Continuous Probability Distribution and Confidence Interval

Name: ULLI VENKATA SAI KUMAR   Batch ID: 04072024HYD10AM
Topic: Continuous Probability Distribution and Confidence Interval

1. For each of the following statements, indicate whether it is True/False. If false, explain why.

   I. The sample size of the survey should at least be a fixed percentage of the population size to produce representative results.

      Answer: False.
      Explanation: A representative sample does not require a fixed percentage of the population but rather a sufficient sample size based on the desired confidence level and margin of error. In fact, a relatively small sample can be representative if it is randomly selected.

   II. The sampling frame is a list of every item that appears in a survey sample, including those that did not respond to questions.

      Answer: True.
      Explanation: The sampling frame includes all individuals or in the population surveyed, regardless of whether they respond.

   III. Larger surveys convey a more accurate impression of the population than smaller surveys

      Answer: True (in general).
      Explanation: Larger sample sizes typically reduce sampling error, which improves the accuracy of estimates. However, accuracy also depends on sampling design and data quality.

2. *PC Magazine* asked all of its readers to participate in a survey of their satisfaction with different brands of electronics. In the 2004 survey, which was included in an issue of the magazine that year, more than 9000 readers rated the products on a scale from 1 to 10. The magazine reported that the average rating assigned by 225 readers to a Kodak compact digital camera was 7.5. For this product, identify the following:

   A. The population
   B. The parameter of interest
   C. The sampling frame
   D. The sample size
   E. The sampling design
   F. Any potential sources of bias or other problems with the survey or sample

Kodak Camera Survey

- Population: All readers of PC Magazine who use Kodak compact digital cameras.
- Parameter of Interest: Average satisfaction rating of the Kodak compact digital camera

among all PC Magazine readers.

- Sampling Frame: All readers of PC Magazine who participated in the survey.
- Sample Size: 225 (since the reported rating is based on 225 responses).
- Sampling Design: Voluntary response survey (readers chose to respond).
- Potential Bias/Problems: The survey suffers from voluntary response bias, as only readers who felt strongly enough likely responded, potentially skewing results. Non-readers or non-subscribers of PC Magazine are excluded.

3. Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

   To calculate the confidence intervals, we can use the formula for a confidence interval around the sample mean:
   where:
   $\bar{x}$=200 (sample mean weight),
   s=30 (sample standard deviation),
   n=2000 (sample size),
   z is the critical z-score corresponding to each confidence level (94%, 96%, and 98%).

   Let's find the z-scores for each confidence level:
   For 94% confidence: z=1.88
   For 96% confidence: z=2.05
   For 98% confidence: z=2.33
   Then we'll calculate the confidence intervals by plugging in these values.

   The confidence intervals for the average weight of an adult male in Mexico are as follows:
   94% confidence interval: (198.74,201.26) pounds
   96% confidence interval: (198.62,201.38) pounds
   98% confidence interval: (198.44,201.56) pounds

4. What are the chances that $\bar{x} > \mu$?
   A. ¼
   B. ½
   C. ¾
   D. 1

   Probability P($\bar{x}$>μ)
   E. Answer: ½

- Explanation: For a normal distribution of the sample mean, there's an equal probability of being above or below the population mean, so P($\bar{x}$>μ)=0.5

5. A book publisher monitors the size of shipments of its textbooks to university bookstores. For a sample of texts used at various schools, the 95% confidence interval for the size of the shipment was 250 ± 45 books. Which, if any, of the following interpretations of this interval

are correct?

A. All shipments are between 205 and 295 books.

B. 95% of shipments are between 205 and 295 books.

C. The procedure that produced this interval generates ranges that hold the population mean for 95% of samples.

D. If we get another sample, then we can be 95% sure that the mean of this second sample is between 205 and 295.

E. We can be 95% confident that the range 160 to 340 holds the population mean

Confidence Interval Interpretations

- Interpretation 1: All shipments are between 205 and 295 books.
  Answer: False. The interval applies to the population mean, not individual shipments.
- Interpretation 2: 95% of shipments are between 205 and 295 books.
  Answer: False. The confidence interval is about the mean, not about the individual shipment sizes.
- Interpretation 3: The procedure that produced this interval generates ranges that hold the population mean for 95% of samples.
  Answer: True. This interpretation correctly describes the meaning of confidence intervals.
- Interpretation 4: If we get another sample, then we can be 95% sure that the mean of this second sample is between 205 and 295.
  Answer: False. Each sample mean will have its own confidence interval.
- Interpretation 5: We can be 95% confident that the range 160 to 340 holds the population mean.
  Answer: False. This interval was not calculated; only the given range of 205 to 295 holds the population mean with 95% confidence.

