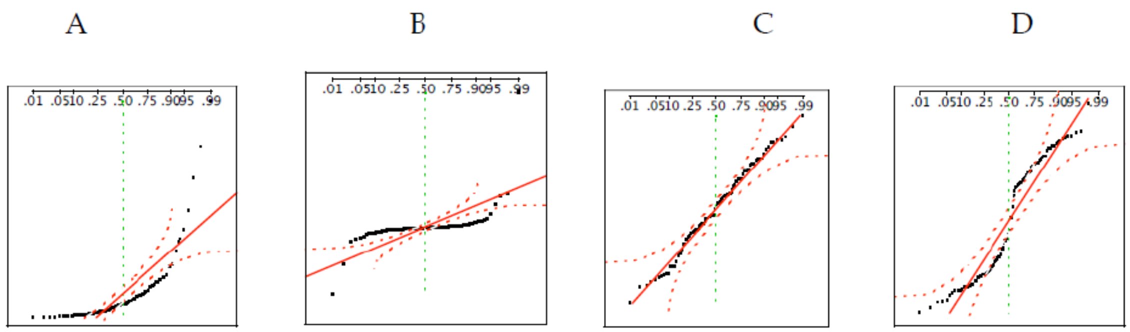
Q1

1. Examine the following normal Quantile plots carefully. Which of these plots indicates that the data …

   I. Are nearly normal?
   II. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.)
   III. Are skewed (i.e. not symmetric) ?
   IV. Have outliers on both sides of the center?



Answer 1 :

The normal quantile plots: Before looking at the actual plots, recall that a normal quantile plot plots the theoretical values on the X-axis and the actual (data) values on the Y-axis. For each point in the data, it calculates the expected theoretical value (or its Z-score) based on the number of points and the properties of the normal distribution. This is the X-coordinate of the point. The Y-coordinate is the actual data point observed. If these two are equal, the point will lie along the diagonal. If a point is above the diagonal, this means that its actual value is higher than the one that would be expected based on the normal distribution (with the mean and SD given by the data). If it lies lower than the diagonal line, it is the opposite. Now we can examine each of the plots.

A. **Skewed**. Explanation: The normal quantile plot has a crescent shape with the points lying above the diagonal both at the lower and upper ends. Further, the points are a lot closer together in the left tail than in the right tail. This implies that the data is right skewed i.e. has a lot more points closer to the mean on the left tail, and has a long tail on the right side. Both of these result in data values that are higher than those expected.

B. **Outliers**. Explanation: There are a couple of points which are quite far away from the rest of the data and have values quite different from those expected from a normal distribution (especially in the left tail. It is not quite so pronounced on the right tail).

C. **Normal**. Explanation: All of the points track the diagonal quite closely and lie within the bounds.

D. **Bimodal**. Explanation: Notice that the density of points near the median is much sparser than on either side of the median. This means significant probability masses on either side of the median with a low probability mass near the median, a shape characteristic of a bimodal distribution.

Q2.

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.

(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.

Answer:

I. **False**. While normality of the weights of individual packages guarantees that the mean weights will be normally distributed, it is not necessary that the weights of individual packets will be normally distributed. Based on the Central Limit Theorem, the mean weights are normally distributed even if the individual package weights don't have a normal distribution so long as the sample size conditions necessary for normality are satisfied. So those conditions should be checked.

II. **True.**

$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1lb.$$

Q3. 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%

Answer 2:

Option **D**.

Here, μ = 50, σ = 40, n = 100. Note that the probability of investigation is the same as probability that the sample mean is beyond $45 and $55.

From Central Limit Theorem, we know that

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}).$$

We first calculate the standard error of the mean, which is

$$\frac{\sigma}{\sqrt{n}} = 4$$

Then, we calculate the z-values corresponding to 45 and 55, which are (-1.25) and (+1.25). Therefore,

Consequently, probability that the sample mean is beyond the limits (and hence there is an investigation) is (1 - 0.79) = **0.21**.

Q4.
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

Answer 4

Option **D**.

In this case, we know that‾. Thus, the z-values corresponding to 45 and 55 are (-1.96) and (+1.96) respectively. Using the definition of z-value, we get

$$1.96 = \frac{55-50}{\sigma_{\bar{x}}}$$

$$\sigma_{\bar{x}} = 2.55$$

Finally, using the definition of‾, we obtain

$$\frac{40}{\sqrt{n}} = 2.55$$

which gives n = 245.86.

We choose 250 as the closest integer greater than this number. Choosing anything smaller than 245.86 will yield a probability of investigation higher than 0.05.

Q5.

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

Answer:

Option **D**.

Again, from Central Limit Theorem, ¯. Hence, the average of many sample means will tend to the population mean, which is 720.