**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?

C:  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.

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

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

1. Are skewed (i.e. not symmetric) ? A,C,D

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

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

Outliers are data points that significantly deviate from the overall pattern in the plot. If there are outliers on both sides of the center, 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.

This is true. This is because the Central Limit Theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, only if the population distribution is approximately normal. Therefore, if the weights of individual packages are not normally distributed, it may affect the validity of using a normal model for the sampling distribution of the average package weights.

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

True. The standard error of the sample mean (SE(x̅)) is a measure of the variability of sample means from different samples in relation to the population mean. If the standard error of the daily average is given as SE(x̅) = 1, it indicates that the average sampling error of the daily means is 1 unit. This value provides information about the precision of the sample mean estimate and helps assess the reliability of the sample mean as an estimate of the population mean.

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 is D

Population mean (μ) = $50

Population standard deviation (σ) = $40

Sample size (n) = 100

Range for no investigation: $45 to $55

import scipy.stats as stats

p = 2 \* (1 - stats.norm.cdf(55, loc=50, scale=40/100\*\*0.5))

P=2×0.1056=0.2112

Converting this probability to a percentage, we get 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

We can use the formula for the standard error of the mean (SE) to calculate the required sample size:

SE= σ/sqrt(n)

5=40/sqrt(n)

Sqrt(n)=40/5

Sqrt(n)=8

N=64

the options provided do not include 64, which means that the provided options might not be correct for this scenario. Therefore, the correct answer is:

E. Not enough information

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.

False

While the population standard deviation is 120, the standard deviation of scores within any sample may vary depending on the composition of the sample and its size. It will not necessarily be exactly 120 for each sample.

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

False

The standard deviation of the sample means across several samples, also known as the standard error of the mean, is calculated as Sigma/sqrt(n), where σ is the population standard deviation and n is the sample size. It will generally be less than 120 because of the effect of sample size.

1. The mean score in any sample will be 720.

False. While the population mean is 720, the mean score in any sample may vary due to sampling variability.

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

True.

The average of the sample means across several samples will tend to approach the population mean of 720 as the number of samples increases. This is a property of the Central Limit Theorem.

1. 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, is calculated using the formula mentioned in option B and depends on the sample size. It will not be 0.60 unless the sample size is specifically set to give that value.