Stat 243 Lab – Simulating the Central Limit Theorem

Background:
This lab will investigate the Central Limit Theorem using the data simulation feature of Minitab.

Part One (Simulating a die roll).
Begin by loading MINITAB. In the MINITAB worksheet, label column A1 as "Data".

The instructions below will show you how to use MINITAB to simulate a set of 600 data points. The simulation will replicate the process of

rolling a die and recording a 1 each time the die comes up ⚅
and recording a 0 if the die shows anything else. A data set like this is said to come from a "Bernoulli Distribution". Note that the chance of

getting a ⚅ on a single die roll is 1/6 which is approximately 0.16667, so roughly 1/6 of the 600 data points should have the value of 1.

From the top menu select "Calc > Random Data > Bernoulli".

- In the "Number of Rows of Data to Generate" box, input 600
- In the "Store in Column" box, select "C1 Data"
- In the "Event Probability" box, input 0.16667

Finally, click "OK".

There should now be data in the first 600 rows of column A1. The data should all be 0's and 1's. Each data point should have had a 1/6 chance to be a 1 and a 5/6 chance to be a 0.

Use MINITAB to create a histogram for your data set and include it in your lab report. Then answer the questions below and include them in your lab report also.

Question 1: What is the expected percentage of 1's in the sample data set? (Hint: You may want to review the "Sampling Distribution of $\hat{p}$" information from section 6.4 of the text).

Question 2: What is the standard deviation for the percentage of 1's in the sample data set? (Hint: As with Question 1, you may want to review the "Sampling Distribution of $\hat{p}$" information from section 6.4 of the text).

Question 3: What was the observed percentage of 1's that showed up in your sample data set?

Question 4: Is the number of 1's in your data set within two standard deviations of the expected value? Explain.

Question 5: Roughly what percentage of students doing this lab should have found that the number of 1's in their data set was within two standard deviations of the expected value? Explain.

Question 6: Does the histogram of the data set look like a normal curve?

Question 7: Considering what the Central Limit Theorem says, should you expect the histogram for this data set to look like a normal curve? Explain.

Part Two (Adding More Dice).

Clear the MINITAB worksheet of the data from Part One.

In Part Two of the lab, you will produce a new data set. This data set will also have 600 data points, but each data point will be the number of dice that show up ⚃ when three dice are rolled. (For instance, if dice show [⚃⚂⚃] then the recorded number would be 2 since two of the dice came up ⚃, and if the dice show [⚅⚄⚀] the recorded number would be 0.

Notice that this means that each entry is like the sum of three different Bernoulli results like you produced in Part One of the lab.

A data set like this is said to come from a “Binomial Distribution”.

From the top menu select “Calc > Random Data > Binomial”.

- In the “Number of Rows of Data to Generate” box, input 600
- In the “Store in Column” box, select “C1 Data”
- For the “Number of Trials”, select 3.
- In the “Event Probability” box, input 0.16667

Finally, click “OK”.

As with Part One, use Minitab to create a histogram for your data set and include it in your lab report. Then answer the questions below and include them in your lab report as well.

Question 8: What is the largest number that could theoretically show up in this data set?

Question 9: Did this theoretical maximum actually show up in the data set, and if so, how many times?

Question 10: What is the smallest number that could theoretically show up in this data set?

Question 11: Did this theoretical minimum actually show up in the data set, and if so, how many times?

Question 12: Explain why there is such a large difference between your answer to question 9 and your answer to question 11.

Question 13: What is the expected value for the average of this sample data set? (Hint: You may want to review the “Mean for a Binomial Random Variable” information from section 4.4 of the text).

Question 14: What is the standard deviation for the average of this sample data set? (Hint: You may want to review the “Standard Deviation for a Binomial Random Variable” information from section 4.4 of the text).

Question 15: Use MINITAB to find the actual average for your sample data set and record it in your lab report. Is your actual average within two standard deviations of the expected average? Explain.

Question 16: Does the histogram of the data set look like a normal curve?

Question 17: Considering what the Central Limit Theorem says, should you expect the histogram for this data set to look like a normal curve? Explain. (Hint: You may want to review the "Sampling Distribution of $\hat{p}$" information from section 6.4 of the text and calculate the value of $n\hat{p}$ for this sample).

Part Three (Adding Even More Dice).

Repeat Part Two of the lab, except this time modify the instructions so that each of your 600 data points represents the number of times a ⚅ is observed in 90 die rolls.

As in the prior parts of the lab, create a histogram of the resulting data and include it in your lab report. Then answer the following questions.

Question 18: What is the largest number that could theoretically show up in this data set?

Question 19: Did this theoretical maximum actually show up in the data set, and if so, how many times?

Question 20: What is the smallest number that could theoretically show up in this data set?

Question 21: Did this theoretical minimum actually show up in the data set, and if so, how many times?

Question 22: Explain your answers to questions 19 and 21 in the context of what would have to occur in dice rolls for these extreme values to be observed.

Question 23: What is the expected value for the average of this sample data set?

Question 24: What is the standard deviation for the average of this sample data set?

Question 25: Use MINITAB to find the actual average for your sample data set and record it in your lab report. Is your actual average within two standard deviations of the expected average? Explain.

Question 26: Does the histogram of the data set look like a normal curve?

Question 27: Considering what the Central Limit Theorem says, should you expect the histogram for this data set to look like a normal curve? Explain. (Hint: You may want to review the "Sampling Distribution of $\hat{p}$" information from section 6.4 of the text and calculate the value of $n\hat{p}$ for this sample).

Question 28: Consider the extreme values you identified in Questions 18 and 20. Determine how many standard deviations each of these extreme values are from average.