

Activity: Describing a List of Numbers with the TI-82/83

(Original by K Johnson and T Ordoyne. Revised Su'97 by B White.)

This lesson contains information that will enable you to organize and describe a list of numbers and to summarize your data using appropriate graphs. You will also gain experience with the graphing calculator. After working carefully through this activity, you will need to practice these methods with additional lists of numbers, either data you collect yourself, or any of the data sets generated in class or provided by the instructor.

When you see <enter>, this means to press the **ENTER** key.

The following are the weights in pounds of 40 selected members of the Carolina Panthers football team, obtained from the Spartanburg *Herald Journal*.:

195	216	207	183	295	195	199	305	240	178
240	257	185	225	178	240	208	242	188	224
294	310	215	320	275	200	218	288	181	195
290	214	315	188	226	226	185	210	248	335

To input this list in the calculator, press **STAT**. Under **EDIT**, choose **1:EDIT** to edit the data. You should see three columns (label headings L_1 , L_2 , L_3) in which you may enter data. If you \rightarrow to the right, you will see that there are a total of six lists in which you can store data. Additionally, on the TI-83, you have the option of having user-named lists. You can have as many lists as the memory allocated for lists can hold. Now return to L_1 . If it is not empty, you can clear it by putting the cursor on the heading and pressing **CLEAR**, then <enter>. Now enter data values one at a time, each followed by <enter>. **The importance of entering the data correctly cannot be stressed enough; incorrect data makes all computation done with the data incorrect.** Put all the data into L_1 . Note that the calculator numbers the entries as you enter them. For example $L_1(4)$ means List 1, 4th entry.

To arrange your data from lowest to highest (ascending order), you use the sort feature. Before you sort the data, you may want to keep the data in its original order of entry. You can do this by copying one list to another. Here's how to copy L_1 to L_2 : Be sure L_2 is empty, then arrow and put the cursor on the label heading. Press **2nd 1** (L_1); notice that the line at the bottom of the screen says $L_2 = L_1$. <enter> and you should see that L_1 is duplicated into L_2 . To sort the data in ascending order, use **STAT** \downarrow **2:SortA**(<enter>. You will see SortA(with the blinking cursor. The cursor is asking which list to sort. To sort List 1, enter L_1 and <enter>. The display tells you that the task is DONE. **SortD**(puts the data in descending order. To view your list, again do **STAT, EDIT** \downarrow **1:Edit**. You can now easily see that the lightest weight of the Panthers on our list is 178 pounds, and if you arrow down to view the entire list, you should see that the heaviest Panther on our list is 335 pounds. Using standard statistical notation, you would write: $\min = 178$, $\max = 335$. NOTE: Having the list in order makes it much easier to group the data for a frequency table or histogram; or to find minimum and maximum values, medians, or quartiles for a boxplot.

Now create a frequency table for your data. This means that you are going to sort your data into intervals or classes. Make sure that you have at least five and no more than twenty classes (usually six to ten classes). Also make sure that the classes are of equal width and start or end at convenient endpoints. To determine the appropriate class width for your data set, you first examine the range of the data. The range is a number obtained by computing $\text{max} - \text{min}$. For this data set, the range = $335 - 178 = 157$ pounds. Divide the range value by the number of classes and round up to a neat number for your class width. For example, if you want to use seven classes, you would divide the range 157 by 7 to get 22.4, round up to 25, and set up classes of 175-199, 200-224, 225-249, ..., 325-349. If you want nine classes, divide 157 by 9, get 17.4, round up to 20 and use classes 160-179, 180-199, 200-219, ..., 320-339. Now, sort your data values into the classes you have chosen and count the frequencies - the number in each class. Record the data below:

<u>Weight Classes (pounds)</u>	<u>Frequency</u>	<u>Histogram</u>	<i>(draw here)</i>
--------------------------------	------------------	------------------	--------------------

