AP Statistics Name____________________

This worksheet provides instructions that allow you to carry out simulations using your calculator, to investigate the sampling distribution of sample proportions. The examples on pp. 1-2 walk you through one simulation (HW for Fri), and the exercises on p. 3 will involve carrying out two more (HW for Mon).

Example A:

A large state university has a student body consisting of 60% females and 40% males. Suppose you want to randomly select eighty students to participate in a survey regarding campus issues. Predict the answers to the following questions *before* going on to the Procedure section.

What proportion of the eighty students do you expect to be female? __________

How different from 60% might the proportion of females be?____________________

What would be an exceptionally large proportion of women in the sample of eighty? _________

What would be an exceptionally small proportion? ___________

Procedure:

The instruction `randBin(80,.6)`* will produce the number of women in a sample of eighty students chosen from the student body that is 60% female. Execute this instruction several times to see that the number of females varies, but the results are usually close to 48.

To obtain the proportion of females, execute `randBin(80,.6)/80`. Repeat this instruction several times to observe that sample proportions tend to be between .50 and .70.

To gain information about the distribution of the sample proportions, it is helpful to perform many simulations. Then the results can be summarized with graphs and summary statistics. Adding a third argument to the `randBin(` instruction causes the calculator to repeat the simulation the desired number of times. For example, to perform 100 of these simulations and store the resulting sample proportions in L1, execute `randBin(80,.6,100)/80 STO> L1`. This will take awhile; go and have a snack, and it should be done when you get back.

Now execute `1-Var Stats L1` to observe the summary statistics for your simulations. (Remember that every value in your list is a sample proportion.)

What is the mean value of the sample proportions? ___________

What is the minimum proportion value? _________

What is the maximum value? ________

What is the standard deviation of the sample proportions? _________

---

* Find the `randBin(` instruction by pressing `MATH`, then hit the right arrow 3 times to get to the `PRB` menu, and `randBin(` is option `7`.

Example B:
Describe the overall shape of the distribution of sample proportions generated in Example A.

Procedure:
A histogram provides the most detailed information about the shape of the distribution of sample proportions. Make a histogram of the data in `L1`, and use `ZOOM 9` to set an initial viewing window for your histogram. Tweak the window settings (`Xmin`, `Xmax`, `Xscl`) to be "nice"[†] numbers. Study the histogram to answer the following questions:

Are the minimum and maximum values set off from other values?________

Do you consider them to be outliers? __________

Describe the shape of the histogram:

Example C:
The histogram created above is likely to be approximately normal in appearance. Determine how well the sample proportions from the simulation satisfy the Empirical Rule. Is it reasonable to conclude that sample proportions are approximately normally distributed?

Procedure:
To compare simulation outcomes to the Empirical (68-95-99.7) Rule, you must first calculate the endpoints of the intervals. The first interval will extend from $\bar{x} - s_x$ to $\bar{x} + s_x$, the second from $\bar{x} - 2s_x$ to $\bar{x} - 2s_x$, and the third from $\bar{x} - 3s_x$ to $\bar{x} + 3s_x$. Then sort the sample proportions in ascending order using `SortA(L1)` and count the entries that fall within the relevant endpoints.

Record information about your simulation below.

Interval that corresponds to $\bar{x} \pm s$: __________________

Percentage of proportions within one standard deviation of the mean: ___________

Interval that corresponds to $\bar{x} \pm 2s$: _________________

Percentage of proportions within two standard deviations of the mean: __________

Interval that corresponds to $\bar{x} \pm 3s$: _________________

Percentage of proportions within three standard deviations of the mean: __________

How well does your sample adhere to the Empirical Rule?

[†] Try rounding `Xmin` down to the nearest tenth, `Xmax` up in the same manner, and `Xscl` to the nearest hundredth.

Exercises (Use extra paper where necessary)

1. Repeat the examples above (A, B, C) to investigate the distribution of sample proportions based on samples of size 100 and samples of size 200. Store the sample proportions resulting from samples of size 100 in `L2` and the proportions resulting from samples of size 200 in `L3`. (It will take several minutes to generate each of these samples.) What is the most obvious difference in the distributions? Record the mean and the standard deviations associated with each of these simulations.

   Mean of sample proportions from sample size 100: _____________

   Standard deviation of sample proportions from sample size 100: _____________

   Mean of sample proportions from sample size 200: _____________

   Standard deviation of sample proportions from sample size 200: _____________

2. Mathematical theory states that the sampling distribution of sample proportions is approximately normal and that the distribution becomes more like the normal as the sample size increases. Do your simulations seem consistent with this theory?

3. Create parallel boxplots of the sample proportions stored in `L1`, `L2`, and `L3`. (Use the first boxplot option, which shows outliers as separate points.) Sketch a copy on your extra paper. How do the medians compare? How do the spreads compare?

4. You should now have a standard deviation for simulations based on samples of size 80, 100 and 200. As the sample size increases, how does the standard deviation change? Does this fit what you expected, from our class discussions?

5. Mathematical theory states that the mean of the sampling distribution of sample proportions equals p, the proportion for the population from which the samples are drawn. The theory also states that the standard deviation of the sampling distribution is $\sqrt{\frac{p(1-p)}{n}}$, where n is the sample size and p is the population proportion. So in the examples above, the mean of the sample proportions should always be close to 0.6 and the standard deviation for a sample of size n should be approximately $\sqrt{\frac{(0.6)(0.4)}{n}}$. Use this formula to compute the standard deviation you would expect for simulations based on $n = 80$, $n = 100$, $n = 200$. Compare the standard deviations you recorded in Exercise 1 to the theoretical result.