**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 =**   **Answer is C.**

**Nearly Normal Distribution:** In a normal quantile plot, if the data points closely follow a straight line without any significant deviations or bends, it suggests that the data is nearly normally distributed.

ii. **Answer is B.**

**Bimodal Distribution:** A bimodal distribution will have two distinct peaks or modes in the plot, indicating that the data has two different groups or sub-populations.

iii. **Answer is A, C and D.**

**Skewed Distribution:** A skewed distribution will have a longer tail on one side of the plot, suggesting that the data is not symmetric around the center.

iv. **Answer is A.**

**Outliers:** Outliers are data points that significantly deviate from the overall pattern in the plot. If there are outliers on both sides of the centre, it indicates that the data has outliers in both the lower and upper tails.

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.

**Ans = Statement 2 is False:**  The standard error of the daily average cannot be determined. The standard error depends on the sample size and the population standard deviation. In this above question the population standard deviation is (5) but sample size (n) is not specified, without knowing the sample size its impossible to calculate standard error.

**=====================================================================================**

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 =** To determine the probability of an investigation we need to calculate the probability that the mean transaction amount falls outside the range of 45 to 55.

We are given that the population mean is 50 dollar and the population standard deviation is 40. When we take a sample of 100 transaction the sampling distribution of the mean will have a standard error.

SE = σ/√5 = 40/√100 = 40/10 = $4

We can use the standard normal (Z) distribution to find Probabilities

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

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

We find the probability corresponding to these Z-score using a standard normal distribution table

P(Z < -1.25) = 0.1056 (approx)

P(Z > -1.25) = 1- P(Z < -1.25) = 1-0.8944 = 0.1056(approx)

Now we add these two probabilities to find the total probability.

P(mean < $45 or mean > $50) = P(Z<-1.25) = 0.1056 + 0.1056 = 0.2112

To express this as a percentage, multiply by 100

0.2112\*100 = 21.12%

So, the closest answer **is Option, 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 = Option D: 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

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

This statement is not likely to be true. In a random sample of individuals, the standard deviation of GMAT scores is likely to vary. The standard deviation within a sample depends on the variability of GMAT scores among the individuals in that specific sample, which can be different from the population standard deviation.

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

This statement is not likely to be true either. When you take multiple random samples from a population and calculate the mean of each sample, the standard deviation of the means (standard error) will be smaller than the standard deviation of the population. It is given that the population standard deviation is 120, so the standard deviation of the sample means will be less than 120.

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

This statement is likely to be true. When you take a random sample from a population, the mean of that sample is expected to be close to the population mean, which is 720 in this case.

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

This statement is likely to be true. When you take multiple random samples and calculate the mean for each sample, the average of those means is expected to be close to the population mean, which is 720.

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

This statement is not likely to be true. The standard deviation of the sample means, known as the standard error, will be less than the population standard deviation but will not be as small as 0.60. The actual value of the standard error depends on the sample size and the population standard deviation.

So, **the likely true statements are C and D.**

**It is likely not possible to decide for E**

**=====================================================================================**