Chapter 4
Numerical Methods for Describing Data

4.0 Introduction

Statistics is, of course, a mathematical science not only of graphic representation of data, but of numeric representation of data. The TI-83 family of calculators lumps its statistical calculations into one screen for easy reference, and you will see this screen very often when you are doing statistics. To find this screen, grab that calculator, press the **On** button (That's a joke. – we know you know that…) and then execute the following sequence:

Stat > **`Calc`**

You should now see a screen with calculation options – and there are sure a lot of them! (Again, notice that not all of these will actually appear in your calculator window at one time; that little arrow means you will have to scroll down to see all the options.

```
EDIT  CALC  TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg
8:LinReg(a+bx)
9:LnReg
0:ExpReg
A:PwrReg
B:Logistic
C:SinReg
```

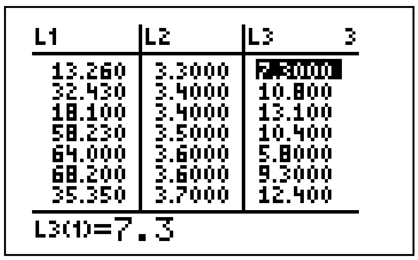
In Chapter 4, we are interested in describing the center and variability of a univariate data set, and using that information to construct a graphic representation of these aspects of data – the box plot. Let's pause briefly to mention some of the other selections in this screen. Not the choices <u>less</u> traveled by, the choices <u>later</u> traveled by. For the statistics in Chapter 4, our choice in this screen will always be "`1:1-Var Stats.`" This is TI-Speak for "Single Variable Statistics," our topic in Chapter 4. Choice 2, Two Variable Stats, will be rarely used; we will not use choice 4, the Median-Median line, nor Choices B and C, Logistic and Sine Regression. Therefore, this seemingly large number of choices is already cut down to a manageable size.

Before we can explore the single variable statistics options on the TI-83 we need some data. We already know how to enter data into the calculator (Remember, we did that in Chapter 3 with the mercury contamination data. You may still have that data in your calculator in List1, and for our purposes we will assume you still want to keep that data around. Therefore, we will enter our data in List2. Our real reason for using List2 is not that the mercury contamination is a really important data set, but to show you how to enter data if you don't want to obliterate data you have previously entered. It is frequently the case in data analysis that you will be working with more than 1 data set at a time – therefore, knowing how to work with more than one list, or at least work with something other than the default List1 is a necessary skill.

In any case, execute the sequence:

Stat > **Edit**

and you should see the following screen (Lest you have forgotten or otherwise lost your mercury data and don't remember, that data in List1 is the mercury contamination data.)



Now, let's calculate some statistics. Our first problem will be to find some measures of the center of a distribution: the mean and median.

4.1 Measures of Center

Examples 4.3 and 4.4: Number of visits to a class website

Forty students were enrolled a section of STAT 130, a general education course in statistical reasoning, during Fall quarter 2002 at Cal Poly, San Luis Obispo. The instructor made course materials, grades and lecture notes available to students on a class website, and course management software kept track of how often each student accessed any of the web pages on the class site. One month after the course began, the instructor requested a report that indicated how many times each student had accessed a web page on the class site. The 40 observations were:

20	37	4	20	0	84	14	36	5	331
19	0	0	22	3	13	14	36	4	0
18	8	0	26	4	0	5	23	19	7
12	8	13	16	21	7	13	12	8	42

We will use these data to demonstrate how to find the mean and median on your calculator, specifically using List2. By now your calculator may have turned itself off to conserve its batteries, so execute the following sequence:

Stat > Edit > Enter

(Remember, if the calculator's cursor is already on **Edit** – which it will be in this case, you need not "arrow over" to Edit, just press **Enter**.) You should once again now see the Edit screen with the mercury data in List1. Press the "right arrow" key to place the cursor in List2, and enter the data from Example 4.3. When you have completed that task, we are on the way to calculating the mean and median.

There is one slight adjustment we have to make because our data is in List2, however. The TI people, concerned about making the calculator efficient and easy to use, have built in what are known in the computer/calculator biz as "defaults." These defaults are the most commonly used keystroke sequences, and are designed to lessen the number of keys you actually have to press in your usual use of the calculator. As far as the TI folks are concerned, your "usual" use of the calculator would be to calculate statistics for List1.

Press the following sequence:

Stat > Calc > 1-Var Stats > 2nd > L_2 > Enter

and the following screen (complete with little arrow, not printed below) awaits you. As you might expect, there are lots of possible statistics that can be calculated for a univariate data set, and you will have to search the list of statistics for the mean and median. (You will have to search for the median via that down arrow, ▼.) Fortunately, it is pretty clear which statistics are the mean and median. $\bar{x}$ is the standard notation for the mean, and the median is abbreviated, naturally enough, "Med." But just in case, we point them out below.

