**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?



I. Ans -> C  
II. Ans -> B and D  
III. Ans -> A, B and D  
IV. Ans -> A and 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.
2. The standard error of the daily average SE() = 1.

i. The statement is false because it is not necessary to confirm that the weights of individual packages are normally distributed in order to use a normal model for the sampling distribution of the average package weights. The Central Limit Theorem (CLT) states that, under certain conditions, the sampling distribution of the sample mean (x̅) becomes approximately normally distributed, even if the population from which the samples are drawn is not normally distributed. The key conditions for the CLT are that the sample size should be sufficiently large (typically n > 30 is considered sufficient), and the samples should be drawn randomly and independently from the population. As long as these conditions are met, the manager can rely on the normal model for the sampling distribution of the average package weights.

ii. The statement is false because The statement is not necessarily true, and it depends on the sample size (n) and the population standard deviation (σ). The formula for the standard error of the sample mean SE() is

SE() = \frac(*σ*) (sqrt(n))

SE() = \frac(5) (sqrt(25)) = 1

So the statements is True if the conditions provided are met, but it depends on the specific values of *σ* and n.

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%

To calculate the probability of an investigation, you need to determine the probability that the mean transaction amount of the sample falls outside the range of $45 to $55.

The standard error of the mean (SEM) can be calculated as:

SEM = Standard Deviation/ sqrt (Sample size) = 4 / sqrt (100) = 4

Now, we have to find the probability that the sample mean falls outside the range of $45 to $55. To do this, we can use the z-score formula:

Z = X – Population Mean / SEM

X is the sample mean you want to test (in this case, $45 and $55).

The population mean is $50.

The SEM is 4.

For Z45 = 45 – 50 / 4 = -1.25

For Z55 = 55 – 50 / 4 = 1.25

Now, we can find the probabilities associated with these z-scores using a standard normal distribution table or calculator.

The probability of Z being less than -1.25 or greater than 1.25 (which corresponds to the sample mean being outside the range of $45 to $55) is approximately 0.2119 (or 21.19%).

So, the probability that there will be an investigation in any given week is approximately 21.19%.

The answer is **option D**

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

To maintain the probability of investigation at 5% while not changing the thresholds of 45 and 55, you can use the normal distribution and the z-score formula.

Given that the auditors want a 5% chance of investigation, this corresponds to the upper tail of the distribution. In other words, they want to find the number of transactions that fall in the top 5% of the distribution.

First, find the z-score corresponding to the 95th percentile (since 100% - 5% = 95% falls below the threshold):

Z-score for 95th percentile ≈ 1.645 (you can find this value using a standard normal distribution table or calculator).

Now, use the formula for the z-score:

Z = (X - μ) / σ

Where:

X is the value you want to find (the minimum number of transactions to sample).

μ is the mean threshold (which is (45 + 55) / 2 = 50).

σ is the standard deviation (which is given as 10).

Plug in the values:

1.645 = (X - 50) / 10

Now, solve for X:

X - 50 = 1.645 \* 10

X - 50 = 16.45

X ≈ 66.45

To maintain a 5% chance of investigation, the minimum number of transactions they should sample is approximately 66.45. Since you can't sample a fraction of a transaction, you'll need to round up to the nearest whole number:

X ≈ 67

So, the minimum number of transactions they should sample is 67.

None of the provided answer choices match this calculation, so it appears there may be a discrepancy in the options provided or additional information needed to answer the question accurately.

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

A Ans. This statement is not likely to be true. The standard deviation within any sample will depend on the distribution of GMAT scores within that specific sample. Since the scores are distributed between 650 and 790 with substantial skewness, the standard deviation within each sample can vary and may not necessarily be exactly 120.

B Ans. This statement is also not likely to be true. The standard deviation of the sample means across several samples, also known as the standard error of the mean (SEM), depends on both the standard deviation of the population (120 in this case) and the sample size. It decreases as the sample size increases. So, the standard error of the mean will generally be smaller than the population standard deviation of 120 for larger samples.

C Ans. This statement is likely to be true. Since you are targeting individuals with an average GMAT score of 720, it's reasonable to expect that the mean score within any random sample of these individuals will be close to 720.

D Ans. This statement is likely to be true as well. If you take multiple random samples from the same population of individuals with an average GMAT score of 720, the average of the sample means across those samples should be close to 720, assuming that the sampling process is unbiased.

E Ans. This statement is not likely to be true. The standard deviation of the mean across several samples is the standard error of the mean (SEM), and it depends on both the population standard deviation (120 in this case) and the sample size. It will be influenced by the sample size and is unlikely to be as small as 0.60 unless the sample size is very large.

So, the statement D are likely to be true, while the statements A, B, C and E are not

likely to be true based on the given information.