**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

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

* B

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

* A,C,D

1. Have outliers on both sides of the center?

* A



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.

* TRUE , In this case at least 30 samples must be collected and weight every day. According to central limit theorem the sampling distribution of sample mean approach normal distribution as sample size bigger than 30

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

* TRUE , Standard error is equal to standard deviation divided by square root of sample size . Standard error = 5/sqrt(25) =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%

* To calculate the probability that there will be an investigation, we can first find the distribution of the sample mean using the Central Limit Theorem. Since the sample size is sufficiently large (n=100), we can assume that the distribution of the sample mean is approximately normal.
* The mean of the sample mean is equal to the population mean, which is $50. The standard deviation of the sample mean is equal to the population standard deviation divided by the square root of the sample size, which is $40/sqrt(100) = $4.
* To find the probability that the sample mean falls outside of the range of $45 to $55, we can calculate the z-score for the lower and upper limits:

z\_lower = ($45 - $50) / $4 = -1.25

z\_upper = ($55 - $50) / $4 = 1.25

* Then, we can use a standard normal distribution table or calculator to find the probability of a z-score less than -1.25 or greater than 1.25. Alternatively, we can use the cumulative distribution function (CDF) of the standard normal distribution:

P(z < -1.25) = 0.1056

P(z > 1.25) = 0.1056

* The probability that the sample mean falls outside of the range of $45 to $55 is the sum of these probabilities:

P(-1.25 < z < 1.25) = 1 - (P(z < -1.25) + P(z > 1.25))

= 1 - (0.1056 + 0.1056)

= 0.789

* Therefore, the probability that there will be an investigation in any given week is 0.789, or about 78.9%. The probability of not being investigation = P(45<x<55)= 0.789

so probability of not being investigation in given week = 1-0.789

= 0.2113

=21.13%

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

* To maintain a probability of investigation at 5%, we need to make sure that the expected number of errors in the sample is greater than or equal to the threshold of 5% of the total number of transactions,

which is 0.05 x 500 = 25.

Let n be the minimum number of transactions that need to be sampled. The expected number of errors in a sample of size n can be calculated as:

Expected number of errors = 0.02n + 0.01(500 - n)

where 0.02n represents the number of errors in the selected transactions and 0.01(500 - n) represents the number of errors in the remaining transactions.

To satisfy the condition that the expected number of errors is greater than or equal to 25, we can set up an inequality:

0.02n + 0.01(500 - n) ≥ 25

Simplifying this inequality, we get:

0.01n + 5 ≥ 25

0.01n ≥ 20

n ≥ 2000/100

n ≥ 20 x 100

n ≥ 200

Therefore, the minimum number of transactions that should be sampled to maintain a probability of investigation at 5% is 200.

However, we are given that we cannot change the thresholds of 45 and 55. Since the thresholds are based on the sample statistics,

changing the sample size will change the thresholds. Therefore, none of the options provided is correct.

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. The standard deviation of the scores within any sample will depend on the size of the sample. As the sample size increases, the standard deviation of the sample means will decrease according to the central limit theorem.

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

* False. The standard deviation of the mean across several samples will depend on the size of the samples and the population standard deviation. As the sample size increases, the standard deviation of the sample means will decrease according to the central limit theorem. However, the standard deviation of the mean across several samples will not be the same as the population standard deviation of 120.

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

* False. The mean score in any sample will vary depending on the individuals included in the sample. The population mean is 720, but the sample mean can be different from this value.

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

* True. The average of the mean across several samples will be equal to the population mean of 720, assuming that the samples are randomly chosen and representative of the population.

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

* False. The standard deviation of the mean across several samples is given by the standard error of the mean, which is equal to the population standard deviation divided by the square root of the sample size. It is not possible to determine the standard deviation of the mean across several samples without knowing the sample size.