**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?



Ans:

* + 1. Are nearly normal?

C.

* + 1. Have a bimodal distribution?

B and D.

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

A, B and D.

* + 1. Have outliers on both sides of the center

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

Ans:

1. False.

Since we are working on a sample data which is 25 packages, as in the question mean for the sample is 22 lbs. and standard deviation of 5 lbs. which variance is 25(approx..)

The mean is 22 lbs. and the variance is 25lbs it means that the individual package weights are spread out over a wider range around the mean. This could affect the shape of the distribution and potentially impact the reliability of certain statistical assumptions

Each package is having a average weight vary between + or -5 with respect to mean. even if the sample package is normally distributed we cannot say the whole shipments made by the deliver team is normally distributed, because the sample is 25 which is very low sample to build a model.

1. True.

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 standard deviation is not given we take the t-test

T45 = (45-50)/40/sqrt(100)

= - 5/40/10

= - 1.25

T55 = (55-50)/40/sqrt(100)

= 5/40/10

= 1.25

**[Refer IPYNB for Better Reference]**

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.

The formula for the z-score is

Z = (sample mean – population mean)/(population standard deviation)/sqrt(sample size)

If we rearrange the formula we get

Sample size = [(sample mean – population mean)/ (population standard Deviation)/ Z]2

As we simplify, we get

n = 249.92(approx..) for both mean 45 & 55

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.

D. The average of the sample means across several samples (the sampling distribution of the sample means) should be close to the population mean, which is 720.