**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, D

Bimodal Distribution: Having 2 Peaks in the curve

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

A, B, D

1. Have outliers on both sides of the center?

A, B, D



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

The central limit theorem tells us that if we have a large number of independent, identically distributed variables, the distribution will approximately follow a normal distribution.

It doesn’t matter what the underlying distribution is.

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

True

Standard Error (SE) = σ/√n

= 5/√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%

Standard Error (SE) = $40 / sqrt (100)

= $40 / 10 = $4

For $45:

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

For $55:

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

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

= 0.1056 + 0.1056 = 0.2112

= 0.2112 \* 100

= 21.12%

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

Given, Margin of Acceptance = 5%

Margin of Acceptance = Z \* Standard Error

Standard Error = Standard deviation / √n

Margin of Acceptance = Z \* (Standard deviation / √n)

n = ((Z \* Standard Deviation) / Margin of Acceptance )2

Z-scores with a 5% probability on each tail are approximately ±1.96

n = [(1.96-40)/5]2

n = 246.54

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.

False

The standard deviation of the ‘mean across several samples’, known as the standard error of the mean decreases as the sample size increases. It is not the same as the population standard deviation.

1. The mean score in any sample will be 720.
2. The average of the mean across several samples will be 720.

True

This is likely to be true due to the central limit theorem. As the number of samples increases, the average of the sample means is expected to approach the population mean.

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

True

i.e., Standard Error = 120/sqrt (40000)

= 120/200

= 0.6