**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:

I. C

II. B

III. A, C and D

IV. A

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. True

Before using a normal model for the sampling distribution of the average package weights (sampling distribution of the mean), it is necessary to confirm that the weights of individual packages are normally distributed. According to the Central Limit Theorem, the sampling distribution of the mean becomes approximately normal as the sample size increases, even if the underlying population distribution is not normal. However, for small sample sizes, it's important to check the normality assumption.

1. True

The standard error of the mean (SE(x̄)) is calculated as σ/√n, where σ is the population

standard deviation and n is the sample size. In this case, the standard error would be 5/√25=1. Therefore, the statement SE(x̄) = 1 is true. Not false. The standard error is a measure of the variability of the sample mean and is affected by both the population standard deviation and the sample size.

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:

To find the probability of an investigation, you can use the standard normal distribution because the sample size is sufficiently large (n = 100). The distribution of the sample mean (x̄) can be approximated by a normal distribution due to the Central Limit Theorem.

First, calculate the standard error of the sample mean (SE(x̄)):

SE(x̄) = σ/√n

Where:

* σ is the population standard deviation,
* n is the sample size.

In this case:

SE(x̄) = 40/√100 = 4

Now, we need to find the probability that the mean transaction amount falls between $45 and $55. To do this, we need to convert these values into z-scores using the formula:

Z = (x- *μ*)/SE(x̄)

Where:

* x is the value,
* *μ* is the mean,
* SE(x̄) is the standard error of the sample mean.

For $45:

Z\_1 = (45-50)/4 = -1.25

For $50:

Z\_2 = (55-50)/4 = 1.25

Now, we need to find the probability corresponding to these z-scores using a standard normal distribution table.

P(-1.25 < Z < 1.25)

Using a standard normal distribution table or a calculator, we find the probability to be approximately 0.7887.

Now, let’s calculate the probability of an investigation:

P(Investigation) = 1 - P(-1.25 < Z < 1.25)

P(Investigation) = 1 - 0.7887

P(Investigation) ≈ 0.2113

The probability that in any given week there will be an investigation is approximately 21.13%.

Hence, Option D. 21.13 % is correct.

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:

To find the minimum sample size (n) to maintain a probability of investigation at 5%, we need to use the critical z-value associated with a 5% significance level (two-tailed test).

Let's denote the critical z-value as z\_critical, which is the z-value that separates the central region from the tails. For a 5% significance level, z\_critical is approximately 1.96 (you can find this value in a standard normal distribution table).

Now, use the formula for the minimum sample size:

n = ((z\_critical \* σ) / Margin of Error)^2

The margin of error in this case is half the width of the interval, which is (55 - 45)/2 = 5.

Substitute the values into the formula:

n = ((1.96 \* 40) / 5)^2

Calculate the result to find the minimum sample size.

n ≈ ((1.96 \* 40) / 5)^2

n ≈ (196 / 5)^2

n ≈ (39.2)^2

n ≈ 1536.64

Now, let’s round up to the nearest whole number because we can't have a fraction of a sample:

n ≈ 1537

So, the minimum number of transactions that they should sample to maintain a 5% probability of investigation is not provided among the given options. The correct answer is:

E. Not enough information

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:

The statements which are likely to be true are:

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

- This statement is correct. The standard deviation of the population (σ) is the same as the standard deviation within any sample.

D. The average of the mean across several samples will be 720.

- This statement is likely to be true. The expected value of the sample mean across several samples will be equal to the population mean (*μ*).