, return to a cleared home screen, press [2nd] [STAT] and select 〈**CALC**〉 to obtain (9.4.1).



(9.4.1)

2. Next select 〈**OneVa**〉 to obtain (9.4.2) and then type in ERA to obtain (9.4.3).



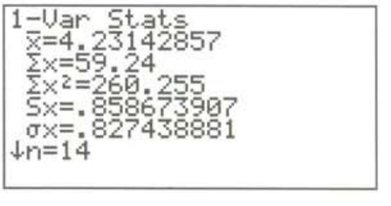
(9.4.2)

(9.4.3)

3. Press [ENTER] to obtain (9.4.4). The down arrow beside S_x indicates that we should use the down arrow key [▼] to scroll for more values.

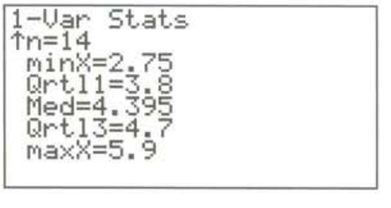
(9.4.4)

4. However, we first press [EXIT] twice to remove the two menu lines as in (9.4.5) so that we can view more of the statistics at one time.



(9.4.5)

5. Then as in (9.4.6), we scroll to view the remaining statistics.

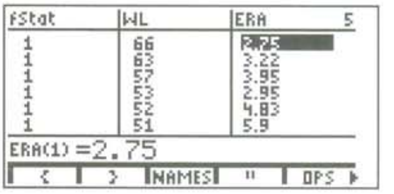


(9.4.6)

§5 – More on OneVar and the LIST Editor

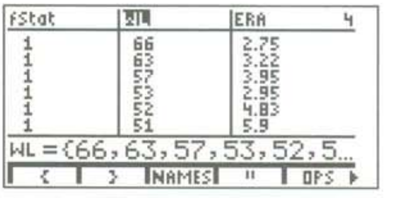
When we do a **OneVar** command, a duplicate of the x-list is copied to the xStat list, and a duplicate of the f-list is copied to the fStat list. Thus, in a one-variable statistical analysis, the calculated statistical results always match the data found in the xStat and fStat lists. We verify this fact for the present data; and, at the same time, show how to choose the lists to be shown in the LIST editor as well as the order in which they will be shown.

1. With the display as in (9.4.6), press [2nd] [STAT] and select 〈**EDIT**〉 to obtain what will probably will be (9.5.1).



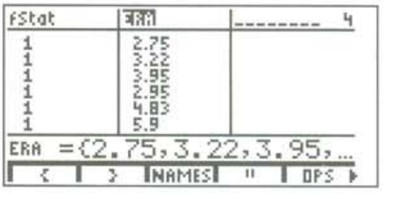
(9.5.1)

2. We would like to remove the WL list from the editor and have the three lists fStat, ERA, and xStat be the ones in view. To remove the WL list from the editor, move the block cursor to the top line of the WL column as shown in (9.5.2).



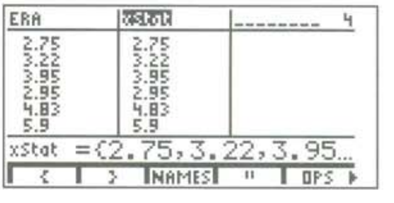
(9.5.2)

3. Then press [DEL] to obtain (9.5.3). The WL list has been removed from the LIST editor, but it has not been removed from memory. It can be recalled to the LIST editor just as was done in §1.



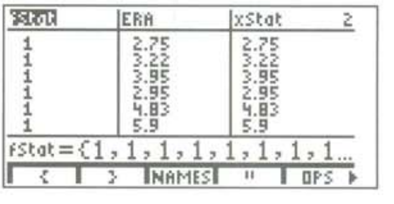
(9.5.3)

4. Then, as illustrated in (9.1.4)–(9.1.8) except select 〈**xStat**〉 rather than 〈**WL**〉, place the xStat list in the LIST editor as shown in (9.5.4).



(9.5.4)

5. Press [◄] twice to obtain (9.5.5). You can then use the down arrow key [▼] to scroll down the lists and verify that the ERA and xStat lists are identical and that all 14 entries of the fStat list are 1. Incidentally, we can place a list between two existing lists by using the [2nd] [INS] key combination when the block cursor is on the top line and to the right of the list to be added.

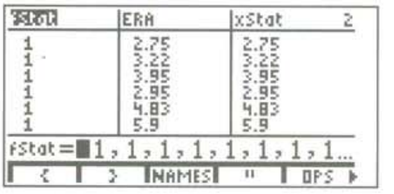


(9.5.5)

§6 – Editing Lists Using the LIST Editor

In this section, we explore additional important features of the LIST editor. We begin by clearing the fStat list; *i.e.*, we reduce the fStat list to an empty list.

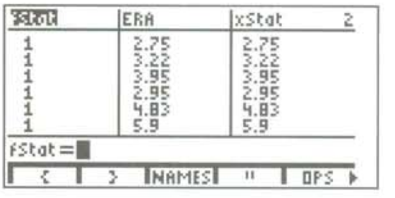
1. With the display as in (9.5.5), press ENTER to obtain (9.6.1) in which the blinking rectangular cursor is over the { to the right of fStat=.



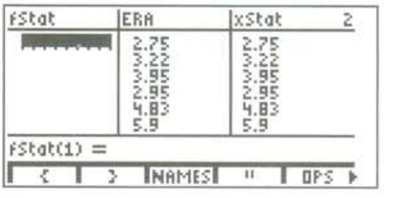
(9.6.1)

2. Press CLEAR to obtain (9.6.2), and then press ENTER to obtain (9.6.3). We have just reduced the fStat list to an empty list!

