**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. 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 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.

False : A sampling distribution is a probability distribution of a statistic obtained from a large number of samples drawn from a specific population. In our case the samples contain 25 packages and the larger number of samples contain of each such 25 packages taken into different samples (25+25+25+25…and so on). The mean for one these samples is 22lbs and standard deviation of 5lbs which means each individual package is having a weight varying between + or – 5lbs with respect to mean(22lbs). Hence it is invalid to take a weight of individual packages and confirm that it follows normal distribution before using a normal model for the sampling distribution. The Sample Central Limit Theorem states that the sampling distribution of the samples mean approaches normal distribution as the sample size is large enough.

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

 True : As SE(Standard Error) = sample standard deviation / Square root of (number of sample) SE = 5 / (25)^1/2 SE = 1

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: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

Ans:D

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. This statement is unlikely

to be true. The standard deviation within each sample is determined by the variability of scores

present in that specific sample. Since the scores are distributed between 650 and 790 with a

long and thin tail towards the higher end, the standard deviation within any sample may vary

depending on the composition of that sample. Therefore, the standard deviation within a

sample is not fixed at 120.

B. The standard deviation of the mean across several samples will be 120. This statement is also

unlikely to be true. The standard deviation of the mean (standard error) across several samples

can be estimated using the formula: SE = SD / √(sample size). However, the sample size is not

specified in the question. Without the sample size, it is not possible to calculate the standard

deviation of the mean across several samples accurately.

C. The mean score in any sample will be 720. This statement is possible and likely to be true. The

population average GMAT score is given as 720. When randomly selecting samples from the

population, each sample should have an average (mean) that is centered around the population

mean. Therefore, it is likely that the mean score in any sample will be close to 720.

D. The average of the mean across several samples will be 720. This statement is likely to be

true. If we calculate the mean of each sample and then average those means across several

samples, the resulting value should be close to the population mean. Since each sample is

randomly selected from the population, the average of the means across several samples is

expected to be around 720.

E. The standard deviation of the mean across several samples will be 0.60. This statement is

unlikely to be true. The standard deviation of the mean (standard error) across several samples

is determined by the standard deviation of the population and the sample size. In this case, the

standard deviation of the population is given as 120, and the sample size is not specified.

Without knowing the sample size, it is not possible to calculate the exact standard deviation of

the mean across several samples.

Based on the analysis above, it is likely that statement C (&quot;The mean score in any sample

will be 720&quot;) and statement D (&quot;The average of the mean across several samples will be 720&quot;)

are true for randomly chosen samples of aspirants.