1-Var Stats
$\bar{x}$=23.10000 ← The mean
$\sum x$=924.00000
$\sum x^2$=128158.000
S_x=52.33364
σ_x=51.67533
n = 40.00000
minX=0.00000
Q_1 = 4.50000
Med=13.00000 ← The median
Q_3 = 20.50000
MaxX=331.00000

4.2 Measures of Variability

Fresh from our success at measuring the center of a distribution, we will now tackle the variability in the form of the standard deviation. In what turns out to be a terrific stroke of luck on your part, finding the standard deviation is just like finding the mean. In fact, we have already done the necessary keystrokes needed to find it.

Example 4.3 again!: The standard deviation

Remember this example? Why, of course – it's the number of visits to the website that we just entered into List2. Just as before, we would enter our data in List2 and do the following sequence:

Stat > **Calc** > **1-Var Stats** > **2nd** > L_2 > **Enter**

Unless we stumble on our keystrokes we should see the screen below. Now, looking for the measures of variability is just like looking for the measures of center for a distribution. There is one slight hitch, in that the TI-83 will report 2 different standard deviations, and we need to keep them straight…

The sample standard deviation:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

The population standard deviation:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

1-Var Stats
$\bar{x}$=23.10000
$\sum x$=924.00000
$\sum x^2$=128158.000
S_x=52.33364
σ_x=51.67533
n = 40.00000
minX=0.00000
Q_1 = 4.50000
Med=13.00000
Q_3 = 20.50000
MaxX=331.00000

There is one little notation problem between our usual notation for the sample standard deviation and the population standard deviation. The TI puts a little subscript after our usual symbols, giving s_x and σ_x, but that seems to be a minor inconvenience.

4.2 Quartiles and the box plot calculations

Fresh from our earlier successes with the mean, median, and standard deviation, we will now construct the box plot. It will come as no surprise that you already know how to calculate all the statistics needed to plot the box plot – they are all in that box of statistics we used earlier when we did the mean, median, and standard deviation. What may come as a pleasant surprise is that the TI-83 will not only do the calculations for the box plot, it will draw the box plot as well. The data we will use for the box plot seems particularly appropriate – rectangles!

Example 4.11: Golden Rectangles

As you will no doubt recall from the text, observations were made on the variable x = width/ length for a sample of n = 20 beaded rectangles used in Shoshoni Indian leather handicrafts.

.553	.570	.576	.601	.606	.606	.609	.611	.615	.628	.654	.662
.668	.670	.672	.690	.693	.749	.844	.933				

Enter these data in List3 (or whatever List you choose) and we'll construct a boxplot on the TI-83. (If needed, you may wish to review the directions for entering data from our earlier discussion in Chapter 3.) The data is univariate, so the data entry is a repeat of what we've done already. Once again, we will point out the relevant statistics, just in case you might need them. Very soon we will get the boxplot without doing any arithmetic by hand.

	1-Var Stats
	$\bar{x}$=.66050
	$\sum x$=13.21000
	$\sum x^2$=8.88781
	S_x=.09251
	σ_x=.09017
	n = 20.00000
The smallest observation →	minX=.55300
The lower quartile →	Q_1 = .60600
The median →	Med=.64100
The upper quartile →	Q_3 = .68100
The largest observation →	MaxX=.93300

The quantities needed for constructing the modified boxplot are all here:

$$\begin{aligned}
&\text{MinX } = .553 \\
&\text{Lower quartile } (Q_1) = .606 \\
&\text{Median (Med) } = .641 \\
&\text{Upper quartile } (Q_3) = .681 \\
&\text{MaxX } = .933
\end{aligned}$$

From these quantities we can calculate…

$$\begin{aligned}
\text{iqr } (Q_3 - Q_1) &= .681 - .606 \\
&= .075 \\
1.5\text{iqr } &= .1125 \\
3\text{iqr } &= .225
\end{aligned}$$

…and now the mild and extreme outlier cutoffs:

$$\begin{aligned}
\text{Upper mild outlier cutoff} &= Q_3 + 1.5\text{iqr} \\
&= Q_3 + 1.5(Q_3 - Q_1) \\
&= .681 + .1125 \\
&= .7935
\end{aligned}$$

$$\begin{aligned}
\text{Lower mild outlier cutoff} &= Q_1 - 1.5\text{iqr} \\
&= Q_1 - 1.5(Q_3 - Q_1) \\
&= .606 - .1125 \\
&= .4935
\end{aligned}$$

$$\begin{aligned}
\text{Upper extreme outlier cutoff} &= Q_3 + 3\text{iqr} \\
&= Q_3 + 3(Q_3 - Q_1) \\
&= .681 + .225 \\
&= .906
\end{aligned}$$

$$\begin{aligned}
\text{Lower extreme outlier cutoff} &= Q_1 - 3\text{iqr} \\
&= Q_1 - 3(Q_3 - Q_1) \\
&= .606 - .225 \\
&= .381
\end{aligned}$$

4.3 Graphing the box plot.

From the quartiles and other quantities calculated above you can sketch a box plot on your paper with good accuracy, or the calculator can do it for you. The graphing calculator is not as good at representing box plots as you are – it has only a limited screen space, as we have seen. However the graph is perfectly reasonable for many classroom purposes.