   Note: *Obviously, we must be very careful when using* CLEAR *in the LIST editor, since it is very easy to make an ill-advised press of the* CLEAR *key and lose important data.*

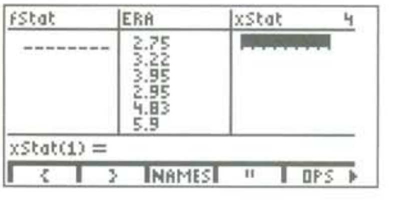


(9.6.2)



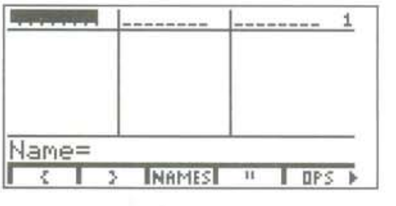
(9.6.3)

3. Use the same procedure to reduce xStat to an empty list as shown in (9.6.4).



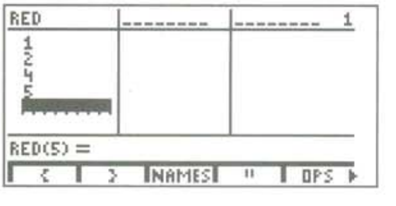
(9.6.4)

4. Next, delete all existing lists from the LIST editor using the DEL key as explained in §5 to obtain the display (9.6.5).



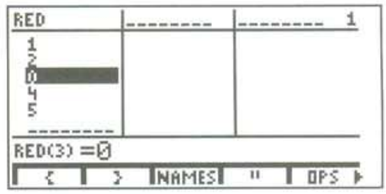
(9.6.5)

5. Then, obtain (9.6.6) by entering the list {1, 2, 4, 5} named RED. We will use the RED list to show how to delete, add, and change elements of a list.



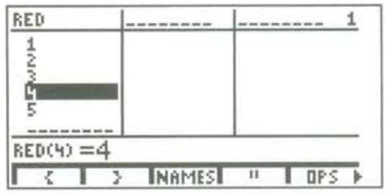
(9.6.6)

6. We begin by inserting the number 3 between 2 and 4 in the list. With the display as in (9.6.6), move the block cursor over 4 and then press [2nd] [INS] to obtain (9.6.7), in which 0 has been added by default.



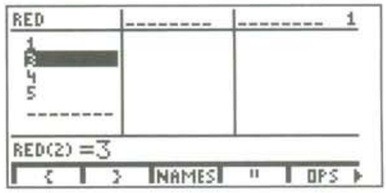
(9.6.7)

7. Type in 3 and press [ENTER] to obtain (9.6.8).



(9.6.8)

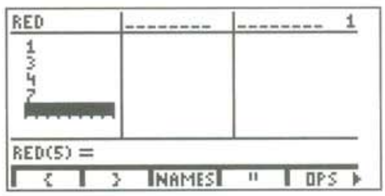
8. To delete the element 2 from the list in (9.6.8), move the block cursor to 2 and press [DEL] to obtain (9.6.9).



(9.6.9)

9. To replace the element 5 with the element 7, move the block cursor to 5, type 7, and press [ENTER] to obtain (9.6.10).

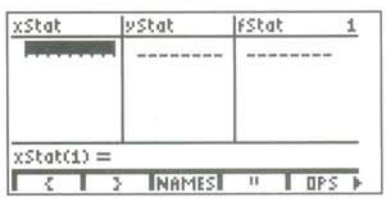
   Incidentally, if after typing in 7 in the previous operation, we decided not to change 5 to 7, we would have pressed [CLEAR] [ENTER] *to cancel the change.*



(9.6.10)

10. We want to remove all existing lists from the LIST editor and place the default system lists xStat, yStat, and fStat back into the editor in columns 1 through 3, respectively. The **SetLEdit** command performs this operation. From the home screen press [2nd] [LIST] ⟨**OPS**⟩ [MORE] [MORE] [MORE] and select ⟨**SetLE**⟩ to paste the command **SetLEdit** on the home screen. Then press [ENTER] to execute the command and obtain (9.6.11). Access the LIST editor to verify that the command has accomplished the stated purpose and has produced (9.6.12). Finally, before proceeding to the next section, access MEMORY and delete the list RED from memory.

(9.6.11)

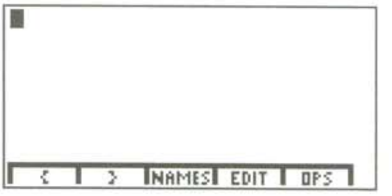


(9.6.12)

§7 – Using sum to Compute Summary Statistics

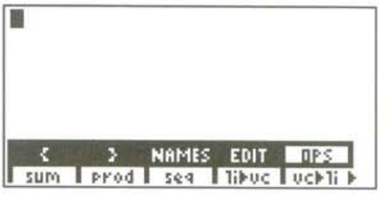
In this section we will illustrate another method to compute the first five summary statistics shown in (9.4.5), using the LIST arithmetic capabilities of the TI-86 and the **sum** command. Recall that the **sum** command sums the elements of the specified list. For example, **sum** ERA sums the elements of the ERA list.

1. To enter and compute this sum on the calculator, proceed in the following fashion. Return to the home screen and clear it. Press [2nd] [LIST] to obtain (9.7.1).



