**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?

Solution: In Quantile plots if datapoints are nearly distributed in the form of line without much gaps or bends then the points are said to be normally distributed. Therefore, **Figure C** contains data points that are nearly normal

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

Solution: A bimodal distribution is the one in which the distribution has 2 modes or peaks. There are 2 prominent peaks in the distribution. Therefore, **Figure B** has a bimodal distribution

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

Solution: If the plot is said to have longer tails on one side compared to the other side then it is said to be skewed. Therefore, **Figures A,B and D** are skewed

1. Have outliers on both sides of the center?

Solution: Outliers are points that significantly deviate from the overall pattern. Therefore, **Figure A and B** have outliers.



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.

Solution: True. Before using the normal model for the sampling distribution of the average package weights, the weights of the individual packages are to be normally distributed as per the CLT (Central Limit Theorem). CLT states that the population should be approximately normal distributed for the sampling distribution of the mean to be normally distributed

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

Solution: True. The Standard error of mean is given by s/sqrroot(n)

Where s is standard deviation and n is sample size

=5/sqrroot(25)

=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%

Solution: To find the probability that there will be an investigation then we have to find the probability when the mean value obtained is not in the specified range. So we have to find the probability when mean is less than 45 and greater than 55. To find that we have to find the SE about mean.

SE about mean=standard deviation/root (size)

=40/10

=4

Then wkt Z=(X- *μ)*/SE

Z45=45-50/4

=-5/4

=-1.25

Z55=55-50/4

=5/4

=1.25

P(Z45)+P(Z55)=0.106+0.106

=0.212=21.2%

Therefore, **Option D** is the answer.

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

Solution: Wkt Z=(X- *μ)*/SE

Where se=standard deviation/root(n)

Since 5% is the significance level in this case the standard value of Z is -1.645

For X=45

-1.645=(45-50)/(40/root n)

-1.645=-5/(40/root n)

(-1.645\*40)/5=root n

=173.2

The answer is close to 196. Therefore the answer is **Option 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

Solution: **Option C and D** will be TRUE. Since the mean of the population is 720 the mean in any sample will be 720.

WKT the average of several samples of a normally distributed data are to be found closer to population mean according to the CLT. Therefore option D is also TRUE.