CHAPTER 5 SAMPLING DISTRIBUTIONS

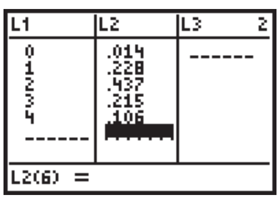
Section 5.1 ▪ Sampling from a Population

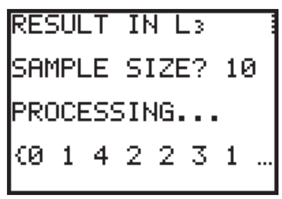
Sampling with Random Digits

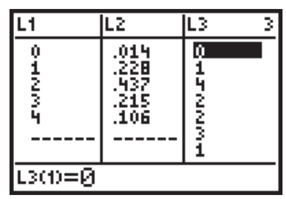
The opening discussion in this section, regarding the assignment of random digits, applies to random digit tables as well as the calculator. For example, to use Display 5.2 on page 269 of the student text, you can repeatedly use **randInt(0,9,3)** to create triples of random digits. You could also use rand, which generates a random number between 0 and 1, and use the first three digits of the result. Recall that you find the random number generators in the [MATH] **PRB** submenu.

Sampling with a Program

You can use the program **SAMPLE**, which selects **N** values at random from a distribution given in a relative frequency table. Before running the program, enter the values of *x* into list **L1** and enter the probabilities associated with the values of the variable *x* (PF) or the cumulative probabilities up to and including each value of *x* (CDF) into list **L2**. Run the program by pressing [PRGM] **EXEC**, choosing **SAMPLE**, and pressing [ENTER]. At the prompt, enter the sample size, **N**, and press [ENTER]. The program stores the random sample in list **L3** and pauses to display **L3** on the screen. You can use the right and left arrow keys to scroll through the list. Press [ENTER] to end the program.



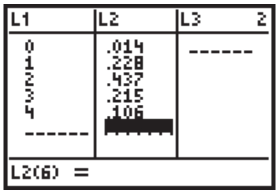





```
PROGRAM:SAMPLE
ClrHome
Disp "SAMPLING"
Disp ""
Disp "VALUES IN L1"
Disp "PF OR CDF IN L2"
Disp "RESULT IN L3"
DelVar LCDF
SetUpEditor L1,L2,L3
ClrList L3
If sum(L2)=1
Then
cumSum(L2)→LCDF
Else
L2→LCDF
End
Disp ""
Input "SAMPLE SIZE? ",N
Disp ""
Disp "PROCESSING..."
dim(L1)→M
For(I,1,N,1)
rand→X
For(J,1,M,1)
If X≤LCDF(M-J+1)
Then
L1(M-J+1)→L3(I)
End
End
End
DelVar LCDF
Disp ""
Pause L3
Stop
```

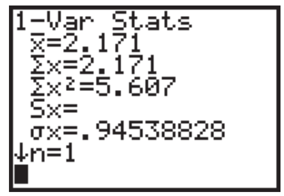
(continued)

The Mean and Standard Deviation of a Population



The TI-83 and TI-83 Plus can calculate the mean and standard deviation for probability distributions as well as for data distributions. Data values should be entered into one list, say, list **L1**, and relative frequencies should be entered into a second list, say, list **L2**. For example, here are the data for the number of motor vehicles from Display 5.1 on page 268 of the student text.

Execute **1-Var Stats** for these lists.



Notice these three characteristics of the output:

a. The value **x̄=2.171** is the mean of the distribution, although the calculator does not label it μ.

b. The sample standard deviation is blank. The calculator assumes that this is a population calculation because the relative frequencies sum to 1.

c. The value $n = 1$ indicates the sum of the relative frequencies.

Section 5.2 ▪ Generating Sampling Distributions

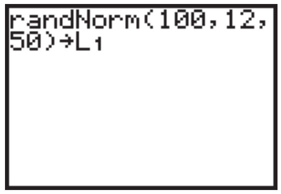
Many statistics computer programs efficiently perform sampling from data sets and offer the option of sampling with and without replacement. Parallel programs for the TI-83 and TI-83 Plus can be written and executed but frequently without the efficiency and control of statistics computer programs.

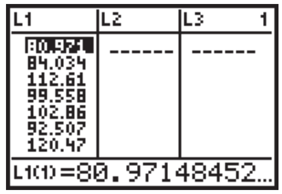
Below is a procedure that generated a sampling distribution without requiring a program. This example uses the **randNorm(** command to create samples from a normally distributed population. You can also use the **randInt(** command or the **randBin(** command to create three uniformly distributed populations or binomially distributed populations, respectively. All three commands are under the [MATH] **PRB** submenu.

This example takes 50 random samples of size 5 from a normal distribution with mean 100 and standard deviation 12. Then the sample means are calculated and displayed.