The histogram is the graph that displays the information found in the frequency table. The values of your variable (in this case, the weights) go on the horizontal axis, with the frequencies on the vertical axis. Both axes **must** be labeled. The vertical axis should start at zero, so that the heights of the rectangles show how many data values are in each class. The TI will draw the histogram for you, but you have to tell it which classes to use. Suppose you are using the nine classes of width 20 with classes 160-179, ..., 320-339. To prepare to produce the histogram, choose **WINDOW** and enter $X_{\text{min}} = 160$ (lower end of the first class), $X_{\text{max}} = 340$ (upper end of last class + 1), and $X_{\text{scl}} = 20$ (class width). Since the y-values for a histogram are the frequencies, you would look at the frequency table you just made to see the largest frequency of any class. Set $Y_{\text{min}} = 0$, $Y_{\text{max}} =$ a value larger than the highest frequency (so that the top of the histogram doesn't get "chopped off"), and Y_{scl} appropriate for readability of the finished graph (usually 1).

Before producing any statistical graphs, clear or deselect all equations from the function (Y=) list, then press **STAT PLOT** (2nd Y=). If any of plots 1,2 or 3 is on, press **4:PlotsOff** and <enter> to turn it off. Then, return to **STAT PLOT** to set up your histogram. Choose **1: Plot 1**, turn Plot1 ON, and choose the histogram for the Type, L₁ for the Xlist, and 1 for the frequency. Now, press **GRAPH**, and your histogram should appear. After you have produced a histogram, the **TRACE** key and directional arrows will let you check the frequencies in each class. Draw the histogram in the area provided on the previous page.

The five-number summary consists of the minimum value, 1st quartile, median, 3rd quartile, and the maximum value (min, Q₁, med, Q₃, max). The min and the max are easily read from the sorted data list. The median separates the bottom half of the list from the top half. If the sample size n is odd, there is one score that is right in the middle of the list; however, if the sample size n is even, there is no “middle” score, and we use the average of the two middle scores. How do we quickly find the median for a long list? Calculate .5n, which will not be an integer if n is odd. In that case we round up to the next integer, call it k, and use the kth entry, or L₁(k). If n is even, then .5n is an integer, which we’ll call k. In that case the median will be the average of entries k and k+1, or $[L_1(k)+L_1(k+1)]\div 2$.

We use a similar strategy to locate the quartiles. The first quartile, Q₁, is the entry that separates the bottom fourth of the list from the top three-fourths. So, we calculate .25n and proceed as with the median, depending on whether .25n is an integer or not. We calculate .75n to locate the third quartile, Q₃, the entry which separates the top fourth of the list from the bottom three-fourths. For example, if there are 50 numbers in our list (n = 50):

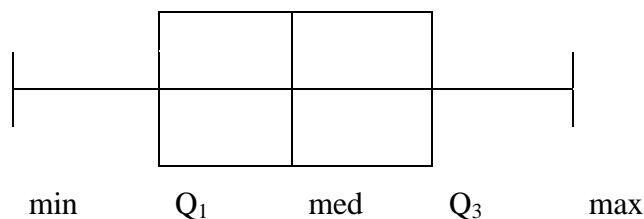
$.25n = .25(50) = 12.5$ (not an integer) - so we round up to 13, and Q₁ will be the 13th entry, L₁(13)

$.5n = .5(50) = 25$ (an integer) - so our median will be the average of the 25th and 26th entries, $[L_1(25) + L_1(26)] \div 2$

$.75n = .75(50) = 37.5$ (not an integer) - so we round up to 38, and Q₃ will be the 38th entry, L₁(38)

For your list of football players’ weights, n = 40. To locate the median, you calculate .5(40) = 20, so you use the average of the 20th and 21st entries. L₁(20) = 218 and L₁(21) = 224, so the median is halfway between 218 and 224, or 221. To locate the first quartile, you calculate .25(40) = 10, so you use the average of the 10th and 11th entries. L₁(10) = 195 and L₁(11) is also 195, so the first quartile is 195. Similarly, the third quartile is the average of the 30th and 31st entries - halfway between 257 and 275, or 266. Your five-number summary for the Panthers’ weights would read: min = 178, Q₁ = 195, med = 221, Q₃ = 266, max = 335. You can now describe the distribution of weights as follows: “The Panthers’ weights range from a low of 178 to a high of 335 pounds. Half of the Panthers in the sample weigh less than 221 pounds; half weigh more than 221 pounds. Only a fourth of the players weigh less than 195; a fourth weigh more than 266.” Another way to state the information given by the quartiles would be: “The middle half of the players weigh between 195 and 266 pounds”, or “The interquartile range is from 195 to 266 pounds.”

