**CBA: Practice Problem Set 2**

**Topics: Sampling Distributions and Central Limit Theorem**

1. Examine the following normal Quantile plots carefully. Which of these plots indicates that the data …
2. Are nearly normal?

Ans:- C

1. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.)

Ans:- B & D

1. Are skewed (i.e. not symmetric) ?

Ans:- A,B & D

1. Have outliers on both sides of the center?

Ans:- A & B



1. For each of the following statements, indicate whether it is True/False. If false, explain why.

The manager of a warehouse monitors the volume of shipments made by the delivery team. The automated tracking system tracks every package as it moves through the facility. A sample of 25 packages is selected and weighed every day. Based on current contracts with customers, the weights should have *μ* = 22 lbs. and *σ* = 5 lbs.

1. Before using a normal model for the sampling distribution of the average package weights, the manager must confirm that weights of individual packages are normally distributed.

Ans- True

It's essential to confirm that the weights of individual packages are normally distributed before relying on a normal model for the sampling distribution of the average package weights. If the weights of individual packages are not normally distributed, the Central Limit Theorem (CLT) may not apply, and the assumptions for using a normal distribution for the sample mean may not be met. However, as mentioned earlier, for large sample sizes, the CLT suggests that the sampling distribution of the sample mean tends to be approximately normal regardless of the distribution of the individual observations, but it's still a good practice to confirm the normality of the individual observations whenever possible.

1. The standard error of the daily average SE() = 1.

Ans:- True

The standard error (SE) of the sample mean is calculated as the standard deviation of the population divided by the square root of the sample size (σ/√n). Given that the population standard deviation (σ) is 5 lbs and the sample size (n) is 25 packages, the standard error would be SE = σ/√n = 5/√25 = 5/5 = 1 lb. Therefore, this statement is True.

窗体顶端

窗体底端

1. Auditors at a small community bank randomly sample 100 withdrawal transactions made during the week at an ATM machine located near the bank’s main branch. Over the past 2 years, the average withdrawal amount has been $50 with a standard deviation of $40. Since audit investigations are typically expensive, the auditors decide to not initiate further investigations if the mean transaction amount of the sample is between $45 and $55. What is the probability that in any given week, there will be an investigation?
2. 1.25%
3. 2.5%
4. 10.55%
5. 21.1%
6. 50%

Ans:- [Input]

import scipy.stats as stats

mean = 50

std\_dev = 40

sample\_size = 100

lower\_limit = 45

upper\_limit = 55

z\_lower = (lower\_limit - mean) / (std\_dev / (sample\_size \*\* 0.5))

z\_upper = (upper\_limit - mean) / (std\_dev / (sample\_size \*\* 0.5))

probability = stats.norm.cdf(z\_upper) - stats.norm.cdf(z\_lower)

probability\_percentage = (1- probability) \* 100

print("Probability of investigation in any given week:", np.round((probability\_percentage),1) "%")

[Output] :- Probability of investigation in any given week: 21.1 %

1. The auditors from the above example would like to maintain the probability of investigation to 5%. Which of the following represents the minimum number transactions that they should sample if they do not want to change the thresholds of 45 and 55? Assume that the sample statistics remain unchanged.
2. 144
3. 150
4. 196
5. 250
6. Not enough information

Ans:- [C]

[Input]:-

std\_dev = 40

alpha = 0.05 # Probability of investigation

margin\_of\_error = (55 - 45) / 2

z\_alpha\_2 = stats.norm.ppf(1 - alpha/2)

required\_sample\_size = ((z\_alpha\_2 \* std\_dev) / margin\_of\_error) \*\* 2

print("Minimum number of transactions to sample:", round(required\_sample\_size))

[Output]:- 246

1. An educational startup that helps MBA aspirants write their essays is targeting individuals who have taken GMAT in 2012 and have expressed interest in applying to FT top 20 b-schools. There are 40000 such individuals with an average GMAT score of 720 and a standard deviation of 120. The scores are distributed between 650 and 790 with a very long and thin tail towards the higher end resulting in substantial skewness. Which of the following is likely to be true for randomly chosen samples of aspirants?
2. The standard deviation of the scores within any sample will be 120.
3. The standard deviation of the mean of across several samples will be 120.
4. The mean score in any sample will be 720.
5. The average of the mean across several samples will be 720.
6. The standard deviation of the mean across several samples will be 0.60

Ans:[ D]

[Input]:-

population\_mean = 720

population\_std\_dev = 120

sample\_size = 100

np.random.seed(42)

sample\_scores = np.random.normal(population\_mean, population\_std\_dev, sample\_size)

1. The standard deviation of the scores within any sample will be 120.

[input]:-

sample\_std\_dev = np.std(sample\_scores)

print("Standard deviation of scores within the sample:", np.round((sample\_std\_dev),2))

[Output]:- Standard deviation of scores within the sample: 108.43

[B]. The standard deviation of the mean across several samples will be 120.

[Input]:-

standard\_error = population\_std\_dev / np.sqrt(sample\_size)

print("Standard error of the mean across several samples:", standard\_error)

[Output]:- Standard error of the mean across several samples: 12.0

[C]. The mean score in any sample will be 720.

[Input]:-

sample\_mean = np.mean(sample\_scores)

print("Mean score in the sample:", np.round((sample\_mean),2)

[Output]:- Mean score in the sample 707.54

[D]. The average of the mean across several samples will be 720.

[Input]:-

num\_samples = 1000

sample\_means = [np.mean(np.random.normal(population\_mean, population\_std\_dev, sample\_size)) for \_ in range(num\_samples)]

average\_of\_means = np.mean(sample\_means)

print("Average of the mean across several samples:", average\_of\_means)

[Output]:- Average of the mean across several samples: 720.13

[E]. The standard deviation of the mean across several samples will be 0.60.

[Input]:-

print("Standard deviation of the mean across several samples:", standard\_error)

[Output]:- Standard deviation of the mean across several samples 12.0