(continued)

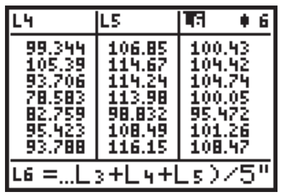
a. Enter **randNorm(100,12,50)→L1** to load 50 randomly selected numbers from the population into list **L1**.





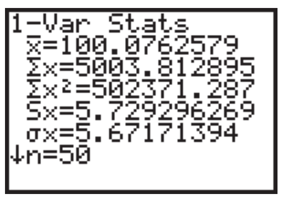
b. Repeat step a for lists **L2**, **L3**, **L4**, and **L5**.

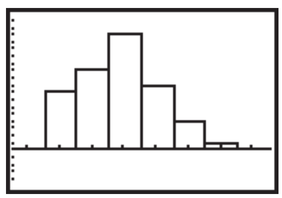
c. Each row of the List Editor screen constitutes a random sample of size 5. Define list **L6** as the sum of the rows divided by 5.



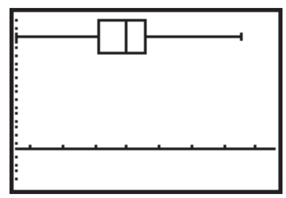
List **L6** now contains 50 sample means from a normal population with mean 100 and standard deviation 12.

d. Analyze the sampling distribution in list **L6** both numerically and graphically.





[84.89, 121.49, 4.58, −4.81, 18.72, 1]



[84.89, 121.49, 4.58, −4.81, 18.72, 1]

Needless to say, a greater number of samples would produce a more accurate picture of the sampling distribution.

Section 5.3 ▪ Sampling Distribution of the Sample Mean

The Central Limit Theorem

In order to demonstrate the Central Limit Theorem for a Sample Mean, you can use a calculator procedure similar to the one described in the section above. Look back at the 1-Var Stats screen for the example. Notice that, for a population that was normally distributed with mean 100 and standard deviation 12, 50 samples gave $\mu_{\bar{x}} = 100.076$ and $\sigma_{\bar{x}} = 5.729$.

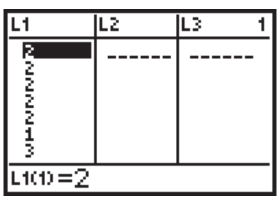
(continued)

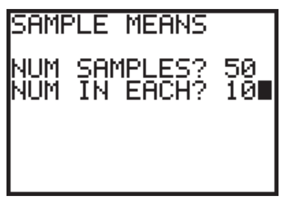
These values are consistent with the values dictated by the Central Limit Theorem, $\mu_{\bar{x}} = \mu = 100$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 5.367$. The slight differences can be attributed to the fact that only 50 samples were considered.

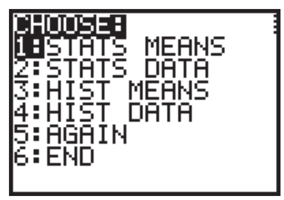
To demonstrate the Central Limit Theorem for the Sample Mean from a skewed population, you can perform the same procedure on a highly skewed population, such as a binomial distribution with $p = 0.20$.

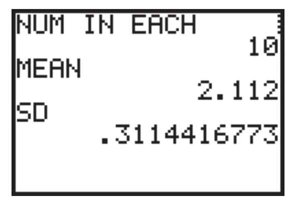
As an alternative to the sampling procedure, you can use the program SAMPMEAN, which selects random samples (with replacement) from a population stored in list L1, displays a sampling distribution, and calculates the mean of the sampling distribution and the standard error of the mean. Although the program may be slow, it is valuable for demonstrating the Central Limit Theorem for both normal and non-normal distributions.

Before running the program, enter the population data into list L1. Run the program by pressing [PRGM] EXEC, choosing SAMPMEAN, and pressing [ENTER]. At the prompts, enter the number of samples, S, and the sample size, N. Press [ENTER] after each value. The program collects samples and stores the sample means in list L2. A menu will appear from which you can view the statistics for the sampling distribution (mean and standard error), the statistics for the population data (mean and standard deviation), a histogram of the sampling distribution, or a histogram of the population. After viewing any of the statistics or histograms, press [ENTER] to return to the menu. To collect a new sampling distribution, choose 5:AGAIN. To end the program, choose 6:END.









