**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

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

**Ans**: D

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

**Ans**: A,B and C

1. Have outliers on both sides of the center?

**Ans**: 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.

**Ans**: False , The Central Limit Theorem (CLT) allows us to use a normal model for the sampling distribution of the average package weights, even if the individual data points (weights of individual packages) are not normally distributed. As long as the sample size is large (typically n > 30), the sampling distribution of the average package weights tends to be approximately normal, So no need to check the for every individual package.

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

**Ans**: True , It is calculated as the standard deviation of the population (σ) divided by the square root of the sample size (n). In this case, if σ (population standard deviation) is 5 lbs and the sample size (n) is 25,

then SE(x̄) = 5 / (25^1/2)

= 5 / 5

= 1. So, the statement is 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% (**D is correct answer**)
6. 50%

**Ans**: standard error of the sample mean (SE(x̄)) using the population standard deviation (σ) and the sample size (n): mean\_population = 50 # Average withdrawal amount over the past 2 years

std\_population = 40 # Standard deviation of withdrawal amounts

sample\_size = 100

lower\_limit = 45

upper\_limit = 55

SE(x̄) = σ / (n^1/2)

= 40 / (100^1/2)

= 4

Z for $45 = ($45 - $50) / $4

= -5 / $4

= -1.25

Z for $55 = ($55 - $50) / $4

= 5 / $4 = 1.25

Using python

stats.norm.cdf(1.25)= 0.8943502263331446

stats.norm.cdf(-1.25)= 0.10564977366685535

P = 1 – (0.8943502263331446)-(0.10564977366685535)

=21.129%

There is a 21.129% probability that there will be an investigation in any given week.

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 **Option E** is correct answer

Ans:

To calculate the minimum sample size required to maintain a 5% probability of investigation, you can use the following formula for the margin of error in the sampling process:

Margin of Error = Z \* (Standard Deviation / sqrt(n))

Where:

Z is the Z-score associated with the desired level of confidence (for a 95% confidence level, Z ≈ 1.96).

Standard Deviation is the population standard deviation (assuming it's known).

n is the sample size.

If we don't want to change the thresholds of 45 and 55, this means that the width of the range of acceptable values is 10 units.

So, we want to find the minimum sample size (n) that keeps the margin of error within 5 units. Therefore:

5 = 1.96 \* (Standard Deviation / sqrt(n))

Since the sample statistics remain unchanged, we can assume that the standard deviation remains the same.

5 = 1.96 \* (Original Standard Deviation / sqrt(n))

Now, solve for n:

sqrt(n) = 1.96 \* Original Standard Deviation / 5

n = (1.96 \* Original Standard Deviation / 5)^2

We don't have the value of the original standard deviation, so we can't calculate the exact value of n. Therefore, the answer is E (Not enough information). To find the minimum sample size, you would need to know the standard deviation of the population.

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

**Option D is True**. The average of the means across several samples should be close to the population mean of 720