**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?

ZZ

Ans:I. C i.e., there is a one-to-one relationship between the data.

II. One way recognize a bimodal shape is a “gap” in the spacing of adjacent data

values. Here, B and D has the gap in the spacing of adjacent data values.

III. A,B and D are not symmetric it means the shape of a graph peaks to the left or

the right of the center.

IV. A and B has the outliers on both sides of the center.

2.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.

i)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.

Ans: True. In this case, at least 30 sample packages must be selected and weighted everyday. Based on the central limit theorem, the sampling distribution of the sample mean approach normal distribution as the sample size become bigger(over 30).

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

Ans: True. Standard error equal to standard deviation divided by square root of sample size=5/sqrt(25)=1

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 find the probability of an investigation, you can use the properties of the normal distribution, given the mean and standard deviation of the withdrawal amounts.

The mean withdrawal amount over the past 2 years is $50 with a standard deviation of $40. This information suggests a normal distribution with μ = $50 and σ = $40.

The auditors will initiate an investigation if the mean transaction amount of the sample is between $45 and $55. To calculate the probability of this occurring, you need to find the Z-scores for $45 and $55 using the formula:

*Z*=*X*−*μ*​/ σ

Where:

* *X* is the value of interest ($45 and $55 in this case).
* *μ* is the population mean ($50).
* *σ* is the population standard deviation ($40).

For $45:

*Z*45​=40-45/40​=−0.125

For $55:

*Z*55​=55−50​/40=0.125

Now, using a standard normal distribution table or a calculator, find the probability that Z lies between -0.125 and 0.125.

The probability between -0.125 and 0.125 is approximately 0.0987.

So, the probability of not initiating an investigation (mean transaction amount between $45 and $55) is 0.0987 or 9.87%.

Given the options, the closest match is C. 10.55%.

Top of Form

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

Ans:

Given: Population size (individuals who have taken GMAT in 2012 and interested in top 20 b-schools) = 40,000 Population mean GMAT score = 720 Population standard deviation = 120 Score range = 650 to 790 Data distribution has a long and thin tail towards the higher end, indicating substantial skewness.

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

Ans: This statement might not be true. In a sample, the standard deviation might differ from the population standard deviation due to random sampling. It's likely that the standard deviation within a sample might be close to 120 but not necessarily exactly the same.

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

Ans: The standard deviation of the sample means across multiple samples is described by the standard error of the mean (SEM). The SEM is calculated as the population standard deviation divided by the square root of the sample size.

SEM = Population standard deviation / √sample size

As the samples are drawn from the same population with the same characteristics, the standard deviation of the sample means across several samples is likely to be less than 120 because the standard error of the mean decreases as the sample size increases.

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

Ans: This statement is not neccesarily true. In each individual sample, the mean might not be exactly 720 due to random sampling. However, the mean of all the sample means might converge to the population mean, which is 720, as the number of samples increases.

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

Ans: This statement is likely to be true. The average of the sample means across multiple samples is expected to be close to the population mean (720) due to the central limit theorem.

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

Ans: This statement is not true. The standard deviation of the mean across several samples is the standard error of the mean. It will not be 0.60, but rather a value less than 120 due to the effect of the sample size.

So, the most likely true statement among the options provided is: D. The average of the mean across several samples will be 720.