**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?



**Answer-**

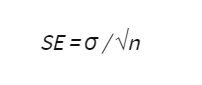
1. C plot indicates that data distribution is nearly normal, because maximum data points are around or on the straight line.
2. B plot have bimodal distribution.
3. A plot is skewed.
4. D plot having outlies on both the sides of the center.
5. For each of the following statements, indicate whether it 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.

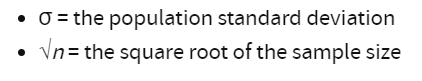
1. Before using a normal model for the sampling distribution of the average package weights, the manager must confirm that weights of individual packages are normally distributed.
2. The standard error of the daily average SE () = 1.

**Answer-**

1. ***False***- According to central Limit theorem Distribution of sample mean will always follow normal distribution irrespective of Distribution of population mean is normal or not. So, weight of the individual packages will always be normally distributed even if average package weight of sample is not distributed normally.
2. ***True***- Standard error can be calculated by using formula,



Where,



Substituting standard deviation and sample size,

SE = 5/sq.root 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%

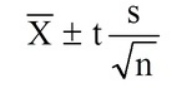
**Answer-**

Defining Hypothesis,

Null Hypothesis – There will be no investigation

Alternate Hypothesis- There will be an investigation.

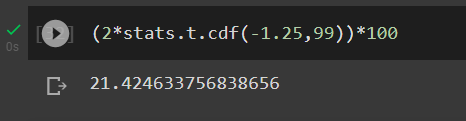
We have confidence interval [45,55]



45= 50-(t\*40/10)

t=1.25

since given problem statement comes under two tail test, multiplying stats.t.cdf(t,df) with 2.



Since P value is more than alpha (5%), we need to accept alternate Hypothesis. i.e. There is a need of investigation.

Therefore, Probability of investigation is 21.42%

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

**Answer**- D.

5% probability of investigation means 95% confidence interval. For 95% confidence interval Z value is 1.96.

Substituting Z score value in confidence interval formula,

45=50-z\*40/sq.root N

N = 245.86≈250.

Therefore, minimum transaction is 250.

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

**Answer-**

1. ***False***- Standard deviation of scores within any sample will not be 120. Because we can take any sample which has less variability or standard deviation i.e., Less variation is GMAT score.
2. ***True***- The standard Deviation of mean of several samples can be 120, because Maximum GMAT scores can be covered in this case.
3. ***False***- The mean score in any sample will not be 720, because distribution od GMAT score is Positively skewed and hence mean cannot be same if any sample is taken into consideration.
4. ***True***- The average of mean across several samples can be 720, because It will cover maximum GMAT scores.
5. ***False***- The standard Deviation of mean across several samples will not be 0.6, since variability of population is 120, and if we take mean standard deviation of several samples, the value will come around 120.