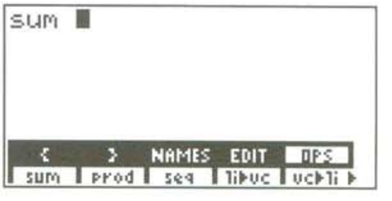
(9.7.1)

2. Select ⟨**OPS**⟩ and press [MORE] to obtain (9.7.2).



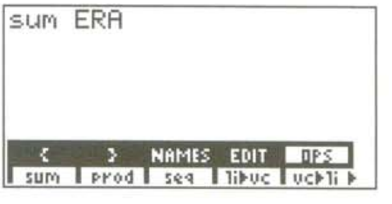
(9.7.2)

3. Select ⟨**sum**⟩ to paste the sum command onto the home screen as shown in (9.7.3).



(9.7.3)

4. Type in ERA as shown in (9.7.4).



(9.7.4)

5. Then press [ENTER] to obtain (9.7.5). From (9.7.5) we see that $\Sigma x = 59.24$.

(9.7.5)

6. Proceeding in a similar manner, we see from (9.7.6) that $\overline{x} = 4.23142857143$. Also, note from (9.7.6) that we have stored this value as AVG.

(9.7.6)

7. Display (9.7.7) shows that $\Sigma x^2 = 260.255$. We note that the new list ERA^2 is the list obtained by squaring each element of the ERA list.

(9.7.7)

8. Figure (9.7.8) shows that $S_x = 0.858673907325$. We remark that the list $(ERA - AVG)^2$ is obtained by subtracting AVG from each element of the ERA list to obtain an intermediate list and then squaring each of the resulting entries in this intermediate list to find the final list.

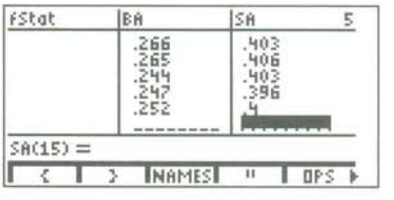
(9.7.8)

9. Finally, display (9.7.9) verifies that $\sigma_x = 0.827438881151$. It should be noted that the values for $\bar{x}$, S_x, and σ_x found in this section display more digits than those in (9.4.5). Later in §10, we shall see how to use the VARS submenu of the STAT menu to display the same number of digits as shown in this section.

(9.7.9)

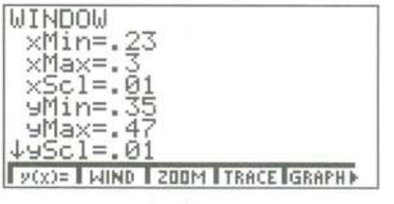
§8 – Scatter and xyLine Plots Using STAT PLOT

1. Use the LIST editor to enter the BA and SA lists as indicated in (9.8.1). We are going to use the paired BA, SA data to illustrate a two-variable statistical analysis. We begin by plotting a scatter plot and an xyLine plot for the data.



(9.8.1)

2. Set the viewing window as in (9.8.2). This is a reasonable viewing window since the batting averages range from 0.236 to 0.283, and the slugging percentages range from 0.363 to 0.460.

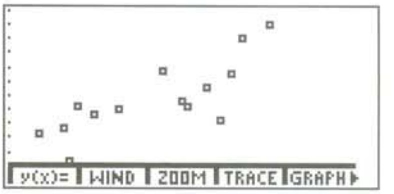


(9.8.2)

3. Then access the PLOT1 editor to obtain what will probably still be (9.3.6). Select scatter for the type plot, and finish the set-up to obtain (9.8.3).

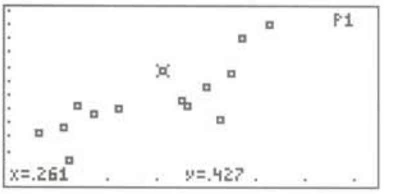
(9.8.3)

4. Press [GRAPH] and select ⟨**GRAPH**⟩ to obtain the scatter plot in (9.8.4).



(9.8.4)

5. Then select ⟨**TRACE**⟩ to track the data pairs as indicated in (9.8.5).

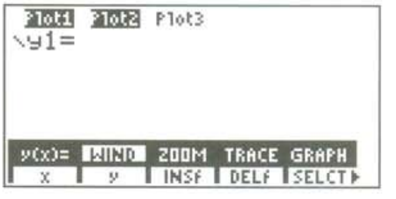


(9.8.5)

6. An xyLine plot is a scatter plot in which the data points are plotted and connected in order of their appearance in Xlist and Ylist. We will set up an xyLine plot for the BA, SA data. Access the PLOT2 editor, select xyLine for type, and finish the plot setup as indicated in (9.8.6).

(9.8.6)

7. Press [GRAPH] ⟨**y(x)=**⟩ to obtain (9.8.7). Observe in (9.8.7) that both Plot1 and Plot2 are highlighted indicating that they will both be graphed if we select ⟨**GRAPH**⟩. We can use the PLOT1 editor to turn off Plot1.

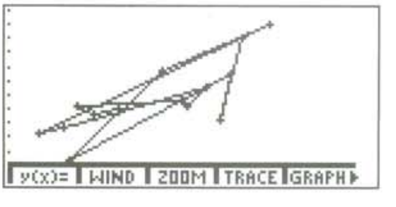


(9.8.7)

8. However, it is easier to turn off Plot1 from the graphing editor. With the display as in (9.8.7), use the arrow keys to make Plot1 a blinking highlight. Then select ⟨**SELCT**⟩ to turn off Plot1. Check that Plot1 is turned off by pressing the down arrow key to obtain (9.8.8). The ⟨**SELCT**⟩ selection used in this context acts as a toggle between on and off, just as it does in selecting and deselecting functions to graph.

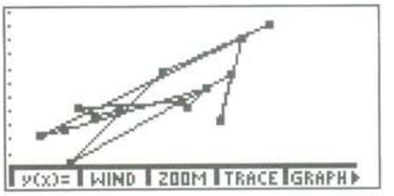
(9.8.8)

9. Then select ⟨**GRAPH**⟩ to obtain the xyLine plot in (9.8.9).



(9.8.9)

10. In (9.8.10), we have showed the result of leaving both Plot1 and Plot2 turned on. Evidently the + mark is just enough to fill in the open rectangle mark!



