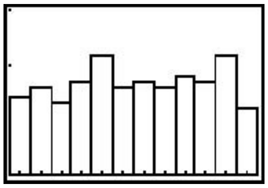
CHAPTER

2 Exploring Distributions

Calculator Note 2A: Generating a Distribution of Random Numbers

The TI-83 Plus and TI-84 Plus have two commands that generate approximately uniform distributions. Press MATH and arrow over to **PRB**. The command **rand(***n***)** randomly selects *n* numbers from the open interval (0, 1), and **randInt(***start, end, n***)** selects *n* integers from the closed interval [*start, end*]. For example, if you enter **randInt(1,12,200)→L1** into the Home screen, a histogram of list **L1** will look something like the screen shown here. This particular example could represent the number of births per month for a sample of size 200. (For more information about histograms, see Calculator Note 2C.)



[0.5, 12.5, 1, 0, 30, 10]

Calculator Note 2B: Graphing a Normal Distribution

normalpdf(

To graph a normal curve, go to the Y= screen and define a function in the form **Y=normalpdf(X,** *mean, standard deviation*). You find the **normalpdf(** command by pressing 2ND [**DISTR**] and selecting **1:normalpdf(** from the **DISTR** menu. If you do not specify the mean and standard deviation, the calculator assumes they are 0 and 1. The suffix “pdf” stands for *probability density function*. Using **normalpdf(** as the definition of a function provides the *y*-coordinates of the normal curve.

There is no zoom command that gives a “friendly” graph of a normal curve. Instead, use these guidelines to set the Window screen:

Xmin $= mean - 3SD$

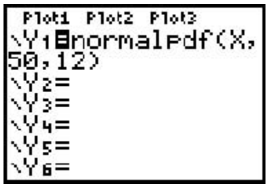
Xmax $= mean + 3SD$

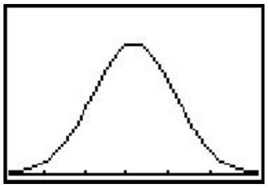
Xscl $= SD$

Ymin $= 0$

Ymax $= \frac{1}{2}\,SD$

Yscl $= 0$





[14, 86, 12, 0, 0.042, 0]

If you are interested in graphing more than one normal curve to show, for example, the effect of changes in standard deviation, you can enter two separate functions into the Y= screen. Or you can enclose the standard deviations in braces and use a single function. In either case, use the following "friendly" window:

Xmin = *mean* − 3(*Largest SD*)

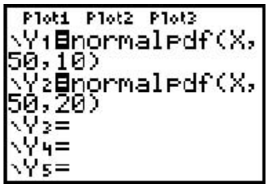
Xmax = *mean* + 3(*Largest SD*)

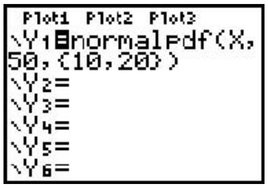
Xscl = *Largest SD*

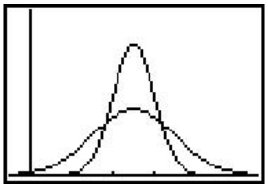
Ymin = 0

Ymax = $\frac{1}{2}$ *Smallest SD*

Yscl = 0







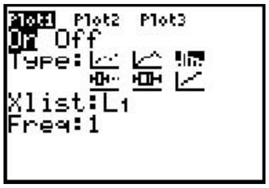
[−10, 110, 20, 0, 0.05, 0]

Calculator Note 2C: Histograms

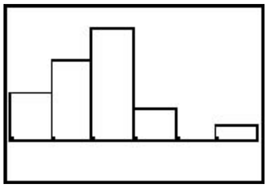
The TI-83 Plus and TI-84 Plus display frequency histograms and allow control over the attributes of the plot via the Window screen.

First, enter the data values into a list, say, list L1. If you have frequencies associated with the data values, enter the frequencies into another list, say, list L2. (See Calculator Note 0B for information about entering data into lists.) The examples that follow use the speeds of mammals from Display 2.24 on page 43 of the student book. The data were entered in list L1.

