Use the link in the Jupyter Notebook activity to access your Python script. Once you have made your calculations, complete this discussion. The script will output answers to the questions given below. You must attach your Python script output as an HTML file and respond to the questions below.

In this discussion, you will apply the central limit theorem and use principles of the Normal distribution to calculate probabilities. You will demonstrate two key parts of the central limit theorem:

* The distribution of *sample means* is approximately Normally distributed (bell-shaped) as the sample size increases and we repeatedly draw these samples, regardless of the shape of the population distribution from which the samples are drawn.
* The average of all sample means is equal to the population mean. In practice, the average of all sample means will closely approximate the population mean.

You will generate a population data set representing total precipitation (TPCP) in tenths of a millimeter using Python's numpy module. The distribution of this data set will be skewed. This data set will be unique to each student, and therefore the answers will be unique as well. Run Step 1 in the Python script to generate your unique population data.

In your initial post, address the following items:

1. In the Python script, you created a histogram for the dataset generated in Step 1. Check to make sure that this data distribution is skewed and included in your attachment. See Step 2 in the Python script.
2. What is the mean of the TPCP population data? See Step 3 in the Python script.
3. In the Python script, you selected a random sample with replacement, of size 50 (note that this is a sufficiently large sample), from the TPCP population. What is the mean of your random sample? Does this sample mean closely approximate the TPCP population mean? See Step 4 in the Python script.
4. You also selected 1,000 random samples of size 50 and calculated the mean of each sample. Then you stored those means into a dataframe. Check to make sure the output of this step is in your attachment. See Step 5 in the Python script.
5. Review the plotted data distribution for these 1,000 means. Does this approximate a Normal distribution? Does this confirm the first part of the central limit theorem? Why or why not? See Step 6 in the Python script.
6. What is the "grand" mean and standard deviation of these 1,000 means? Does the grand mean closely approximate (on a relative basis) the mean of the original distribution? Does this confirm the second part of the central limit theorem? Why or why not? See Step 7 in the Python script.

Recall that your distribution of 1,000 means is Normally distributed even though the population distribution was skewed and the grand mean closely approximates the population mean. In your follow-up posts to other students, review your peers' results and provide some analysis and interpretation.

1. Is their population distribution skewed? Is their distribution of 1,000 sample means approximately Normally distributed? Does this confirm the first part of the central limit theorem? Why or why not?
2. Does *their* grand mean closely approximate *their* population mean? Does this confirm the second part of the central limit theorem? Why or why not?
3. Based on this discussion activity, what have you learned about the central limit theorem?

Remember to attach your Python output and respond to all questions in your initial and follow-up posts. Be sure to clearly communicate your ideas using appropriate terminology. Finally, be sure to review the [Discussion Rubric](https://learn.snhu.edu/d2l/common/dialogs/quickLink/quickLink.d2l?ou=1230325&type=content&rcode=snhu-702316) to understand how you will be graded on this assignment.

1. In the Python script, you created a histogram for the dataset generated in Step 1. Check to make sure that this data distribution is skewed and included in your attachment. See Step 2 in the Python script.

Chart, histogram

Description automatically generated

1. What is the mean of the TPCP population data? See Step 3 in the Python script.
   1. The population mean is 489.33
2. In the Python script, you selected a random sample with replacement, of size 50 (note that this is a sufficiently large sample), from the TPCP population. What is the mean of your random sample? Does this sample mean closely approximate the TPCP population mean? See Step 4 in the Python script.
   1. The mean of my random sample is 541.8
   2. There is a difference of 52.27 during the random sample and the population mean.
3. You also selected 1,000 random samples of size 50 and calculated the mean of each sample. Then you stored those means into a data frame. Check to make sure the output of this step is in your attachment. See Step 5 in the Python script.
4. Review the plotted data distribution for these 1,000 means.
   1. Does this approximate a normal distribution? Yes
   2. Does this confirm the first part of the central limit theorem? Yes
   3. Why or why not? Because you have a normal distribution, and the means align with the middle of the bar graph and the overall distribution.
5. What is the "grand" mean and standard deviation of these 1,000 means? The grand mean is 487.75, with a standard deviation of 49.82. Also giving a probability that a specific meaning is 450 or less is 0.2243.
   1. Does the grand mean closely approximate (on a relative basis) the mean of the original distribution? The original population mean was 489.33, with the sample mean of 541.8. The grand mean is closer to the original population mean with a deviation of 1.58.
   2. Does this confirm the second part of the central limit theorem? No it does not confirm the 2nd part, the numbers do not match equally to the standard deviation equal to a population standard deviation divided by the square root of sample size. Therefore, it does not confirm the 2nd deviation degree mean in the sample mean.

The mean of TPCP population data is 489.33. It represents the average number of individuals in a population. It can be calculated by adding up all the values in a dataset and then dividing the total by the number of values. Step 3 in the Python script shows how to calculate the mean.

In the Python script, a random sample of size 50 with replacement is selected from the TPCP population. The mean of the random sample is 541.8. There is a difference of 52.27 between the random sample mean, and the TPCP population mean. Step 4 in the Python script can be seen to check the approximation of the sample mean to the population mean.

In step 5 of the python script, 1,000 random samples of size 50 are selected from the TPCP population, and the mean of each sample is calculated and stored in a data frame. The output of this step is included in the attachment—the means in the data frame range from 485.64 to 516.92. You can check the step to ensure the output is correct and stored in the data frame.

The means

0 485.64

1 482.70

2 482.40

3 415.12

4 510.26

.. ...

995 473.68

996 437.44

997 498.02

998 433.46

999 516.92

In step 5 of the analysis, the plotted data distribution for the 1,000 means is reviewed. The distribution approximates a normal distribution, which confirms the first part of the central limit theorem. This is evident because the means align with the middle of the bar graph, creating a bell-shaped curve, which is a characteristic of a normal distribution. Additionally, the overall distribution also confirms this. Although this does not necessarily prove that the population from which the sample is drawn has a normal distribution, it shows that the sampling distribution of means of any large random samples from any people will be expected.