(9.8.10)

The xyLine plot in (9.8.9) is not very useful. However, if we sorted the data points (x, y) according to the ascending values of x before we did an xyLine plot, then a much more useful graphical interpretation would result. This sorting can be accomplished by using the **Sortx** command of the TI-86. However, if we sort the lists BA and SA in terms of ascending values of the BA list, then these two lists would no longer mesh with the WL and ERA lists. Therefore, we postpone dealing any further with this sorting problem until §11, when a very natural solution for the problem will be evident.

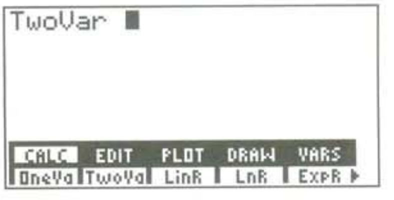
§9 – Using TwoVar to Compute 2-Variable Summary Statistics

The **TwoVar** command is used to compute the summary statistics for two-variable data. This command has the syntax

TwoVar *x-list, y-list, f-list.*

The *f*-list is optional. If none is given, then the TI-86 assumes that each data point (x, y) has a frequency of 1. If, in addition, neither an x-list nor a y-list is specified, then by default the TI-86 takes x-list to be xStat and y-list to be yStat.

1. To compute the summary statistics for the paired BA, SA data, return to a cleared home screen, press [2nd] [STAT] and select ⟨**CALC**⟩ to obtain (9.4.1). Next select ⟨**TwoVa**⟩ to obtain (9.9.1).



(9.9.1)

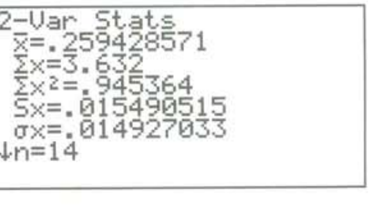
2. Complete the command shown in (9.9.2) by typing in BA and SA.

(9.9.2)

3. Then press [ENTER] to obtain (9.9.3).

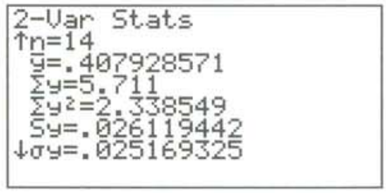
(9.9.3)

4. Obtain (9.9.4) by using [EXIT] to eliminate the menu lines as was done in §4.

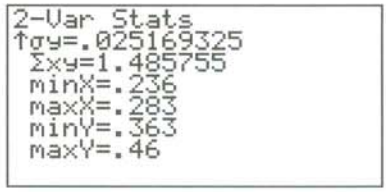


(9.9.4)

5. Scrolling then displays the remainder of the summary statistics shown in (9.9.5) and (9.9.6).



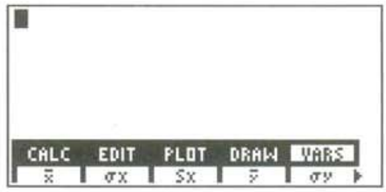
(9.9.5)



(9.9.6)

§10 – Recalling Summary Statistics

1. The summary statistics computed in a **OneVar** or a **TwoVar** command are stored in the calculator and can be conveniently recalled by using the VARS submenu of the STAT menu. With the display as in (9.9.6), press [CLEAR] to clear the home screen and then [2nd] [STAT] ⟨**VARS**⟩ to obtain (9.10.1).

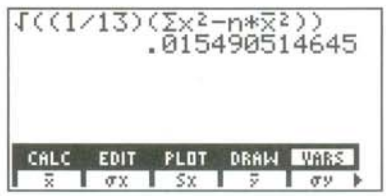


(9.10.1)

2. Select ⟨$\mathbf{S_x}$⟩ and press [ENTER] to obtain S_x as shown in (9.10.2). Note that the TI-86 has computed S_x much more accurately than was suggested by (9.9.4). This is true in general for the other summary statistics. Press [MORE], select ⟨$\mathbf{\Sigma x^2}$⟩, and press [ENTER] to obtain Σx^2.

(9.10.2)

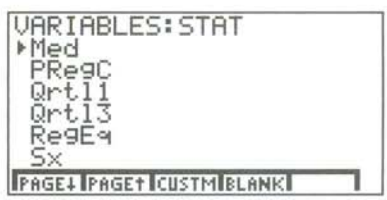
3. In (9.10.3) we illustrate how these stored statistical values may be used to do an arithmetic computation. We recognize the computation in (9.10.3) as the familiar short-cut method for computing S_x.



(9.10.3)

A second, although not as convenient, way to access the statistics is through the CAT/VARS menu. We illustrate this method to obtain S_x again.

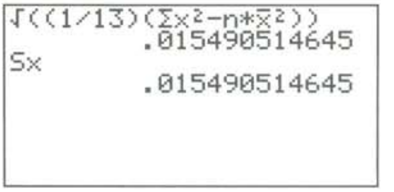
1. With the display as in (9.10.3), press [2nd] [CATLG-VARS] [MORE] [MORE] ⟨**STAT**⟩ to obtain (9.10.4).



(9.10.4)

2. Use the [▾] key to move the triangular pointer to the position shown in (9.10.5). Press [ENTER] to paste S_x on the home screen. Press [ENTER] again to display the value of S_x and obtain (9.10.6).

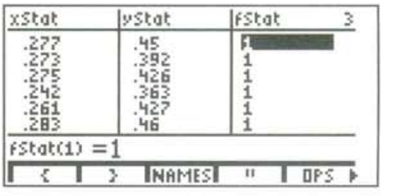
(9.10.5)



(9.10.6)

§11 – More on TwoVar, Sorting Lists, and xyLine Plots

When we do a **TwoVar** command, a duplicate of the x-list is copied to the xStat list, a duplicate of the y-list is copied to the yStat list, and a duplicate of the f-list is copied to the fStat list. Thus, the calculated statistical results match the data found in the xStat, yStat, and fStat lists. Display (9.11.1), obtained by accessing the LIST editor, verifies this for the present data.



