**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
3. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.) Ans: D
4. Are skewed (i.e. not symmetric)? Ans: A
5. Have outliers on both sides of the center? Ans: B



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

When the sample size is large the sample proportion is normally distributed, here the sample size is <30. Manager cannot confirm its normally distributed, on the assumption of normal distribution sample sizes can be taken and tested against various sizes of samples.

As per central limit theorem, the distribution of the sample mean will be normal when the distribution of data in the population is n normal. The distribution of the sample mean will be approximately normal even if the distribution of the data in the population is not normal, if the sample size is fairly large.

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

Ans: True, (s/sqrt. (n)) = 5/251/2 =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

Since we are handling sampling distribution,

= given as a range [45 55]

µ = 50

SE = s/sqrt(n) = 40/sqrt.100 = 4

Since sample size 100>30, we can use this data as per central limit theorem it is normal distribution.

Now find,

P(45<x<55)

Z = 45-50/4

Rcode🡪pnorm (-5/4) = 0.1056498

Z = 55-50/4

Rcode🡪pnorm (5/4) = 0.8943502

P(45<x<55) = 0.8943502 - 0.1056498 = 0.7887

Probability to start investigation in any week = 1-0.7887 = 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

Ans: D

S.E = 40/sqrt (250)

P(45<x<55) = pnorm (+5/2.53) – pnorm (-5/2.53)

P(45<x<55) = 95.18% when sample size is 250

So, probability for investigation with same statistical values = 5%(rounded)

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: Standard Error SE = SD/sqrt(n) = 120/sqrt (40000) = 0.6

1. The standard deviation of the scores within any sample will be 120.- False

The standard deviation of the scores within any sample will not be 120, especially Since we don’t know the sample size.

1. The standard deviation of the mean of across several samples will be 120- False SD of mean of across several samples will not be 120. It will be less, probably about 0.6.
2. The mean score in any sample will be 720. – True.
3. The average of the mean across several samples will be 720. – possible -True – required to calculate with all the sample means.
4. The standard deviation of the mean across several samples will be 0.60 – True- According to SE calculated with the sample size and standard deviation.