**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?

**Ans: C**

**Nearly Normal Distribution:** In a normal quantile plot, if the data points closely follow a straight line without any significant deviations or bends, 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.)

**Ans: 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.

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

**Ans: A , C & 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.

1. Have outliers on both sides of the center?

**Ans: 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.

**Ans: False**

The manager should confirm that weights of individual packages are normally distributed. This is because the Central Limit Theorem, which justifies the use of a normal model for the sampling distribution of the mean, assumes that the underlying population distribution is approximately normal. If the individual package weights are not normally distributed, the sample mean may still be approximately normally distributed if the sample size is large enough.

Samples contain 25 packages and the larger number of samples contain of each such 25 packages taken into different samples (25+25+25+25…). 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.

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

**Ans: True**

Sample standard deviation =5

Number of sample = 25

SE(Standard Error) = Sample standard deviation / Square root of (number of sample)

SE = 5 / √25

SE = 5/5 =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**

X is the sample mean,

μ is the population mean,

σ is the population standard deviation,

n is the sample size.

The z-scores for the lower and upper bounds are calculated as follows:

Zlower​= ​45−50/(40/√100)​=−5/4

Zupper​= 55−50/(40/√100)​=− 5/4

Now, probabilities corresponding to these z-scores using a standard normal distribution table or calculator:

P ( Z < -5/4) = 0.1056

P ( Z< 5/4) = 0.8944

The probability that the sample mean falls between $45 and $55 is given by the difference of these probabilities:

= P ( Z< 5/4) - P ( Z < -5/4)

= 0.8944−0.1056

= 0.7888

= 1-0.7888

= 0.214

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:** The formula for the z-score is:

Z =​ (Xˉ−μ​)/(σ /√n)

Z corresponds to the 5% probability level, you can find the required sample size n

The z-score for a 5% probability is approximately -1.645.

= −1.645=45−50/(40/√n)​

√n ​= 40×1.645​ /5

n = (40×1.645/5)2

= (13.16)2

= 173.3

The options provided, the closest one is 196.

**Answer is C**

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:** SE = ​σ​/ √n

σ is the population standard deviation,

n is the sample size.

= 120 /√40000

=0.6

**Answer is E.**