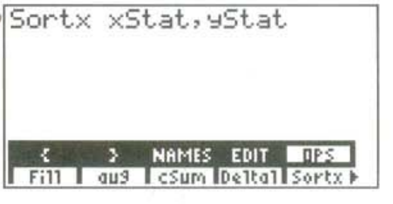
(9.11.1)

In §8, we indicated that we would be able to solve the problem with the xyLine plot without sorting the BA and SA lists. We will accomplish the desired objective by sorting the xStat and yStat lists and then plotting the xyLine plot for the sorted xStat and yStat lists. The sorting will be done by using the **Sortx** command which has the syntax

Sortx *x-list*, *y-list*.

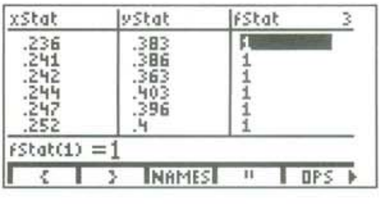
This command sorts the data points (x, y) in ascending order of the x-values.

1. Clear the home screen, and press [2nd] [LIST]〈**OPS**〉 [MORE] [MORE] 〈**Sortx**〉 to paste **Sortx** on the home screen. Then finish the command as illustrated in (9.11.2).



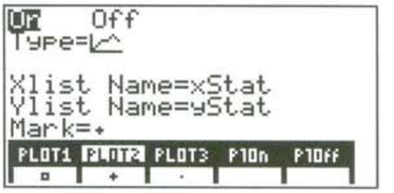
(9.11.2)

2. After pressing [ENTER] to execute the command, return to the LIST editor to obtain (9.11.3). We see that the xStat and yStat lists have been sorted as advertised.



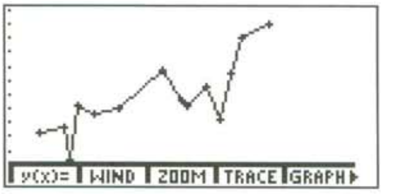
(9.11.3)

3. Set up Plot2 as indicated in (9.11.4). Make sure all other plots are off, and that the viewing window is still (9.8.2).



(9.11.4)

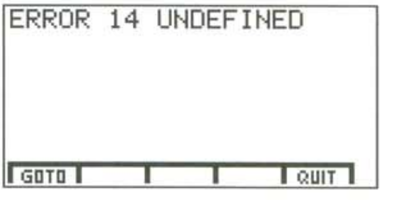
4. Then graph to obtain the xyLine plot in (9.11.5).



(9.11.5)

Note: *In a two-variable analysis, it is important to remember that if you edit the xStat, yStat, or fStat lists, then all of the statistical results stored in the VARS submenu of the STAT menu are cleared from the machine. Similarly, in a one-variable analysis if you edit the xStat or fStat lists then the statistical results are cleared.*

Since we edited the xStat and yStat lists when we sorted them, the summary statistics for the BA, SA data should no longer be stored in the machine. To verify this, from the home screen press [2nd] [STAT] ⟨**VARS**⟩ ⟨$\bar{x}$⟩ [ENTER] to obtain (9.11.6).



(9.11.6)

Before proceeding to the next step it is suggested that you execute the command

TwoVar xStat, yStat

and verify that the summary statistics are identical to those in (9.9.4)–(9.9.6).

§12 – Using LinR to Find a Linear Regression Line

In this section, we will determine the least squares linear regression line $y = a + bx$ for the paired BA, SA data and store this line as $y1$ in the graphing editor. We will also graph $y1$ with the scatter plot given by (9.8.4) in the viewing window [0.23, 0.30, 0.01] × [0.35, 0.47, 0.01] in order to visually see how well the regression line fits the data. The **LinR** command is the TI-86 command to find the least squares regression line. It is located in the CALC submenu of the STAT menu and has the syntax

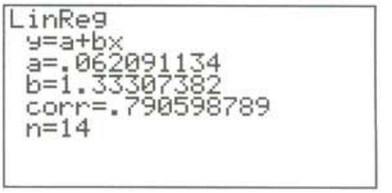
LinR *x-list, y-list, f-list, y-function.*

Except for the optional y-function argument the command has the same syntax as the **TwoVar** command. The optional y-function argument is a function name chosen from $y1$, $y2$, $y3$, etc.; and, if used, results in the regression line being stored under that name in the graphing editor.

1. To determine the least squares regression line and store it as $y1$, enter the command shown in (9.12.1).

(9.12.1)

2. Execute the command and clear the menu lines to obtain the display (9.12.2).

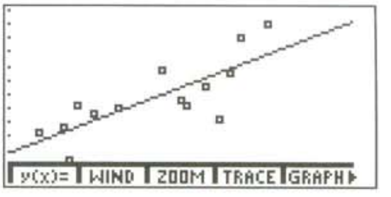


(9.12.2)

3. Then open the ⟨**y(x)=**⟩ graphing editor and note that $y1$ is the regression line. Make the necessary changes indicated in (9.12.3) so that only $y1$ and Plot1 are selected for graphing. (Recall that the Plot1 definition is given by (9.8.3).)

(9.12.3)

4. Finally, graph to obtain (9.12.4).



(9.12.4)

Like the **TwoVar** command, the **LinR** command and the other regression equation commands place duplicates of x-list, y-list, and f-list in xStat, yStat, and fStat, respectively. It is important to note that **LinR** also stores the values for a, b, and $corr$ as well as the same **TwoVar** summary statistics in (9.9.4)–(9.9.6), except for the min and max values in (9.9.6). Unfortunately, some of the other regression commands behave very differently as we will discover in §15.