Second, go to the Stat Plot screen, 2ND [STAT PLOT], and define a histogram. Histograms are the third option under **Type**.

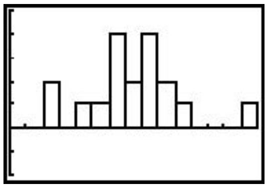


If you press ZOOM and select **9:ZoomStat** from the **ZOOM** menu, the histogram will fill the Graph window. The bars, however, may have unusual widths and dividing lines.

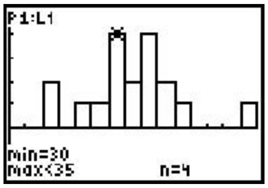


[11, 81.8, 11.8, −2.10, 8.19, 10]

The width of the bars is determined by **Xscl**, starting at **Xmin**. To change the width of the bars, reset the Window screen. For example, the window [0, 75, 5, −2, 5, 1] starts the first bar at 0 and starts a new bar every 5 units. As the bar width changes and data values are redistributed, you may need to adjust **Ymax**, too.



[0, 75, 5, −2, 5, 1]



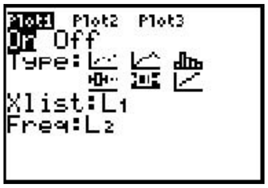
[0, 75, 5, −2, 5, 1]

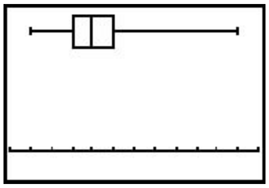
When you trace the histogram, you see the lower and upper bounds of each bar and the number of data values (the frequency) of each bar. Note that a value that falls at the dividing line between two bars is put in the bar on the right.

Unfortunately, the TI-83 Plus and TI-84 Plus do not provide relative frequency histograms. One way to work around this is to use a list to divide each frequency by the total number of data values. For example, enter your data values into list **L_1** and enter your frequencies into list **L_2**. Then arrow up and right to highlight the name of list **L_3** and enter **L_2/dim(L_2)**. Find the **dim(** command by pressing 2ND [**LIST**], arrowing over to **OPS**, and selecting **3:dim(**. (*Note:* If you enter the expression in quotation marks, ALPHA [“], the definition will be dynamic and the values will update whenever list **L_2** changes.) Then create a histogram using lists **L_1** and **L_3**. Remember, however, that the shape of the relative frequency histogram is identical to that of the frequency histogram.

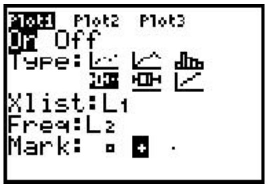
Calculator Note 2D: Boxplots, Outliers, and Five-Number Summaries

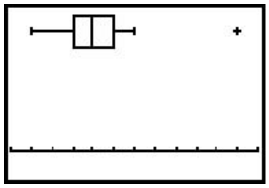
The TI-83 Plus and TI-84 Plus provide two types of boxplots: regular and modified. The regular boxplot (the fifth option for **Type** within the Stat Plot screen) does not indicate outliers, whereas the modified boxplot (fourth option) does. As with histograms, you specify a list for frequencies if your data are contained in a frequency table.





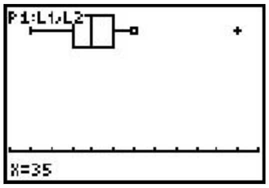
[5, 65, 5, −1, 6, 0]





[5, 65, 5, −1, 6, 0]

Pressing TRACE when a boxplot is displayed traces the values of the five-number summary. For modified boxplots, you can also trace the values of the whiskers and the outliers.



