Topics: Confidence Intervals

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 in order to produce representative results.

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

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

Ans: I. False. The sample size of a survey does not need to be a fixed percentage of the population size in order to produce representative results. In fact, the sample size of a survey should be chosen based on the desired level of precision and the margin of error that is acceptable for the estimate. A larger sample size will generally result in a smaller margin of error and a more precise estimate, but it is not necessary for the sample size to be a fixed percentage of the population size.

II. False. The sampling frame is a list of all the items in the population that are eligible to be included in the sample. It is not a list of every item that appears in the sample, including those that did not respond to questions.

III. True. Larger surveys generally convey a more accurate impression of the population than smaller surveys because they have a smaller margin of error. However, the precision of a survey also depends on the sampling design and the level of response to the survey, as well as the size of the sample. So it is possible for a smaller survey to be more accurate if it is well-designed and has a high response rate.

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

Ans:
A. The population in this survey is all of the readers of PC Magazine.

B. The parameter of interest in this survey is the average satisfaction rating of Kodak compact digital cameras among all of the readers of PC Magazine.
C. The sampling frame for this survey is the list of readers of PC Magazine who participated in the survey.
D. The sample size for this survey is 225, which is the number of readers who rated the Kodak compact digital camera.
E. The sampling design for this survey is not specified, but it could be a simple random sample or a stratified sample, depending on how the readers were selected to participate in the survey.
F. Potential sources of bias or other problems with the survey or sample include:

Self-selection bias: The survey may be subject to self-selection bias because it relies on voluntary participation, which means that only readers who are interested in the topic or who have strong opinions about the products may be more likely to respond. This can lead to a sample that is not representative of the entire population of readers.

Response bias: The survey may be subject to response bias if some readers are more likely to respond than others, or if the way in which the survey is administered or the wording of the questions influences the responses.

Sampling bias: The sample may be subject to sampling bias if the readers who participated in the survey are not representative of the entire population of readers. For example, if the sample is not randomly selected, it may be more likely to include readers who are more or less satisfied with the products than the population as a whole.

Overall, it is important to consider these potential sources of bias and other problems when interpreting the results of the survey.

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

   I. If the 95% confidence interval for the average purchase of customers at a department store is $50 to $110, then $100 is a plausible value for the population mean at this level of confidence.

   II. If the 95% confidence interval for the number of moviegoers who purchase concessions is 30% to 45%, this means that fewer than half of all moviegoers purchase concessions.

   III. The 95% Confidence-Interval for μ only applies if the sample data are nearly normally distributed.

Ans:
I. True. If the 95% confidence interval for the average purchase of customers at a department store is $50 to $110, then $100 is a plausible value for the population mean at this level of confidence. The confidence interval represents the range of values that are likely to include the true population mean, based on the sample data and the level of confidence desired. A value that falls within the confidence interval is considered plausible for the population mean.

II. False. If the 95% confidence interval for the number of moviegoers who purchase concessions is 30% to 45%, this means that there is a 95% chance that the true population proportion of moviegoers who purchase concessions is between 30% and 45%. It does not necessarily mean that fewer than half of all moviegoers purchase concessions.

III. False. The 95% confidence interval for the population mean, μ, can be calculated for any sample data, regardless of whether the data are nearly normally distributed. The confidence interval is based on the sample mean and the standard error of the mean, which are calculated using the sample data. The shape of the sampling distribution of the sample mean depends on the sample size and the population characteristics, but it is not necessary for the sample data to be normally distributed in order to calculate the confidence interval.

4. What are the chances that $\overline{X} > \mu$?

   A. ¼
   B. ½
   C. ¾
   D. 1

Ans: B This is pure assumption. There is a 50% chance that the sample mean is greater than the population mean

5. In January 2005, a company that monitors Internet traffic (WebSideStory) reported that its sampling revealed that the Mozilla Firefox browser launched in 2004 had grabbed a 4.6% share of the market.

   I. If the sample were based on 2,000 users, could Microsoft conclude that Mozilla has a less than 5% share of the market?

   II. WebSideStory claims that its sample includes all the daily Internet users. If that's the case, then can Microsoft conclude that Mozilla has a less than 5% share of the market?

   We are given that WebSideStory claims that its sample includes all the daily Internet users.

Ans.
[I] To determine whether Microsoft can conclude that Mozilla has a less than 5% share of the market based on the sample of 2,000 users, we need to consider the margin of error associated with the sample.
The margin of error is a measure of the precision of an estimate, and it is typically expressed as a percentage of the estimate. The margin of error is

determined by the sample size and the level of confidence desired. A larger sample size and a higher level of confidence will result in a smaller margin of error.
To determine the margin of error for the sample of 2,000 users, we can use the following formula:

margin of error = (z * s) / √n

where:

z is the z-score for the desired confidence level (e.g. 1.96 for 95% confidence)
s is the standard error of the estimate (the standard deviation of the sampling distribution of the estimate)
n is the sample size

Since the sample size is 2,000 and the standard error of the estimate is the standard deviation of the sampling distribution of the estimate, we can calculate the margin of error as follows:

margin of error = (1.96 * 0.046) / √2000

margin of error = 0.0091

The margin of error for the sample of 2,000 users is approximately 0.91%. This means that the sample estimate of 4.6% is accurate within plus or minus 0.91% at a 95% confidence level.