§13 – Using sum to Compute SSD Values

Recall the least squares criterion given below.

> If S = $\{(x_i,y_i)\colon i = 1,\ldots,n\}$ is a set of data points and $f(x)$, a function depending on various parameters, is the type of curve you wish to fit to the data, then the best fitting curve of this type minimizes the sum of the squares of the differences
>
> $$\text{SSD} = \sum_{i=1}^{n} \left[y_i - f(x_i)\right]^2.$$

In §12 we asked the TI-86 to find the least squares regression line for the BA, SA data, and it determined it to be

$$y1 = 0.062091133907 + 1.33307382304x.$$

Thus, if x_i is the ith element of the list BA and y_i is the ith element of the list SA, then no matter how we choose a and b in $f(x) = a + bx$, we know that

$$\sum_{i=1}^{n} \left[y_i - f(x_i)\right]^2 \geq \sum_{i=1}^{n} \left[y_i - y1(x_i)\right]^2.$$

The minimizing SSD sum on the right-hand side of the previous inequality is easily computed on the TI-86 by executing the command

$$\textbf{sum } (\mathrm{SA} - y1(\mathrm{BA}))^2$$

as shown in (9.13.1) to give the minimal

$$\mathrm{SSD} = 0.003325436298.$$

(9.13.1)

§14 – Forecasting

The ⟨**FCST**⟩ selection found in the STAT menu by pressing [2nd] [STAT] [MORE] allows us to compute the predicted y-value for a given x-value using the current regression curve. It can also be used to compute the predicted x-value for a given y-value when the current regression curve is a linear, logarithmic, exponential, or power regression curve.

1. For example, with the display as in (9.13.1), press [2nd] [STAT] [MORE] ⟨**FCST**⟩ to obtain (9.14.1) with the blinking cursor on the x= line.

(9.14.1)

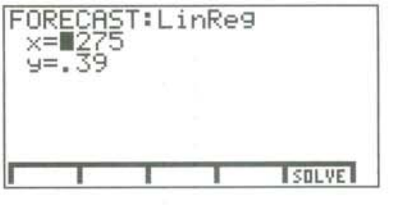
2. Type in .275 for x and move the blinking cursor down to the y= line with the [▼] key to obtain (9.14.2).

(9.14.2)

3. Select ⟨**SOLVE**⟩ to obtain (9.14.3). The rectangular mark to the left of y= indicates that 0.42868643524455 is the predicted value of y for the given x-value 0.275.

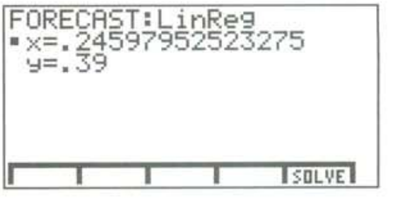
(9.14.3)

4. With the display as in (9.14.3) and the blinking cursor just after = on the y= line, type in 0.39 and move the blinking cursor to the x= line with the [▲] key to obtain (9.14.4).



(9.14.4)

5. Select ⟨**SOLVE**⟩ to obtain (9.14.5). Thus, 0.24597952523275 is the predicted x-value for the given y-value 0.39.



(9.14.5)

6. In (9.14.6) we see that we also could have used the home screen to predict the y-value for a given x-value.

```
y1(.275)
            .428686435245
```

(9.14.6)

§15 – An Introduction to Other Regression Models

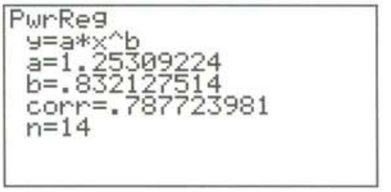
In addition to the linear least squares regression line, the TI-86 will also perform:		Represented on the TI-86 as Menu Item:
Logarithmic regression	$y = a + b\ln(x)$	**⟨LnR⟩**
Exponential regression	$y = a(b^x)$	**⟨ExpR⟩**
Power regression	$y = a(x^b)$	**⟨PwrR⟩**
Sinusoidal regression	$y = a\sin(bx + c) + d$	**⟨SinR⟩**
Logistic regression	$y = a/\left(1 + b*e^{cx}\right) + d$	**⟨LgstR⟩**
Quadratic polynomial regression	$y = ax^2 + bx + c$	**⟨P2Reg⟩**
Cubic polynomial regression	$y = ax^3 + bx^2 + cx + d$	**⟨P3Reg⟩**
Fourth degree polynomial regression	$y = ax^4 + bx^3 + cx^2 + dx + e$	**⟨P4Reg⟩**

The polynomial regression models are least squares fits based on the original (x_i, y_i) data. The logarithmic, exponential, and power regression models are based on a least squares line using transformed data $(\mathbf{X}_i, \mathbf{Y}_i)$. In the logarithmic case, the transformed data is given by $\mathbf{X}_i = \ln(x_i)$, and $\mathbf{Y}_i = y_i$. In the exponential case, the transformed data is given by $\mathbf{X}_i = x_i$, and $\mathbf{Y}_i = \ln(y_i)$. In the power regression case, the transformed data is given by $\mathbf{X}_i = \ln(x_i)$, and $\mathbf{Y}_i = \ln(y_i)$. The sinusoidal and logistic regression models are iterative least-squares fits. The logistic regression model is used in Exercise 3(b). The sinusoidal regression model will not be used in this book. All the regression model commands, except for the sinusoidal command, have the same syntax as **LinR**.