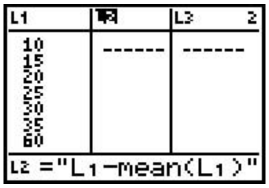
[5, 65, 5, −1, 6, 0]

Calculator Note 2E: Calculating the Standard Deviation Step by Step

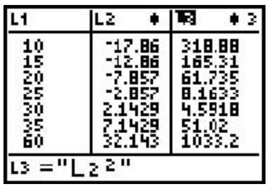
Many calculators and computers have built-in functions for calculating the standard deviation. Nonetheless, performing the calculations by hand may help you understand the meaning in the formula. The TI-83 Plus and TI-84 Plus can support hand calculations with the spreadsheet capabilities of the List Editor.

a. Enter the data into list L1.

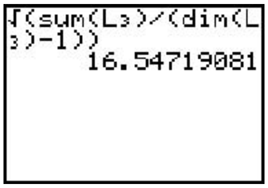
b. Define list L2 to be the deviations, $x - \bar{x}$. You do this with either the expression L1−**mean(**L1) or L1−**sum(**L1)/**dim(**L1). Find **mean(** and **sum(** by pressing 2ND [LIST] and arrowing over to MATH. Find **dim(** by pressing 2ND [LIST] and arrowing over to OPS. (*Note:* If you enter the expression in quotation marks, ALPHA ["], the definition will be dynamic and the values will update whenever list L1 changes.)



c. Define list L3 to be the squared deviations, $(x - \bar{x})^2$. Use the expression $L2^2$.

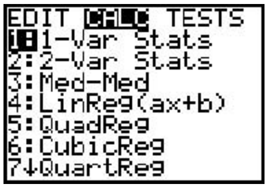


d. Complete the calculations on the Home screen. The expression $\sqrt{(\text{sum}(L_3)/(\text{dim}(L_3)-1))}$ calculates the sum of the square deviations, divides by 1 less than the number of data values, and takes the square root.

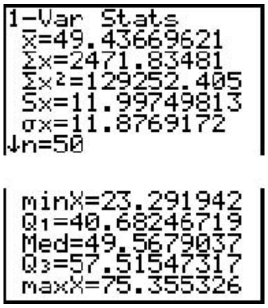


Calculator Note 2F: Summary Statistics **1-Var Stats**

Using the **1-VarStats** command, you can calculate a variety of summary statistics for any data set stored in a list. This command is found by pressing [STAT], arrowing over to **CALC**, and selecting **1:1-VarStats**. The summary statistics include mean, median, standard deviation, and quartiles.



In general, you enter **1-Var Stats** into the Home screen followed by the name of the list.



Computing Summary Statistics for Data in a Frequency Table

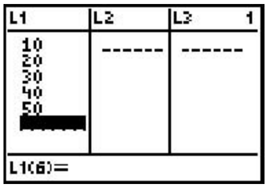
If your data set is contained in a frequency table, you enter the names of two lists, separated by a comma—first, the list that contains the data values and, second, the list that contains the frequencies.

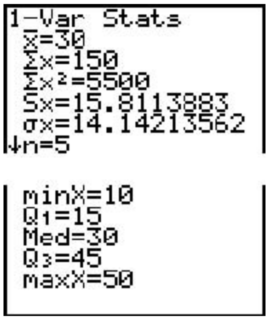
Please note that the mean is listed as $\bar{x}$. The calculator does not distinguish between sample mean and population mean. The calculator does, however, provide two standard deviations. **Sx** is the sample standard deviation, calculated with division by $(n - 1)$. **σx** is the population standard deviation, calculated with division by n.

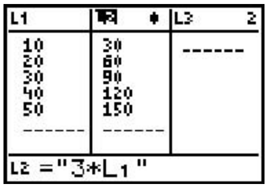
Note also that the complete five-number summary is displayed on the lower portion of the **1-Var Stats** screen.

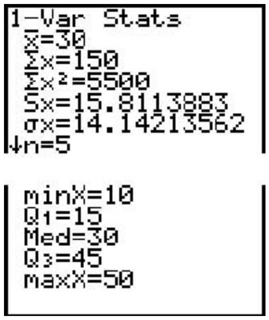
Calculator Note 2G: Exploring the Effects of Recentering or Rescaling

The **1-Var Stats** command makes it relatively easy to explore and recall the effects of recentering or rescaling. For example, to determine the effect of tripling each data value in any data set, first enter a small, hypothetical data set into list **L1**. Calculate **1-Var Stats** for the original data set. Then define list **L2** as the triple of each value, **3*L1**, and calculate **1-Var Stats** for the rescaled data. You can see which of the summary statistics are likewise tripled.





