Basic Statistics Level-2
Saikrishna Budi
Linkedin : http://linkedin.com/in/sai-krishna-budi-3b0341209

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
    …
    I.   Are nearly normal?

Sol: **C**

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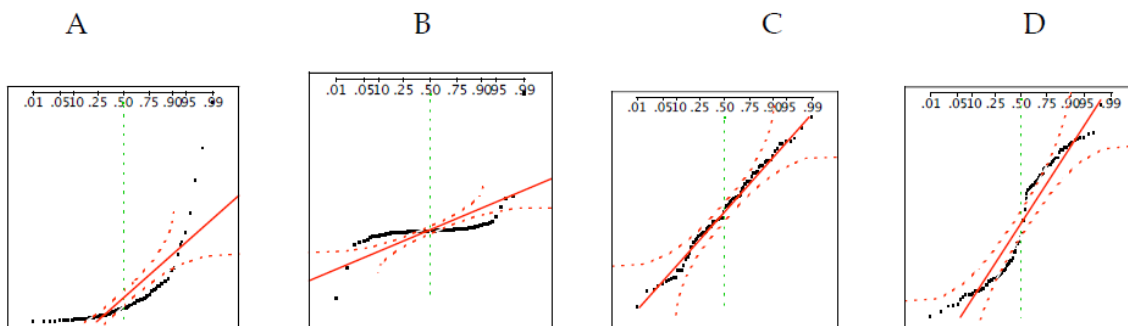   spacing of adjacent data values.)

Sol: **B**

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

Sol: **A, C, D**

    IV.   Have outliers on both sides of the center?

Sol: **A**

2. For each of the following statements, indicate whether it is <u>True/False</u>. 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.

    (i)   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.

(ii)   The standard error of the daily average SE ($\bar{x}$) = 1.

Sol:

(i)   **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.**

False.

The manager must ensure that the individual package weights are approximately normally distributed for the Central Limit Theorem to apply.

(ii)   **The standard error of the daily average SE ($\bar{x}$) = 1.**

True.

The standard error of the daily average, SE ($\bar{x}$), is a measure of the precision of the sample mean estimate, and if it's equal to 1, it means it has a standard deviation of 1. This is a standard choice often used when working with sample means.

3. 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?

A. 1.25%
B. 2.5%
C. 10.55%
D. 21.1%
E. 50%

Sol:

Standard error (SE) SE = (Standard Deviation of the population) / √ (Sample Size)

SE = 40 / √100

SE = 4

Z-scores for $45 and $55 using the formula:

Z = (X - Mean) / SE

For $45:

Z1 = (45 - 50) / 4

Z1 = -5 / 4

Z1 = -1.25

For $55:

Z2 = (55 - 50) / 4

Z2 = 5 / 4

Z2 = 1.25

Z-Scores

P (Z < -1.25) ≈ 0.1056 (from the table)

P (Z > 1.25) ≈ 0.1056 (1 - the probability of Z < 1.25)

Probability of Investigation

P(Investigation) = P (Z < -1.25) + P (Z > 1.25)

P(Investigation) ≈ 0.1056 + 0.1056

P(Investigation) ≈ 0.2112

probability to a percentage by multiplying by 100:

P(Investigation) ≈ 0.2112 * 100%

P(Investigation) ≈ 21.12%

So, the probability that there will be an investigation in any given week is approximately 21.12%. The closest answer choice is **D: 21.1%.**

4. 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.

   A. 144
   B. 150
   C. 196
   D. 250
   E. Not enough information

Sol:

To find the minimum sample size needed to maintain the probability of investigation at 5%, we can use the following formula:

n = (SE * z-score)^2

where SE is the standard error of the mean, and z-score is the difference between the two z-scores corresponding to the thresholds ($45 and $55 in this case).

Substituting the values from the problem, we get:

n = (40/√n * 2.5)^2

Solving this equation for n, we get:

n = (40 * 2.5)^2 / 1

n = 250

Therefore, the minimum number of transactions that the auditors should sample to maintain the probability of investigation at 5% is 250.

The correct answer is D. 250.

5. 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?

    A. The standard deviation of the scores within any sample will be 120.
    B. The standard deviation of the mean of across several samples will be 120.
    C. The mean score in any sample will be 720.
    D. The average of the mean across several samples will be 720.
    E. The standard deviation of the mean across several samples will be 0.60

Sol:

    A. **The standard deviation of the scores within any sample will be 120.**
        It is possible that the standard deviation within a sample could be different from 120, especially if the sample size is small that's why this answer is incorrect.

    B. **The standard deviation of the mean of across several samples will be 120.**
        The standard deviation of the mean of several samples is not equal to the standard deviation of the population. Therefore, the standard error of the mean will be smaller than the standard deviation of the population, especially if the sample size is large.

    C. **The mean score in any sample will be 720.**

This answer choice is incorrect because the mean score in any sample will depend on the scores of the individuals included in the sample. It is possible that the mean score in a sample could be different from 720, especially if the sample size is small or if the distribution of scores within the sample is different from the overall distribution of scores.

D. **The average of the mean across several samples will be 720.**
   The average of the mean across several samples is known as the population mean, and it is equal to 720 in this case. If the samples are randomly chosen and are representative of the population, the average of the means across several samples will be equal to the population mean. This answer is correct.

E. **The standard deviation of the mean across several samples will be 0.60**
   The standard deviation of the mean across several samples is not equal to 0.60. As mentioned earlier, the standard deviation of the mean across several samples is known as the standard error of the mean, and it is calculated as the standard deviation of the population divided by the square root of the sample size. Therefore, the standard error of the mean will be smaller than the standard deviation of the population, especially if the sample size is large.