When we select **⟨LnR⟩**, **⟨ExpR⟩**, or **⟨PwrR⟩** to determine a regression equation, it is important to note that the corresponding statistics Σx, Σx^2, $\bar{x}$, S_x, σ_x, Σy, Σy^2, $\bar{y}$, S_y, σ_y, Σxy, and *corr* are the statistics for the transformed $(\mathbf{X}_i, \mathbf{Y}_i)$ data rather than for the original (x_i, y_i) data. In the cases **⟨P2Reg⟩**, **⟨P3Reg⟩**, and **⟨P4Reg⟩**, the statistics computed are the ones for the original (x_i, y_i) data. It is suggested that you use either the **TwoVar** or the **LinR** command if you want to compute the two-variable statistics for the original (x_i, y_i) data.

1. We illustrate these points by using ⟨**PwrR**⟩ to compute the power regression curve for the data in the BA and SA lists. In (9.15.1) and (9.15.2) we give the important displays in the process of finding the power regression curve.

(9.15.1)



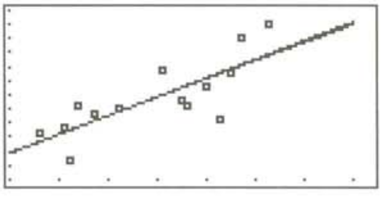
(9.15.2)

2. In (9.15.3) we use the STAT ⟨**VARS**⟩ selection to find Σxy, and in (9.15.4) we show that the value of Σxy is indeed based on the transformed data ($\ln(x_i)$, $\ln(y_i)$).

(9.15.3)

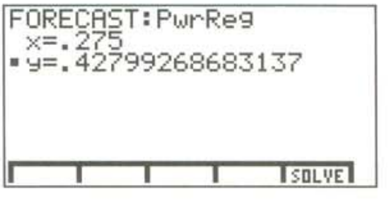
(9.15.4)

3. In (9.15.5) we have graphed the scatter diagram, the linear regression curve and the power regression curve in the same viewing window. Graphically, it appears that the linear and power regression curves are almost identical.



(9.15.5)

4. In (9.15.6) we used the ⟨**FCST**⟩ selection to predict the y-value for the given x-value 0.275 based on the power regression curve. Comparing (9.15.6) and (9.14.3), we see that the two models are in close agreement for x-values within the [0.23, 0.30, 0.01] × [0.35, 0.47, 0.01] viewing window.



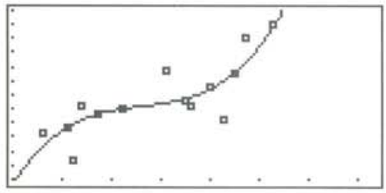
(9.15.6)

5. In (9.15.7) we show the result of finding the cubic polynomial regression curve $y = ax^3 + bx^2 + cx + d$ for the BA, SA data. The list on the fifth line of (9.15.7) gives the coefficients of the polynomial in the form {a, b, c, d}. You can scroll through the list to verify that it is given by

(9.15.7)

{2431.7238346, –1869.81297302, 479.678603882, –40.6532167572}.

6. In (9.15.8) we show the graph of the scatter diagram and the cubic polynomial regression curve.



(9.15.8)

§16 – Selecting the Best Model Based on SSD Values

For the BA, SA data we stored six regression curves in the ⟨**y(x)=**⟩ graph editor, $y1$ is the linear regression curve, $y2$ is the logarithmic regression curve, $y3$ is the exponential regression curve, $y4$ is the power regression curve, $y5$ is the quadratic polynomial regression curve, and $y6$ is the cubic polynomial regression curve.

1. In (9.16.1) and (9.16.2) we compute the SSD sum for each of these six regression curves. From the SSD values we deduce that the cubic polynomial regression curve is the best fit of the six curves, and that the logarithmic regression curve is the worst fit.

```
sum (SA-y1(BA))²
         .003325436298
sum (SA-y2(BA))²
         .003382268427
sum (SA-y3(BA))²
         .003280987956
```

(9.16.1)

```
sum (SA-y4(BA))²
         .003337750764
sum (SA-y5(BA))²
         .003019963752
sum (SA-y6(BA))²
         .002552054754
```

(9.16.2)

2. If we ask the TI-86 to find the fourth degree polynomial regression curve and store it as $y7$ by executing the command

   P4Reg BA, SA, $y7$

   we get the error display (9.16.3). This failure is due to the accuracy limitations of the TI-86 as it attempts to solve for the five polynomial coefficients.

```
ERROR 04 DOMAIN



GOTO                QUIT
```

(9.16.3)

3. We can still obtain the fourth degree polynomial regression curve and the accompanying SSD sum if we give the TI-86 a little help by scaling the BA, SA data by a factor of 10. From (9.16.4)–(9.16.6) we see that the TI-86 can handle the fourth degree polynomial regression curve for the transformed BA1, SA1 data. A little thought and the SSD result for the transformed data in (9.16.6) leads us to conclude that the SSD sum for the original BA, SA data would have been .00227561326263; and by adjusting the decimal point for the polynomial coefficients in (9.16.5), we could easily obtain the fourth degree polynomial coefficients for the BA, SA data.

```
10*BA→BA1
{2.77 2.73 2.75 2.42…
10*SA→SA1
{4.5 3.92 4.26 3.63 …
P4Reg BA1,SA1,y7
```

(9.16.4)

```
QuarticReg
 y=ax⁴+bx³+cx²+dx+e
 n=14
 PRegC=
{147.992042968 -1510…
```

(9.16.5)

```
sum (SA1-y7(BA1))²
        .227561326263
```

(9.16.6)

§17 – Linear Correlation

1. In (9.17.1) and (9.17.2) we see the linear regression results for the WL, ERA data.

```
LinR WL,ERA

 CALC   EDIT  PLOT  DRAW  VARS
OneVa TwoVa  LinR  LnR   ExpR ▸
```