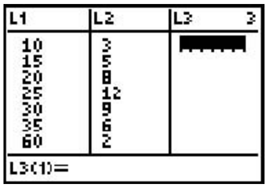




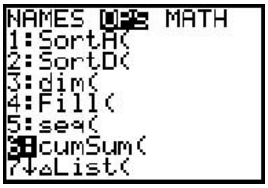
Calculator Note 2H: Cumulative Frequency Plots

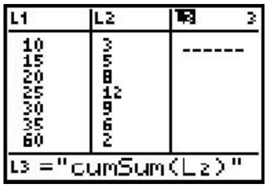
The TI-83 Plus and TI-84 Plus can construct a cumulative frequency plot using this procedure:

a. Enter data values and frequencies into lists **L1** and **L2**, respectively.

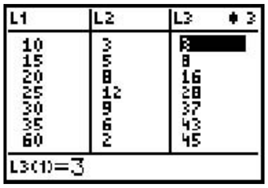


b. Arrow up and right to highlight the name of list **L3**. Enter **cumSum(L2)**. The **cumSum(** command is found by pressing 2ND [**LIST**], arrowing over to **OPS**, and selecting **6:cumSum(**. (*Note:* If you enter the expression in quotation marks, ALPHA ["], the definition will be dynamic and the values will update whenever list **L2** changes.)

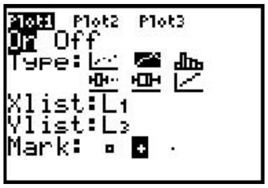




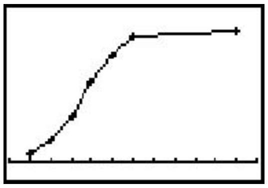
c. Press ENTER to have list L3 calculate the cumulative sums.



d. Press 2ND [**STAT PLOT**] to define an xyline plot that is a cumulative frequency plot. Use lists L1 and L3.



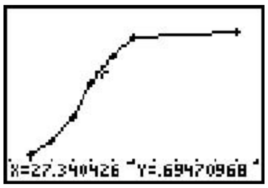
e. Press ZOOM and select **9:ZoomStat** from the **ZOOM** menu to display the cumulative frequency plot.



[5, 65, 5, −4.14, 52.14, 0]

Cumulative Relative Frequency Plots and Percentiles

A cumulative relative frequency plot can be constructed by entering **cumSum(L2)/sum(L2)** as the definition of list L3. Find the **sum(** command by pressing 2ND [**LIST**], arrowing over to **MATH**, and selecting **5:sum(**. With a cumulative relative frequency plot, you can move the cursor to identify the data value for any percentile. For example, to find the value that is at the 70th percentile, move the cursor to the point on the graph whose y-coordinate is approximately 0.70. The x-coordinate is the data value for the 70th percentile.

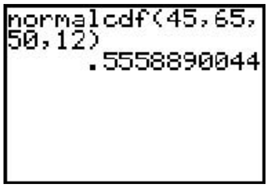


[5, 65, 5, −0.92, 1.159, 0]

Calculator Note 2I: Finding the Proportion of Values Under a Normal Curve on a Given Interval **normalcdf(**

Normal Cumulative Distribution Function

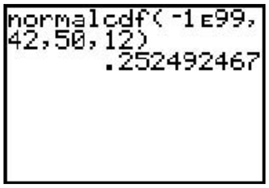
On the Home screen, a command in the form **normalcdf(**x_1, x_2, *mean*, *standard deviation*) returns the area under a normal curve on the interval $[x_1, x_2]$ for the normal distribution specified by *mean* and *standard deviation*. Find the **normalcdf(** command by pressing [2ND] [DISTR] and selecting **2:normalcdf(** from the DISTR menu. The suffix "cdf" stands for *cumulative distribution function*. Technically, a cumulative distribution function returns the percentage of area under a continuous distribution curve from negative infinity to the value of interest. The **normalcdf(** command, however, allows the lower bound of the interval to be specified as any value. For example, to find the percentage of area below a normal curve with mean 50 and standard deviation 12 over the interval [45, 65], enter **normalcdf(45,65,50,12)** into the Home screen.