The boxplot is the graph that displays the information found in a five-number summary. It consists of a number line representing the values from min to max. It is **always** drawn to scale. Shown below is the general form of a boxplot.



If you want to produce the boxplot on the TI, and you have already set up the appropriate window values for the histogram, you may leave those values in **WINDOW**. To produce the boxplot, go to **Stat Plot**. Turn off the plot containing your histogram. If you wish to save the histogram for later use, then set up the boxplot on Plot 2. Again press **Stat Plot** and arrow down to Plot2, turn plot 2 ON, and choose the boxplot with L₁ for the Xlist and 1 for the frequency. Notice on the TI-83 that you have two boxplots available - standard and modified (the one with the *'s on it). Choose the standard boxplot. Now press **GRAPH** and your boxplot should appear.

After you have produced a boxplot, the **TRACE** key and directional arrows left and right will let you check your values for min, Q₁, med, Q₃, and max. If you want to compare two or three lists, set up a boxplot for each, and the calculator will display them in the same screen.

WARNING: occasionally, the calculator will come up with slightly different values for the quartiles than if you calculated them yourself according to the previous instructions. This is because the TI calculates Q₁ as the “median of the values to the left of the median” and Q₃ as the “median of the values to the right of the median”, instead of using the procedure spelled out in this activity. Your instructors prefer you to use the method of this activity, because this method is uniformly accepted, and it generalizes to use in locating other percentiles. For example, if you wanted to locate the 90th percentile for a list of numbers, or the score that separates the top 10% from the bottom 90%, you would calculate $.9n$ and then proceed as in our instructions, depending on whether or not $.9n$ is an integer.

You have already organized your data in a grouped frequency table to show the distribution, and presented that information in a histogram. You have found the five-number summary, which you presented in a boxplot. You may also wish to describe your data set with two other commonly used statistical measures: the mean, which is a measure of central tendency (middle value), and the standard deviation, which measures the spread away from the mean.

You can find the mean and standard deviation using the **STAT** menu. Under **CALC**, choose **1:1-Var Stats** and <enter> to put the command on the screen. Now select the list containing the data, L₁, (on the screen should be **1-Var Stats L₁**) and <enter>. You will see a list of values which uses standard statistical notation. The mean is denoted \bar{x} . If you had to calculate it yourself, you would find the sum of all the values and divide by n. Σx denotes the sum of all the values and n appears at the bottom of the screen.

The standard deviation tells the typical amount of spread away from the mean. In most lists, roughly two-thirds of the entries will be within one standard deviation of the mean; almost all of the entries will be within two standard deviations of the mean. You will find the sample standard deviation under 1-Var Stats as s_x . Record the mean and standard deviation to one decimal place more than the original data values. For example, if your data values are all whole numbers, you would record the mean and standard deviation to the nearest tenth. For your list of Panthers' weights, the mean weight is 233.3 pounds, with a standard deviation of 45.9 pounds. You can now describe the Panthers' weights as follows: The average weight of the players in our sample is 233.3 pounds. If this sample is random (and therefore representative of all Panther weights), about $\frac{2}{3}$ of the weights of all players will be within 45.9 pounds of that value.

Go back to your list and count how many of the players' weights are within one standard deviation of the mean; within two standard deviations of the mean.

"Within one standard deviation of the mean" would be between _____ and _____ pounds; _____ players, or _____% of the players, fall in that interval.

"Within two standard deviation of the mean" would be between _____ and _____ pounds; _____ players, or _____% of the players, fall in that interval.

REMEMBER: it is not enough to "crunch" numbers and record results - you will have to understand and be able to interpret in writing what these numbers tell us.