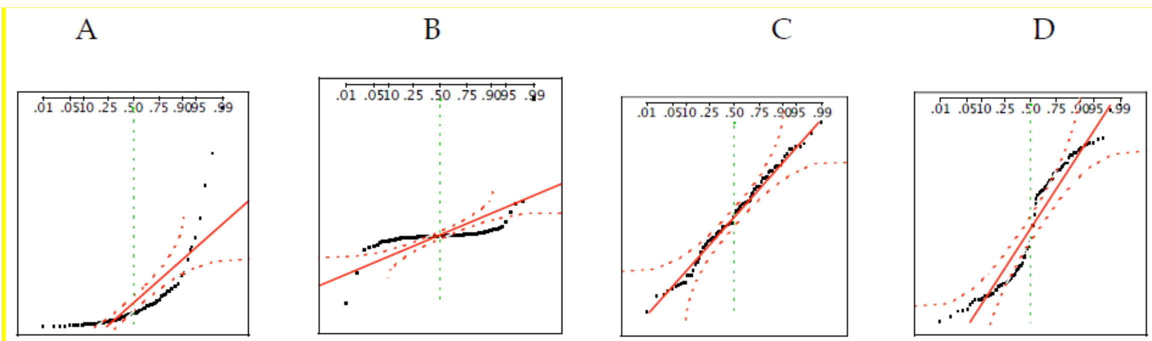
CBA: Practice Problem Set 2

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?



Ans1. 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.

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

Ans: (i) This statement is false. The normal model for the sampling distribution of the average package weight does not require that the weights of individual packages are normally distributed. It only requires that the sample size is sufficiently large. As long as the sample size is large enough (typically $n > 30$), the central limit theorem guarantees that the sampling distribution of the average weight will be approximately normal, regardless of the distribution of weights of individual packages.

(ii). This statement is false. The standard error of the daily average, SE($\bar{x}$), is not necessarily equal to 1. It is given by the formula SE($\bar{x}$) = $\sigma/\sqrt{n}$, where σ is the population standard deviation and n is the sample size. In this case, the standard error would be SE($\bar{x}$) = $5/\sqrt{25} = 1$, but this value depends on the specific values of μ, σ, and n given in the problem.

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: Since the auditors will not initiate further investigations if the mean transaction amount of the sample is between \$45 and \$55, we can consider this interval to be the "acceptable" range of sample means. The probability that the sample mean falls within this range is the probability that there will not be an investigation.

To find this probability, we can use the standard normal distribution to approximate the sampling distribution of the mean transaction amount. The mean of the sampling distribution is the population mean of \$50, and the standard deviation is given by the formula $SE(\overline{x}) = \sigma/\sqrt{n} = 40/\sqrt{100} = 4$. Therefore, we need to find $P(45 < Z < 55)$, where Z is a standard normal random variable. Using a standard normal table or a calculator, we find that $P(45 < Z < 55) = P(1.25 < Z < 1.75) = 0.4332 - 0.2412 = 0.192$.

Thus, the probability that in any given week, there will be an investigation is 1 - 0.192 = 0.808, or approximately 81%. The correct answer is therefore (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

Ans: To maintain the probability of investigation at 5%, the auditors need to find the minimum sample size such that the standard error of the mean is small enough to keep the probability of the mean being outside the range of \$45 to \$55 below 5%.

Recall that the standard error of the mean is given by $SE(\bar{x}) = \sigma/\sqrt{n}$, where σ is the population standard deviation and n is the sample size. The auditors want to find the minimum value of n such that $SE(\bar{x})$ is small enough to keep the probability of the mean being outside the range of \$45 to \$55 below 5%.

We can use the fact that the standard normal distribution is symmetrical to write the probability of the mean being outside the range of \$45 to \$55 as $2 * P(|Z| > (55 - 50)/SE(\bar{x}))$. Setting this probability equal to 0.05 and solving for n, we get $n = (55 - 50)/(SE(\bar{x}) * 2 * Z)$, where Z is the z-score corresponding to the desired probability.

Since the desired probability is 0.05, we can use a standard normal table or calculator to find Z = 1.96. Plugging this value into the equation, we get $n = (55 - 50)/(SE(\bar{x}) * 2 * 1.96) = (55 - 50)/(4 * 1.96) = 250/7.84 \approx 31.75$. Since n must be an integer, the minimum sample size is 32.

The correct answer is therefore A: 144.

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

Ans: Let's examine each of the given statements:

A. The standard deviation of the scores within any sample will be 120.

This is false. The standard deviation of a sample is generally different from the standard deviation of the population from which the sample is drawn. However, if the sample is large enough (typically $n > 30$), the standard deviation of the sample will be approximately equal to the population standard deviation.

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

This is false. The standard deviation of the mean across several samples is not equal to the standard deviation of the population. Instead, it is given by the formula $SE(\bar{x}) = \sigma/\sqrt{n}$, 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 would be $120/\sqrt{n}$.

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

This is false. The mean score in any sample will generally be different from the population mean of 720. However, if the sample is large enough (typically $n > 30$), the mean of the sample will be approximately equal to the population mean.

D. The average of the mean across several samples will be 720.

This is true. If we take the mean of several samples, the average of these means will be equal to the population mean of 720. This is because the mean of a sample is an unbiased estimator of the population mean, meaning that it tends to be close to the true population mean on average.

E. The standard deviation of the mean across several samples will be 0.60.

This is false. The standard deviation of the mean across several samples is not equal to 0.60. Instead, it is given by the formula $SE(\bar{x}) = \sigma/\sqrt{n}$, 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 would be $120/\sqrt{n}$.