Samit Dhawal

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 …

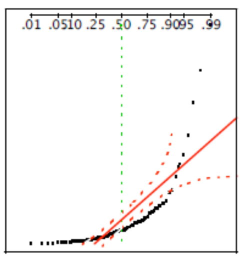
    I. Are nearly normal?

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

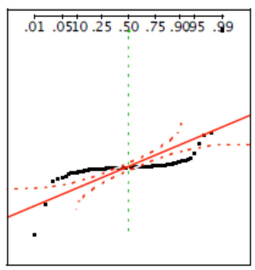
    III. Are skewed (i.e. not symmetric) ?

    IV. Have outliers on both sides of the center?

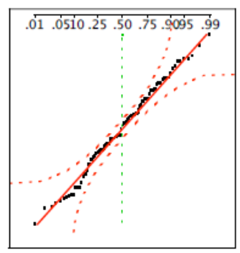
A



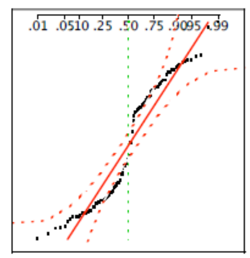
B



C



D



Ans:

I. C

II. B

III. A, C, D

IV. 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.

Ans: True. Before using a normal model for the sampling distribution of the average package weights, the manager must confirm that the weights of individual packages are normally distributed. This is because the sampling distribution of the mean relies on the assumption of normality in the underlying population.

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

Ans: False. The standard error of the daily average (SE($\bar{x}$)) is not necessarily equal to 1. The standard error is calculated as the standard deviation of the population divided by the square root of the sample size. In this case, with μ = 22 lbs. and σ = 5 lbs., the standard error can be computed as $\sigma/\sqrt{n}$ (where n is the sample size). Without knowing the sample size. We cannot determine the specific value of the standard error.

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%

Ans: To calculate the probability of an investigation, we need to determine the probability that the mea transaction amount of the sample falls outside the range of $45 to $55.

First, let's calculate the standard error (SE) of the sample mean,

SE = $\sigma/\sqrt{n}$

Where, σ is the standard deviation ($40) and n is the sample size (100).

SE = $40/\sqrt{100}$

SE = $40/10$

SE = 4

Now, we calculate the z-scores corresponding to the lower and upper limits of the desired range

lower_z_score = (45-50)/4
lower_z_score = -1.25

upper_z_score = (55-50)/4
upper_z_score = 1.25

Now, we find Cumulative probabilities associated with these z-scores.

The probability of the mean transaction amount falling below $45 (below the lower limit) is
P (z < (-1.25)) = 0.1056(approximately)

```
# Q3

from scipy import stats

stats.norm.cdf(-1.25)
✓ 0.0s

0.10564977366685535
```

The probability of the mean transaction amount exceeding $55(above the upper limit) is
P (z > 1.25) = 0.1056(approximately)

```
1-stats.norm.cdf(1.25)
✓ 0.0s

0.10564977366685535
```

Since, we are interested in the probability of an investigation, which is the sum of these two probabilities, the total probability is,

P (investigation) = P (z < (-1.25)) + P (z > 1.25)
P (investigation) = 0.1056 + 0.1056
P (investigation) = 0.2112

Converting to a percentage, the probability of an investigation in any given week is 21.12%.

Therefore, D. 21.1% is correct.

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

Ans: To maintain a probability of investigation at 5%, we need to find the sample size that corresponds to a critical value (z-score) that leaves 5% in the tails of the distribution. Since the auditors want to keep the same thresholds of $45 and $55, we can assume the z-scores remain the same.

The z-score associated with a 5% probability in the tail (2.5% on each side) is approximately 1.96.
The formula to calculate the minimum sample size needed is,
$n = (z^2 * \sigma^2)/e^2$
where, z is the critical value, σ is the standard deviation, and e is the desired margin of error.

$n = (1.96^2 * 40^2)/ (5^2)$
$n = (3.8415999999999997 * 1600)/ 25$
$n= 6146.5599999999995 / 25$
$n = 245.86239999999998 \sim 246$

Rounding up to the nearest whole number, the minimum number of transactions that should be sampled is 246.

So, E. Not enough information (since none of the given options match the calculated minimum sample size).

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.
   Ans: False. The standard deviation of the scores within any sample is not necessarily 120. It will vary depending on the specific individuals included in the sample.

   B. The standard deviation of the mean of across several samples will be 120.
   Ans: True. The standard deviation of the mean across several samples will be 120 divided by the square root of the sample size. As the sample size increases, the standard deviation of the mean decreases, indicating a more precise estimate of the population mean.

   C. The mean score in any sample will be 720.
   Ans: False. The mean score in any sample is not guaranteed to be exactly 720. It will vary from sample to sample due to random sampling variability.

   D. The average of the mean across several samples will be 720.
   Ans: True. The average of the mean across several samples will be approximately equal to the population mean, which is 720 in this case. This is because the sample means are expected to be unbiased estimators of the population mean.

   E. The standard deviation of the mean across several samples will be 0.60
   Ans: False. The standard deviation of the mean across several samples will not be 0.60. It will be equal to the standard deviation of the population divided by the square root of the sample size, which in this case is 120 divided by the square root of the sample size.

   Therefore, the correct answers are,

   B. The standard deviation of the mean across several samples will be 120.
   D. The average of the mean across several samples will be 720.