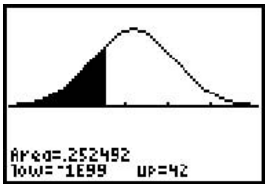


If the interval of interest is bounded by z-scores, then the **normalcdf(** command can be utilized by specifying mean 0 and standard deviation 1. If the command is written with no mean and standard deviation specified, the calculator defaults to the standard normal distribution and assumes the interval is bounded by z-scores. For example, **normalcdf(−1,1)** calculates the area under the standard normal curve between z-scores of −1 and 1.

Using Infinity as a Bound

If one of the bounds of the interval is infinite, use 1×10^{99} or -1×10^{99} to represent positive or negative infinity, respectively. You enter scientific notation using [2ND] [EE] to place E between the coefficient and the exponent on 10. You can use **1E99** and **−1E99** with both **normalcdf(** and **ShadeNorm(**. For example, to find the area below a normal curve with mean 50 and standard deviation 12 over the interval [−∞, 42], use **normalcdf(−1E99,42,50,12)** or **ShadeNorm(−1E99,42,50,12)**.

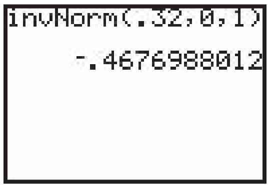




[14, 86, 12, −0.03, 0.042, 0]

Calculator Note 2J: Finding a *z*-Score with a Given Proportion of Values Below It **invNorm(**

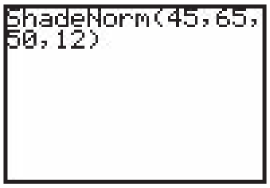
In order to find a *z*-score, the calculator includes the command **invNorm(**, which is found by pressing [2ND] **[DISTR]** and selecting **3:invNorm(** from the **DISTR** menu. This command returns the value of *z* for the specified area to the left of *z*. The command is entered in the form **invNorm(***area, mean, standard deviation***)**.
For example, to find the *z*-score that corresponds to a percentage of 0.32 (left tail area) under the standard normal curve, enter **invNorm(.32,0,1)** into the Home screen. (Because the default mean and standard deviation are 0 and 1, respectively, you could also enter **invNorm(.32)**.)

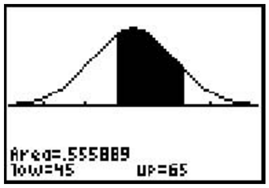


(*Note:* The **invNorm(** command can give the *z*-score that corresponds to the specified left tail area for *any* normal distribution, not just the standard normal distribution.)

Calculator Note 2K: Shading Under a Normal Curve for a Given Interval **ShadeNorm(**

To shade the area on a particular interval under the graph of a normal curve, use the **ShadeNorm(** command. Find this by pressing [2ND] **[DISTR]**, arrowing over to **DRAW**, and selecting **1:ShadeNorm(**. You enter **ShadeNorm(** into the Home screen.





[14, 86, 12, −0.03, 0.042, 0]

Note that the **ShadeNorm(** command does not adjust the window settings. Use these guidelines to set the Window screen.

Xmin $= mean - 3SD$

Xmax $= mean + 3SD$

Xscl $= SD$

Ymin $= -0.03$

Ymax $= \frac{1}{2}SD$

Yscl $= 0$

(*Note:* If you use **ShadeNorm(** repeatedly, press [2ND] **[DRAW]** and select **1:ClrDraw** from the **Draw** menu to clear the drawing after each use.)