**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?
3. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.)
4. Are skewed (i.e. not symmetric) ?
5. Have outliers on both sides of the center?



Ans = 1. the data are nearly normally distributed

1. This gap suggests the presence of two distinct modes or groups within the data, indicating a bimodal distribution.
2. it suggests skewness. A positively skewed distribution will have a long right tail, while a negatively skewed distribution will have a long left tail.
3. No outliers on both side.
4. 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.
2. The standard error of the daily average SE() = 1.
3. Ans = **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.**
   * **True.** This statement is true. Before assuming that the sampling distribution of the average package weights is approximately normal (which is a common assumption due to the Central Limit Theorem), it's important to confirm that the weights of individual packages are normally distributed. If the individual weights are not normally distributed, it could affect the validity of statistical analyses and the use of normal distribution-based methods. Departures from normality in the individual weights can lead to inaccurate results.
4. **The standard error of the daily average SE(x̅) = 1.**
   * **True.** This statement is true based on the information provided in your initial question. You mentioned that the population standard deviation (σ) is 5 lbs., and the sample size (n) is 25 packages. To calculate the standard error of the sample mean (SE(x̅)), you use the formula SE(x̅) = σ / √n, which in this case is SE(x̅) = 5 / √25 = 5 / 5 = 1 lb. So, the statement is correct, and the standard error of the daily average is indeed 1 lb. This standard error quantifies the precision of the sample mean as an estimate of the population mean (μ).
5. 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?
6. 1.25%
7. 2.5%
8. 10.55%
9. 21.1%
10. 50%

Ans = calculate the standard error of the sample mean (SE(x̅)) using the given population standard deviation (σ) and sample size (n):

SE(x̅) = σ / √n SE(x̅) = $40 / √100 SE(x̅) = $4

Now, we can use the Z-score formula to find the Z-scores for $45 and $55:

Z($45) = ($45 - $50) / $4 = -5 / $4 = -1.25 Z($55) = ($55 - $50) / $4 = 5 / $4 = 1.25

Next, we'll use a standard normal distribution table or calculator to find the probabilities associated with these Z-scores. We're looking for the probability that the Z-score is less than -1.25 or greater than 1.25 (because these correspond to being outside the range $45 to $55).

Using a standard normal distribution table, you can find that the probability of Z < -1.25 or Z > 1.25 is approximately 2 \* 10.55% = 21.1%.

So, the probability that in any given week there will be an investigation is approximately 21.1%.

Therefore, the correct answer is **D. 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 = 250

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

* **False.** This statement is not necessarily true. The standard deviation within any sample will depend on the specific individuals chosen in that sample. While the population has a standard deviation of 120, individual samples may have different standard deviations based on the randomness of the selection process.

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

* **False.** The standard deviation of the sample means (also known as the standard error of the mean) across several samples is calculated as the population standard deviation divided by the square root of the sample size (assuming random sampling). In this case, the standard error of the mean is:

SE(x̅) = σ / √n SE(x̅) = 120 / √n

So, the standard deviation of the sample means will decrease as the sample size (n) increases, but it will not remain fixed at 120.

C. **The mean score in any sample will be 720.**

* **False.** While the population mean is 720, the mean score in any specific sample may vary due to random sampling. It is not guaranteed to be exactly 720 in every sample.

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

* **True.** This statement is true. The average of the sample means across several samples (the sampling distribution of the sample mean) is expected to be equal to the population mean. In this case, the average of the sample means will be 720.

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

* **False.** The standard deviation of the sample means (standard error of the mean) across several samples is not 0.60; it depends on the sample size and the population standard deviation, as mentioned in statement B. It will vary depending on the sample size used for each calculation.

So, among the given options, only statement D is likely to be true for randomly chosen samples of aspirants.