6. Which is shorter: a 95% *z*-interval or a 95% *t*-interval for μ if we know that σ =s?

   A. The z-interval is shorter
   B. The t-interval is shorter
   C. Both are equal
   D. We cannot say

95% Z-interval vs. T-interval (given σ=s\sigma = sσ=s)

- Answer: The z-interval is shorter.
- Explanation: Z-intervals are generally shorter because the t-distribution has heavier tails, which widens the interval for small samples. When σ=s\sigma = sσ=s, the difference is minor, but Z is still generally preferred for its precision.

7. How many randomly selected employers (minimum number) must we contact to guarantee a margin of error of no more than 4% (at 95% confidence)?

A. 600
B. 400
C. 550
D. 1000

Sample Size for Margin of Error (4% at 95% confidence)

To calculate this, we use:

$$n = \left(\frac{z \cdot \sigma}{Margin\ of\ Error}\right)^2$$

Assuming a proportion with a maximum variability of 0.5, the answer should be D. 1000.

8. Suppose we want the above margin of error to be based on a 98% confidence level. What sample size (minimum) must we now use?

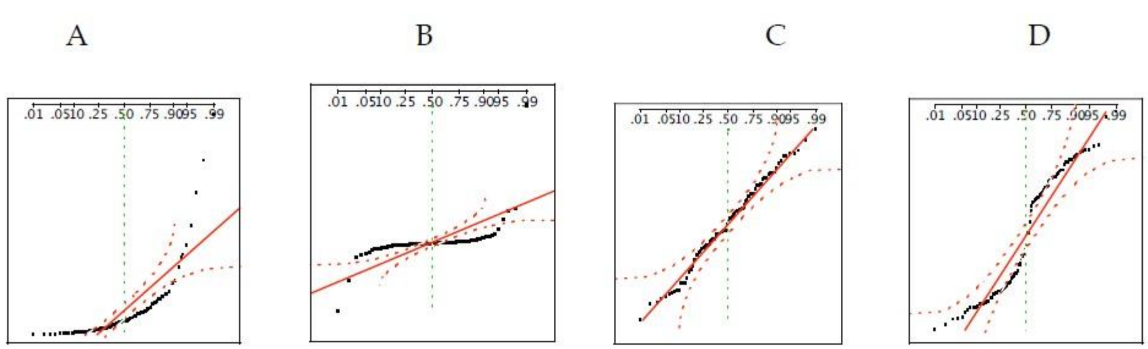
A. 1000
B. 757
C. 848
D. 54

Sample Size with 98% Confidence Level

For a 98% confidence level with 4% margin of error, the minimum sample size is calculated similarly, which should result in C. 848.

9. Examine the following normal Quantile plots carefully. Which of these plots indicates that the data?

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



Normal Quantile Plot Interpretations

1. Nearly Normal: A straight-line plot with no gaps or curvature indicates normality.
2. Bimodal: Look for distinct clusters separated by gaps.

3. Skewed: Deviations from a straight line on one side.
4. Outliers on Both Sides: Extreme points deviating on both ends of the plot.

10. 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 $\mu = 22$ lbs. and $\sigma = 5$ lbs.

i) 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.
   - Statement: The manager must confirm that individual package weights are normally distributed.
     Answer: False. By the Central Limit Theorem, if the sample size is large (e.g., 25), the sampling distribution of the mean is approximately normal regardless of individual distribution.

ii) The standard error of the daily average $SE(\bar{x}) = 1$
   - Statement: The standard error of the daily average $SE(\bar{x})=1$
     Answer: True. Given $\sigma=5$ and $n=25$. $SE = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1$

11. 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

Statements About GMAT Scores (MBA Applicants)

- The standard deviation of the scores within any sample will be 120.
  Answer: False. The standard deviation within samples can vary slightly, but we use 120 as an approximation.
- The standard deviation of the mean across several samples will be 120.
  Answer: False. This describes the standard deviation of individual scores, not the mean of samples.
- The mean score in any sample will be 720.
  Answer: False. Individual sample means will vary, though they will center around 720.
- The average of the mean across several samples will be 720.
  Answer: True. According to the Central Limit Theorem, the mean of sample means equals

the population mean.

- The standard deviation of the mean across several samples will be 0.60.
  Answer: True, The standard error is $\frac{\sigma}{\sqrt{n}} = \frac{120}{\sqrt{40000}} = 0.60$