**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: - **Plot C** is nearly normal. The points on the plot lie close to a straight line, indicating that the data is approximately normally distributed.

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

Ans: - **Plot B** has a bimodal distribution. There is a "gap" in the spacing of adjacent data values, indicating that there are two distinct groups of data

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

Ans: - **Plot A** is skewed to the right. The points on the plot lie above the straight line on the left side of the plot and below the straight line on the right side of the plot, indicating that the data is skewed to the right.

**Plot D is** skewed to the left. The points on the plot lie below the straight line on the left side of the plot and above the straight line on the right side of the plot, indicating that the data is skewed to the left.

1. **Have outliers on both sides of the center?**

Ans: - Plot A has outliers on both sides of the center. There are two points that lie far away from the other points on the plot, one on the left side and one on the right side.

Plots B, C, and D do not have outliers on both sides of the center. All of the points on these plots are relatively close to each other.



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 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."

False. The central limit theorem allows for the use of a normal distribution for the sampling distribution of the average package weights, even if the individual package weights are not normally distributed.

**Statement 2:** "The standard error of the daily average SE(x ̅) = 1."

Cannot be determined without knowing the sample size (n).

**3)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?**

1. **1.25%**
2. **2.5%**
3. **10.55%**
4. **21.1%**
5. **50**%

**Ans-:** To determine the probability that in any given week, there will be an investigation, you need to find the probability that the sample mean of withdrawal transactions falls outside the range of $45 to $55.

First, calculate the standard error (SE) of the sample mean, which is given by:

SE = σ / √n

Where: σ (population standard deviation) = $40 n (sample size) = 100

SE = 40 / √100 SE = 40 / 10 SE = $4

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

Z = (X - μ) / SE

Where: X is the value, you're interested in ($45 and $55 in this case) μ is the population mean ($50) SE is the standard error ($4)

For $45:

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

For $55:

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

Now, you want to find the probability that the sample mean is either less than $45 or greater than $55, which corresponds to the area in the tails of the standard normal distribution.

P(X < 45) + P(X > 55) = P(Z < -1.25) + P(Z > 1.25)

You can find these probabilities using a standard normal distribution table or a calculator:

P(Z < -1.25) is approximately 0.1056. P(Z > 1.25) is also approximately 0.1056.

Now, add these probabilities:

0.1056 + 0.1056 = 0.2112

To express this as a percentage, multiply by 100:

0.2112 \* 100 = 21.12%

So, the probability that there will be an investigation in any given week is approximately 21.12%.

The closest answer choice to this is:

**D. 21.1%**

**4)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.**

1. **144**
2. **150**
3. **196**
4. **250**
5. **Not enough information**
6. Ans-: E. Not enough information

**5)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?**

1. **The standard deviation of the scores within any sample will be 120.**
2. **The standard deviation of the mean of across several samples will be 120.**
3. **The mean score in any sample will be 720.**
4. **The average of the mean across several samples will be 720.**
5. **The standard deviation of the mean across several samples will be 0.60**

**Ans-:** Let's analyse each statement one by one:

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

False. The standard deviation within a sample is not necessarily equal to the population standard deviation. It can vary depending on the specific individuals in that sample. While the population standard deviation is 120, the standard deviation within any individual sample can be different.

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

False. The standard deviation of the sample means across several samples is given by the standard error of the mean (SEM), which is calculated as:

SEM = Population Standard Deviation / √(Sample Size)

In this case, the population standard deviation is 120, but the standard error of the mean across several samples will depend on the sample size. It will be smaller than 120 because it's inversely proportional to the square root of the sample size. As the sample size increases, the SEM decreases.

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

False. The mean score in any individual sample can vary. While the population mean is 720, the mean of any specific sample may not be exactly 720 due to random sampling variability.

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

True. The average of the sample means across several samples is expected to be close to the population mean, which is 720. This is a fundamental principle of sampling, known as the Law of Large Numbers. As you take more and more samples and compute their means, the average of those sample means will tend to converge to the population mean.

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

False. The standard deviation of the mean across several samples, also known as the standard error of the mean (SEM), will not be 0.60 unless there is additional information provided about the sample size. The SEM is dependent on the population standard deviation and the sample size, and it will be larger than 0.60, given the population standard deviation of 120.

So, **option D** is likely to be true, but the other statements are not necessarily true