(9.17.1)

```
LinReg
 y=a+bx
 a=7.60478944
 b=-.06756374
↓corr=-.64624297

 CALC   EDIT  PLOT  DRAW  VARS
OneVa TwoVa  LinR  LnR   ExpR ▸
```

(9.17.2)

2. In (9.17.3) and (9.17.4) we see the corresponding results for the WL, BA data.

```
LinR WL,BA

 CALC   EDIT  PLOT  DRAW  VARS
OneVa TwoVa  LinR  LnR   ExpR ▸
```

(9.17.3)

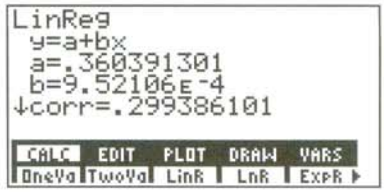
```
LinReg
 y=a+bx
 a=.214937363
 b=8.91097E-4
↓corr=.472465013

 CALC   EDIT  PLOT  DRAW  VARS
OneVa TwoVa  LinR  LnR   ExpR ▸
```

(9.17.4)

3. Finally, in (9.17.5) and (9.17.6) we see the corresponding results for the WL, SA data.

(9.17.5)



(9.17.6)

From the correlation results, we deduce that the highest correlation is between the won-lost percentage and pitching ERA. This is no surprise to baseball enthusiasts since it is a widely held view that good pitching is more important to a team's success than good hitting. It is perhaps surprising that the slugging percentage has such a low correlation with the won-lost percentage.

§18 – A Note on the Cumulative Distribution Programs

In Appendix B you will find cumulative distribution function programs for the Binomial, Poisson, Normal, and Student-*t* distributions.

Exercises

1. The course grades and ACT scores for 24 students taking a precalculus course at the University of South Alabama are tabulated below. The grade coding is the standard A = 4, B = 3, C = 2, D = 1, F = 0.

Course Grade	2	1	3	2	4	4	4	0	2	1	0	2
ACT Score	22	19	29	17	28	25	32	19	18	23	18	22

Course Grade	4	1	3	4	1	4	3	0	2	2	2	0
ACT Score	24	21	19	25	21	23	25	20	25	22	18	19

(a) Plot a histogram for the ACT scores using the viewing window [15, 35, 4] × [0, 12, 2]. Then use ⟨**TRACE**⟩ to find the number of scores in each of the five intervals [15, 19), [19, 23), [23, 27), [27, 31), [31, 35).

(b) Use **OneVar** to find the summary statistics for the ACT data.

(c) If the x and y variables are the course grades and ACT scores, respectively, then find the linear regression line $y = a + bx$ and the correlation coefficient.

(d) Graph the scatter diagram and the linear regression line in the viewing window [–1, 5, 1] × [15, 35, 4].

(e) Use the linear regression line to find the forecasted ACT score for each of the course grades A, B, C, D, and F.

2. In the table below, T is the number of decades since 1790, and P is the population (in millions) of the United States in the indicated year as reported by the U. S. Bureau of the Census.

Year	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
T	0	1	2	3	4	5	6	7	8	9	10	11
P	3.9	5.3	7.2	9.6	12.9	17.1	23.2	31.4	38.6	50.2	63.0	76.2

Graph the scatter diagram for the (T, P) data in the viewing window $[-1, 12, 1] \times [0, 90, 10]$. The scatter diagram indicates that an exponential fit may be an appropriate model. Compare, both graphically and by SSD values, the exponential and linear regression models and deduce that the exponential model is the better fit. Forecast the population in 1990 ($T = 20$) using the exponential model and compare it with the known population of 249 million in 1990 to conclude that the population demographics for the United States have changed since 1900. (Also, see Exercise 3.)

3. (A continuation of Exercise 2) The population data for the United States from 1910 to 1990 is given in the table below.

Year	1910	1920	1930	1940	1950	1960	1970	1980	1990
T	12	13	14	15	16	17	18	19	20
P	92.2	106.0	123.2	132.2	151.3	179.3	203.3	226.5	248.7

(a) Graph the scatter diagram for the 21 (T, P) data points given in Exercises 2 and 3 using the viewing window $[-1, 21, 1] \times [0, 260, 10]$. Compare, both graphically and by SSD values, the exponential and linear regression models and deduce that the linear regression model is the better fit.

(b) From (a) it appears that neither linear regression nor exponential regression is a very good model to describe the population of the United States from 1790 to 1990. The logistic regression model $y = a/\left(1 + b * e^{cx}\right) + d$ is frequently used to model population growth. Use this model ⟨**LgstR**⟩ on the 21 (T, P) data points. Graph the logistic regression curve and the scatter diagram in the viewing window specified in (a), and compute the SSD value for this regression curve to conclude that the logistic model does a good job of modeling the population of the United States from 1790 to 1990.

4. (a) The nine points in the table below all lie on the curve $y = 16 - x^2$.

x	–4	–3	–2	–1	0	1	2	3	4
y	0	7	12	15	16	15	12	7	0

Show that **P2Reg**, **P3Reg**, and **P4Reg** all yield (within round-off error) the same regression curve; namely, $y = 16 - x^2$. Is this to be expected?

(b) If $y1$, $y2$, $y3$, and $y4$ are the linear, quadratic, cubic, and fourth degree polynomial regression curves for a set of n data points (x_i, y_i) and

$$\text{SSD}k = \sum_{i=1}^{n} \left[y_i - yk(x_i)\right]^2$$

for $k = 1, 2, 3, 4$, then explain why $\text{SSD4} \le \text{SSD3} \le \text{SSD2} \le \text{SSD1}$.

(c) For each of the regression algorithms **LinR**, **LnR**, **ExpR**, **PwrR**, **P2Reg**, **P3Reg**, and **P4Reg** give the minimum number of data points (x_i, y_i) required for the algorithm.