Since the sample estimate of 4.6% is less than 5%, and the margin of error is 0.91%, Microsoft cannot conclude that Mozilla has a less than 5% share of the market based on the sample of 2,000 users.

[II] If the sample used by WebSideStory includes all daily Internet users, then the sample estimate of 4.6% would be the true population proportion of Internet users using the Mozilla Firefox browser. In this case, Microsoft could conclude that Mozilla has a less than 5% share of the market based on the sample estimate of 4.6%.

6. 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.

Ans.

The 95% confidence interval for the size of the shipment is 250 ± 45 books, which means that the interval extends from 250 - 45 = 205 books to 250 + 45 = 295 books. This interval can be interpreted as follows:

A: All shipments are between 205 and 295 books. This interpretation is incorrect. The confidence interval represents a range of values that are likely to include the true population mean, but it is not a guarantee that all shipments fall within this range.

B: 95% of shipments are between 205 and 295 books. This interpretation is also incorrect. The confidence interval represents a range of values that are likely to include the true population mean, but it does not specify the percentage of shipments that fall within this range.

C: The procedure that produced this interval generates ranges that hold the population mean for 95% of samples. This interpretation is correct. The confidence interval represents a range of values that are likely to include the true population mean for 95% of samples, based on the procedure used to calculate the interval.

D: If we get another sample, then we can be 95% sure that the mean of this second sample is between 205 and 295. This interpretation is incorrect. The confidence interval represents a range of values that are likely to include the true population mean for the current sample, but it does not apply to future samples.

E: We can be 95% confident that the range 160 to 340 holds the population mean. This interpretation is incorrect. The confidence interval is 250 ± 45 books, not 160 to 340 books.

So the correct answer is C: The procedure that produced this interval generates ranges that hold the population mean for 95% of samples.

7. 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

Ans.

A. The z-interval is shorter.

The z-interval and t-interval are both methods for constructing confidence intervals for a population mean, μ. The z-interval is based on the normal distribution and the t-interval is based on the t-distribution.

If the population standard deviation, σ, is known, the z-interval is used. If the population standard deviation is unknown and must be estimated from the sample standard deviation, s, the t-interval is used.

In general, the t-interval is wider than the z-interval because it takes into account the uncertainty in estimating the population standard deviation from the sample standard deviation. Therefore, if we know that σ = s, the z-interval will be shorter than the t-interval.

So, the correct answer is A: The z-interval is shorter.

Questions 8 and 9 are based on the following: To prepare a report on the economy, analysts need to estimate the percentage of businesses that plan to hire additional employees in the next 60 days.

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

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

Ans: D-1000
Explanation: To determine the minimum number of employers that must be contacted in order to guarantee a margin of error of no more than 4% at 95% confidence, you can use the following formula:

n = (z^2 * p * (1 - p)) / e^2

where:

n is the minimum sample size
z is the z-score for the desired confidence level (1.96 for 95% confidence)
p is the estimated proportion of employers that meet the desired characteristic (e.g. have a certain policy, offer a certain benefit, etc.)
e is the margin of error
Plugging in the values from the question, we get:

n = (1.96^2 * 0.5 * (1 - 0.5)) / (0.04^2)

Solving for n, we get:

n = (3.8416 * 0.5 * 0.5) / (0.0016)

n = 1.9208 / 0.0016

n = 1195.5

So the minimum number of employers that must be contacted in order to guarantee a margin of error of no more than 4% at 95% confidence is approximately 1196.

The correct answer is therefore D: 1000.

9. 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. 543

Ans: A 1000
To determine the minimum sample size needed to achieve a margin of error of no more than 4% at 98% confidence, we can use the same formula as before, but with a different value for the z-score:

n = (z^2 * p * (1 - p)) / e^2

where:

n is the minimum sample size
z is the z-score for the desired confidence level (2.33 for 98% confidence)
p is the estimated proportion of employers that meet the desired characteristic (e.g. have a certain policy, offer a certain benefit, etc.)
e is the margin of error
Plugging in the values from the question, we get:

n = (2.33^2 * 0.5 * (1 - 0.5)) / (0.04^2)

Solving for n, we get:

n = (5.4289 * 0.5 * 0.5) / (0.0016)

n = 2.7144 / 0.0016

n = 1696.5

So, the minimum sample size needed to achieve a margin of error of no more than 4% at 98% confidence is approximately 1697.

The correct answer is therefore A: 1000.