

Appendix

Sampling Distributions

The first section of this chapter introduces the idea central to inference procedures to come; namely, when we have a sample from a population and want to estimate a population parameter, we need to understand what might happen to the sample statistics (like the sample mean or sample proportion or any other statistic of interest) if another sample were taken. Because different individuals would certainly be in the new sample, we would get different statistics. We need to understand the idea of a sampling distribution, the probability distribution of all possible statistic values from samples of the same size from a population.

Calculating probabilities for the sample mean and the sample proportion employs the Normal distribution functions as detailed in Chapter 1. For the Normal approximation to be valid, be sure to check that the necessary conditions hold:

- The sample mean will **always** have a Normal distribution with mean μ and standard deviation σ/\sqrt{n} **if the population is Normal**. The **Normal distribution will be approximate** if the population has any distribution and the sample size is “large enough” (by the central limit theorem).
- The sample proportion will have an approximately Normal distribution with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$ if **both** $np \geq 10$ and $n(1-p) \geq 10$.

Your text discusses examining estimated sampling distributions using simulation. All of these technologies can create simulated distributions using random number generators, but not all can sample from a specified set. Also, some of the technologies require you to create all of the parts for the sampling distribution, while others include built-in modules that greatly simplify the process of creating the sampling distribution.

Recall that you will want to repeatedly obtain samples of size n . The number of samples (repetitions) is your decision; let’s call the number of repetitions k . The value of k can be 100, 1000, or something else, but don’t confuse k with n . The value n is the number of individuals in each sample.

The next step is to use these k samples of size n to find the statistic of interest. Here we are interested in the sample mean, \bar{x} . You will need to find the sample mean for each of the k samples. These k sample means are now your new data set that you want to explore because you have simulated the sampling distribution of the sample mean of samples of size n .

Section exercises give the command sequences for Excel, JMP, Minitab, and R. Below, we detail what can be accomplished with our other three technologies. All have some limitations.



SPSS cannot create repeated random samples from a specified list (its “random sampling” scheme is intended to work only as a means of specifying smaller samples from one very large one for further analysis). You can, however, simulate the sampling distribution of a sample proportion:

1. Page down and enter some value in the first column of an empty worksheet to correspond with the number of random numbers (samples) you wish to generate (this should be in row k). You may have to enter interim values to be able to go down as far as you want.
2. Click **Transform → Compute Variable**.
3. Enter a destination column name.
4. In the Function Group box, select **Random Numbers**.
5. Select **RV.Uniform** or **RV.Binom**.
6. Enter the minimum and maximum values for the uniform or n and p for the binomial. Separate the two parameters by a comma to replace the question marks.
7. Click **OK**.

Turn the observed number of successes per sample into sample proportions (replace “SimCol” below with your name and “n” with your numeric sample size):

8. Click **Transform → Compute Variable**.
9. Enter a destination column name.
10. Enter the function as SimCol/n in the **Numeric Expression** box.
11. Click **OK**.

For Exercise 6.11, generate two columns from the uniform distribution, then compute the means using a function like $(\text{SimCol1} + \text{SimCol2})/2$.

CRUNCH IT!

CrunchIt! cannot create random samples from a specified (integer) list, so it cannot be used to create a sampling distribution of the sample mean. It could be used to simulate the sampling distribution of a sample mean from the Uniform distribution on $[0, 1]$ or sample proportion.

1. Click **Insert → Random Numbers → Uniform** (or **Binomial**).
2. For the Uniform distribution, enter the minimum and maximum values for the uniform distribution and the number of samples, k .
For the binomial distribution enter the size for each sample, n , and the probability of success, p , and the number of samples, k .
3. Enter the number of samples, k .
4. Click **Sample**.

Exercise 6.11 asks you to simulate two uniform random variables and then calculate the mean of those samples of size two. Generate two columns of uniform random variables, then do the following to create a column of means.

1. Click **Insert → Evaluate Formula**.
2. Enter the formula such as
 $([\text{Var5}] + [\text{Var6}])/2$.
3. Click **Evaluate**.

For a proportion, you will have a column that contains the number of “successes” in each trial. Convert those to sample proportions that can be summarized or graphed using the following steps:

1. Click **Insert → Evaluate Formula**.
2. Enter a formula such as
 $[\text{Var5}]/100$.
3. Click **Evaluate**.

Note: CrunchIt! will default to placing a column of inserted values after the last column in the current worksheet. With a blank worksheet, that would be Var5. The square brackets around the variable name in the formula are important!



TI-83/-84

TI calculators can simulate random integers (equally likely), with or without repeats allowed, or observations from Normal and binomial distributions. They cannot sample from lists of specific values, nor can they simulate sample proportions. These options are from the **MATH** PRB menu.

1. Press **STAT****ENTER** to enter the statistics editor.
2. Place the cursor in the list header (so the input line looks like “L1=” instead of “L1(1)=”).
3. Press **MATH****4** to the PRB menu. Option 5 will generate random integers between two specified values, inclusive. Option 7 will generate observations from a binomial distribution with specified n and p . Select the appropriate distribution type. The command shell will transfer to the input line.
4. Enter the parameters separated by commas. For integers, these are the smallest and largest values possible; for the binomial, they are n and p .
5. Enter a comma and the number of values to generate into the list. Press **ENTER**.

Note: These calculators do not have sufficient processing power or memory to generate large numbers of random variables. The maximum number that can reasonably be generated (without taking a very long time) is about 250.