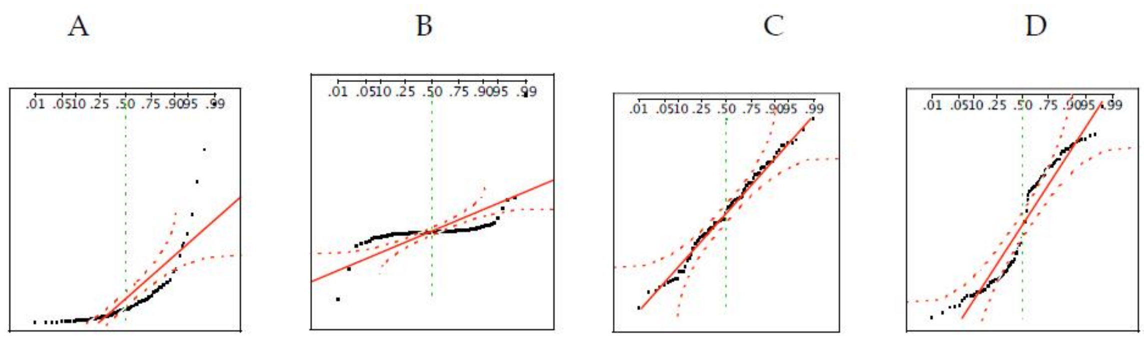
CBA: Practice Problem Set 2

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?



Ans
I. C
II. B&D
III. A,B&D
IV. A&B

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.

1. **True.** Before using a normal model for the sampling distribution of the average package weights, it's essential to confirm that weights of individual packages are normally distributed. The Central Limit Theorem states that the sampling distribution of the sample mean will be approximately normal, regardless of the shape of the population distribution, as long as the sample size is sufficiently large (typically $n \geq 30$) and the population distribution isn't extremely skewed or has outliers. However, for smaller sample sizes or if the population distribution significantly deviates from normality, using a normal model might not be appropriate.
2. **False.** The standard error of the mean (SE) is calculated as the standard deviation of the population divided by the square root of the sample size ($\sigma/\sqrt{n}$). In this scenario, $\sigma = 5$ lbs. and $n = 25$ packages. Therefore, the standard error of the daily average would be $SE(\bar{x}) = 5/\sqrt{25} = 5/5 = 1$ lb., not $SE(\bar{x}) = 1$ lb.

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) d

- Population mean (μ) = $50
- Population standard deviation (σ) = $40
- Sample size (n) = 100

  First, let's find the standard error of the mean (SE): $SE = \sigma / \sqrt{n} = 40 / \sqrt{100} = 40 / 10 = 4$

  Now, we need to find the z-scores corresponding to $45 and $55: For $45: $z = (x - \mu) / SE = (45 - 50) / 4 = -5 / 4 = -1.25$ For $55: $z = (x - \mu) / SE = (55 - 50) / 4 = 5 / 4 = 1.25$

  Next, we look up the corresponding probabilities in the standard normal distribution table or use a calculator. The probability of a z-score being less than -1.25 or greater than 1.25 is the probability of being outside the range $45 to $55.

From the standard normal distribution table or a calculator, the probability of a z-score being less than -1.25 or greater than 1.25 is approximately 2 * (1 - 0.8944) = 2 * 0.1056 ≈ 0.2112.

Therefore, the probability of there being an investigation in any given week is approximately 21.12%, which corresponds to option D.

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) d

- *z* is the critical value (1.96 for a 5% probability on each tail),
- *σ* is the population standard deviation ($40 in this case),
- *E* is the margin of error, which is half the width of the interval ($55 - $45 = $10, so E = \frac{10}{2} = $5).

Substituting the values:

n=(51.96⋅ 40)2

n=(578.4)2

n=(15.68)2

n≈246.74

Rounding up to the nearest whole number because we can't have a fraction of a transaction:

n≈247

Therefore, the minimum number of transactions that the auditors should sample to maintain a 5% probability of investigation is 247, which isn't listed among the options. It seems like there may be a mistake in the options or the problem setup. However, given the options provided, we would choose the closest value, which is option 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

A. **False.** The standard deviation within any sample is unlikely to be exactly 120 because the sample is randomly chosen, and the distribution has substantial skewness. While the population standard deviation is given as 120, the standard deviation within each sample will vary due to sampling variability.

B. **True.** According to the Central Limit Theorem, for large sample sizes, the distribution of sample means will be approximately normal, regardless of the shape of the population distribution. The standard deviation of the sampling distribution of the mean (standard error) is given by $\sigma/\sqrt{n}$, where σ is the population standard deviation and n is the sample size. In this case, the standard deviation of the mean across several samples will indeed be $120/\sqrt{n}$, which approaches 120 as the sample size increases.

C. **True.** Since the population mean is 720, any randomly chosen sample is expected to have a mean score close to 720, assuming that the sampling process is unbiased.

D. **True.** According to the Central Limit Theorem, the average of the means across several samples will be equal to the population mean. Since the population mean is given as 720, the average of the means across several samples will also be 720.

E. **False.** The standard deviation of the mean across several samples will not be 0.60. It will depend on the sample size. As mentioned earlier, the standard deviation of the sampling distribution of the mean is $\sigma/\sqrt{n}$. In this case, it would be $120/\sqrt{n}$, not 0.60.

Therefore, the correct statement is:

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