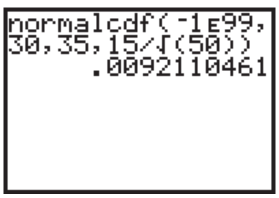

```
PROGRAM:SAMPMEAN
PlotsOff :FnOff :ClrHome
Disp "SAMPLE MEANS"
Disp ""
Lbl C
ClrList L2,L3
1-Var Stats L1
Input "NUM SAMPLES? ",S
Input "NUM IN EACH? ",N
For(J,1,S,1)
For(I,1,N,1)
L1(randInt(1,dim(L1)))→L3(I)
End
mean(L3)→L2(J)
ClrList L3
End
min(L1)-2→Xmin:max(L1)+2→
      Xmax:(max(L1)-
      min(L1))/10→Xscl
Lbl E
ClrHome
Menu("CHOOSE:","STATS
      MEANS",A,"STATS
      DATA",F,"HIST
      MEANS",B,"HIST
      DATA",G,"AGAIN",C,
      "END",D)
Lbl A
ClrHome
Disp "NUM IN EACH",N
Disp "MEAN",mean(L2)
```

(continued)

(PROGRAM: SAMPMEAN continued)

```
Disp "SD",stdDev(L2)
Pause
Goto E
Lbl B
PlotsOff
Plot1(Histogram,L2)
S/2→Ymax:-
      .5→Ymin:1→Yscl:Xscl/2→
      Xscl
DispGraph
Pause
Xscl*2→Xscl
Goto E
Lbl F
ClrHome
Disp "DATA"
Disp "MEAN",x̄
Disp "SD",σx
Pause
Goto E
Lbl G
PlotsOff
Plot1(Histogram,L1)
dim(L1)/2→Ymax:-
      .5→Ymin:1→Yscl
DispGraph
Pause
Goto E
Lbl D
```

Finding Probabilities for Sample Means
Finding Probabilities for Sample Totals

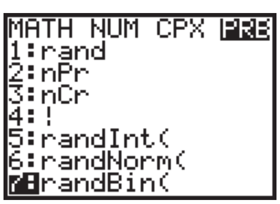


Because the Central Limit Theorem applies to sampling distributions of sample means and sample totals, you know that the respective sampling distributions are approximately normal. Therefore, the **normalcdf(** command can be used to calculate probabilities for each statistic. For example, if you were to conduct random samples of size 50 from a large population that is approximately normal with mean 35 and standard deviation 15, then the probability of a sample mean less than or equal to 30 is found with **normalcdf(−1E99,30,35,15/√(50))**.

Section 5.4 ▪ Sampling Distribution of the Sample Proportion

Activity 5.3 Buckle Up!

On the TI-83 and TI-83 Plus, the random binomial command, **randBin(**, generates a random integer from a specified binomial distribution. A binomial distribution counts the number of successes for a success-or-failure probability experiment, so you can use **randBin(** to create sampling distributions for this activity. (You will learn more about binomial distributions in Chapter 7 of the student text.)

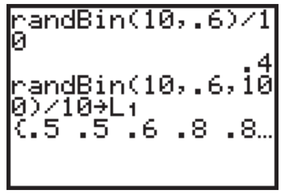


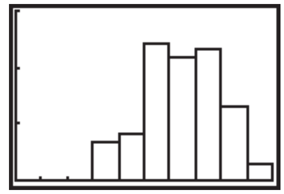
You find the **randBin(** command by pressing [MATH] **PRB 7:randBin(**. The syntax of the command is **randBin(***sample size, probability of success, number of samples***)**. If the number of samples is not specified, the calculator assumes 1.

For example, to select a random sample of size 10 and count the number of successes, enter **randBin(10,.6)** into the Home screen. To calculate the sample proportion, divide the result by 10, or **randBin(10,.6)/10**. To do 100 samples and

(continued)

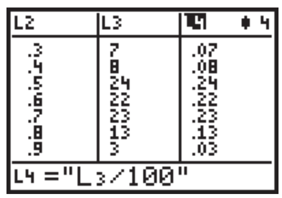
store the sample proportions into list **L1**, enter **randBin(10,.6,100)/10_L1**. You can now create a histogram of list **L1**. When setting the window, use **Xscl** = 1/*sample size.*

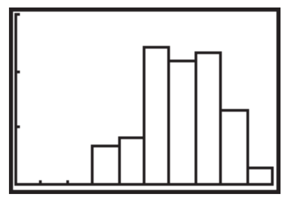




[0, 1, 0.1, 0, 30, 10]

If you want to create a relative frequency histogram, similar to those shown at the bottom of page 301 in the student text, run the **FREQTABL** program (see the "Programs" in the Introduction), which puts data values in list **L2** and frequencies in list **L3**. Then define list **L4** with the expression **L3/100**. Now create a relative frequency histogram using **L2** for **Xlist** and **L4** for **Freq**. Be sure to adjust **Ymin**, **Ymax**, and **Yscl**.





[0, 1, 0.1, 0, 0.3, 0.1]

Follow a similar procedure to do sampling distributions for $n = 20$ and $n = 40$. Change the value of **Xscl** accordingly. If you use relative frequency histograms, you'll need to run **FREQTABL** for each new distribution, and redefine **L4** (unless you use a dynamic definition).

Center and Spread for Sample Proportions

The Central Limit Theorem for a Sample Proportion indicates that the sampling distribution for the sample proportion and the sample number of successes are approximately normal. Therefore, you can use the **normalcdf(** command to calculate probabilities in either context. Be careful to use the correct mean and standard error values.