

## 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.

**Ans:** False.

The sample size of a survey should be determined by statistical power calculations that take into account factors such as the level of precision required, the variability in the population, and the level of confidence desired in the estimates. The sample size does not necessarily need to be a fixed percentage of the population size to produce representative results. In fact, a smaller sample size can be representative of the population if it is properly selected and is large enough to achieve the desired level of precision and confidence in the estimates. However, a larger sample size generally increases the precision of the estimates and reduces the sampling error.

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

**Ans:** False

The sample should contain only those items that have responded because the items with no response adds no value the analysis and just increases the size of data.

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

**Ans:** True

More the size of sample, it captures maximum patterns that helps to build an accurate model

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 - The population in this case would be all the consumers who purchased and used Kodak compact digital cameras.
- B. The parameter of interest - The parameter of interest is the average rating of all Kodak compact digital cameras in the population.
- C. The sampling frame - The sampling frame could be a list of all customers who purchased Kodak compact digital cameras.
- D. The sample size - The sample size is 225 readers who participated in the survey.

- E. The sampling design - The sampling design is not specified in the problem, but it could be a simple random sample of the population of Kodak compact digital camera owners.
- F. Any potential sources of bias or other problems with the survey or sample - There could be potential sources of bias in the survey if the sample is not representative of the population, or if the readers who participated in the survey had different characteristics or opinions than those who did not participate. Additionally, the wording of the survey question or the way in which the respondents were selected or contacted could also introduce bias.

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.

Ans: True.

The confidence interval is constructed in such a way that the true population mean is expected to fall within it with a 95% probability. Therefore, \$100 falls within the interval, and it 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.

Ans: False.

The confidence interval gives a range of plausible values for the true proportion of moviegoers who purchase concessions, not an estimate of the actual proportion. Since the interval includes 45%, it is possible that the true proportion of moviegoers who purchase concessions is 45%, which is more than half of all moviegoers.

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

Ans: False.

The 95% confidence interval for  $\mu$  (population mean) can be constructed using the sample data, regardless of the distribution of the sample mean, as long as the sample size is large enough (typically,  $n \geq 30$ ) due to the Central Limit Theorem. However, if the sample size is small and the population distribution is non-normal, a confidence interval based on the t-distribution may be more appropriate.

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

- A.  $\frac{1}{4}$
- B.  $\frac{1}{2}$

C.  $\frac{3}{4}$

D. 1

Ans: B.  $\frac{1}{2}$

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?

Ans:

(I) Let  $p$  = population proportion share of the market by Mozilla

So, **Null Hypothesis**,  $H_0 : p \geq 5\%$  {means that Mozilla has more than or equal to 5% share of the market}

**Alternate Hypothesis**,  $H_A : p < 5\%$  {means that Mozilla has a less than 5% share of the market}

The test statistics that will be used here is **One-sample z-test for proportions**;

$$\text{T.S.} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

where,  $\hat{p}$  = sample proportion of the share of the market grabbed by Mozilla in 2004 = 4.6%

$n$  = sample of users = 2,000

$$\begin{aligned} \text{So, the test statistics} &= \frac{0.046 - 0.05}{\sqrt{\frac{0.05(1-0.05)}{2,000}}} \\ &= -0.821 \end{aligned}$$

The value of z-test statistics is -0.821.

Since in the question we are not given with the level of significance so we assume it to be 5%. **Now, at 5% level of significance the z table gives a critical value of -1.96 for left-tailed test.**

Since the value of our test statistics is more than the critical value of z, so we have insufficient evidence to reject our null hypothesis as it will not fall in the rejection region.

**Therefore, we conclude that Mozilla has more than or equal to 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?

Ans:

(II) We are given that WebSideStory claims that its sample includes all the daily Internet users. This means that the 4.6% share of the market represents the whole population.

**Hence, we can conclude that Mozilla has a less than 5% share of the market.**

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  $250 \pm 45$  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 correct interpretation of the given 95% confidence interval for the size of shipments of textbooks 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  $\mu$  if we know that  $\sigma = s$ ?

- 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

Because,  $z$  it tell us difference between mean of distribution and data points in standard deviation .

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:**

To determine the minimum number of randomly selected employers required to guarantee a margin of error of no more than 4% at 95% confidence, we can use the following formula:

$$n = (z^* \sigma / E)^2$$

Where:  $z = 1.96$  (the  $z$ -value corresponding to 95% confidence)  $\sigma = 0.5$  (since we do not know the true proportion, we assume a conservative estimate of 0.5)  $E = 0.04$  (the desired margin of error)

Plugging in the values, we get:

$$n = (1.96 * 0.5 / 0.04)^2$$

$$n = 600.25$$

Rounding up to the nearest integer, we get the minimum number of employers to contact is 601, which is option A. Therefore, the answer is A.

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:

$$Z = 2.576$$

$$0.04 = 2.326 * \sqrt{(0.5 * 0.5) / n}$$

$$n = (2.326^2 * 0.5 * 0.5) / (0.04)^2$$

$$= 1.3525 / 0.0016$$

$$= 845.35 \approx 848$$

Option C is correct