You may recall from Chapter 3 our discussion of the general rules for statistical graphing on the TI-83:

1. Store the data somewhere in the calculator
2. Tell the calculator what graph you want
3. Make the graph

We have already entered our data, so the task at hand is to choose a plot. The sequence,

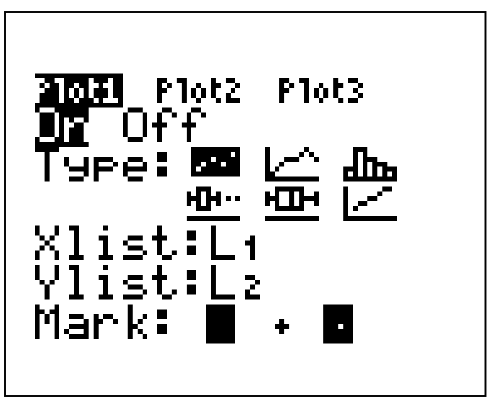
2^{nd} > STAT PLOT

```
STAT PLOTS
1:Plot1…Off
     L1  L2  □
2:Plot2…Off
     L1  L2  □
3:Plot3…Off
     L1  L2  □
4:PlotsOff
5:PlotsOn
```

again brings up the "Stat Plot" screen. If you have been using the calculator for graphs, your screen may differ from ours – do not be unnerved about this, since the plots we choose and how we choose them will fix any problems. First, we will "clear" all our plots by keying the sequence

`4:PlotsOff` > Enter

You may have set up other plots as you have used the TI-83 and wish to preserve them during your learning, so you need not pick the same plot we do. Please choose whichever of the 3 plot options you wish, and we will proceed using Plot1.

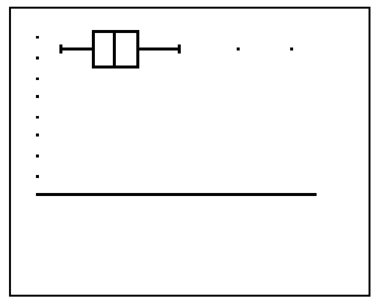


Remember, in our discussion in Section 3.3 we presented the Stat Plot screen, complete with English translation:

Icon→English Translation

Scatter plot	LinePlot	Histogram
Modified Box Plot	Skeletal BoxPlot	Normal probability Plot

What we have here are two box plot options: the modified (i.e. with outliers) and the skeletal (i.e. without outliers.) Generally speaking, we are always interested in knowing if there are outliers in a set of data, so our choice will always be the Modified Box Plot. Arrow over and choose that one. Again, we will inform the calculator which list contains our data for the box plot. Since our data is in List3, we will arrow down to `XList:` and pick L_3.



Now we have another choice to make. Since we are graphing a modified box plot, and are on the lookout for outliers, the TI-83 is giving us an option of how to display the outliers, in the form of "`Mark:`" We saw this before with the scatter plot, and we have the same choices for outlier indicators -- □ + ▪ -- so please pick your favorite and we will proceed to graph the box plot. You can get a fairly good look at the box plot using the **Zoom** > **`ZoomStat`** sequence.

There is still a slight problem, in that we aren't quite sure what the scale is when we graph the box plot. If we press the **WINDOW** key we can see the calculator's choice for a window. Our window is presented at the right

```
WINDOW
Xmin=.515
Xmax=.971
Xscl=1
Ymin=-1.363
Ymax=17.263
Yscl=2
Xres=1
```

The problem here is that the `Xscl` is set at 1, and all our data is between 0 and 1. Therefore, we don't get any tick marks to help us with the graph. Press the **WINDOW** key, change the `Xmin` to .5, and the `Xscl` to 0.1. Now you should have tick marks – of a <u>known</u> size -- that will aid you in interpreting the box plot.

4.3 An afterword about interpreting the box plot.

There actually is another method you can use to help interpret your box plot on screen as well as get numerical values for the minimum, maximum, quartiles, and any outliers – the **TRACE** key. With the box plot shown, press the **TRACE** key and use the ► and ◄ keys to see these different values. This tracing procedure can also remind you what the scale and tick marks are if you have forgotten.