

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.

Answer: **FALSE**, the survey will be better visualized only when it is having a sample size of fixed percentage ... **but** also, we have to make sure that the sample have good confidence interval, margin of error and also variation of data's within dataset.

- 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: **FALSE**, yes it contains all the data contained in the populationIn idle situation ONLY population and sampling frame are same

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

Answer: **TRUE**, yes larger survey provides near to accurate information.

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:** All the readers of PC Magazine.

B. **The parameter of interest:** Here the parameter of interest is the rating given by the readers who use the Kodak compact Digital Camera.

C. **The sampling frame:** The sampling frame is all the readers of the PC Magazine.

D. **The sample size:** The sample size is approximately 9000 for all electronic items listed in question, Kodak compact Digital Camera specific sample is not listed

E. **The sampling design:** Its more like Census based sampling

F. **Any potential sources of bias or other problems with the survey or sample :** Here problems will be only few readers will answer it and out of it there may be chance of non-response because the product he opted may not be there....called as non-response bias Also due to their likeliness or closeness to magazine, sometimes the reader may give good rating because of his relationship with magazine...called response 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.

Answer: **True**, if 95% of confidence interval covers from range of \$50 to \$110, then \$100 is a plausible value for population mean at 95% 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.

Answer: **True** in most cases, Higher Confidence Interval means higher sample size ...so a sample of 95% will give more reliable information, hence the case of purchase concessions of 30% to 45% as given in sample size of 95% confidence interval will suit to whole population too.

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

Answer: **False**, as per the Central Limit Theorem.. when the sample size is large the sampling mean of the sampling distribution becomes approximately distributed regardless of distribution of population...Here 95% confidence interval will definitely have large sample size. Hence it applies.

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

- A. $\frac{1}{4}$
- B. $\frac{1}{2}$
- C. $\frac{3}{4}$
- D. 1

Answer : C

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?

Answer:

REFER: Basic_stats2 Q5.IPYNB

```
n=2000, pnull=0.05, palt=0.046
sm.stats.proportions_ztest(palt*n,n,pnull,alternative='smaller')
(z_value=-0.8539292422983582, p_value0.19657209793658192)
```

Here P_{null} is less than p value , so reject null hypothesis and we do not reject P alternate... Hence the mozilla captures 4.6% market of internet share based on sample of 2000.

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?

Answer: If the WebSideStory claims are true It means it is a good sample ...which includes all type of internet users and this can consider as equivalent to population. Or we can in this perspective it took a complete census. And thus, Microsoft can conclude Mozilla has a less than 5% share of population.

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.

Answer: Here they have given a sample which has 95% confidence having a sample mean as 250 with its confidence intervals lower limit and higher limit are (205,295).

Based on that statement

- A. **False** , we cannot consider all populations range will be in between 205 and 295. It just result of a sample.
- B. **False** , because it not of quantity and its details rather it is a confidence interval to find mean of sample which is used to find the approximate population mean ...Mean can be same with different datasets ... because sample has less data set and population has a large data set so range can be something different ...
- C. **True**, this statement almost reflects the real purpose of the confidence interval. Also this can be conformed by drawing more type of similar samples.
- D. **False** , as we draw more and more samples we can be more confident of the men lying between 205 and 295... we cant be sure by one more sample.
- E. **True** , since 160 to 340 has the interval of 205 to 295 , we can be more sure the population mean will lie in this interval

7. Which is shorter: a 95% z-interval or a 95% t-interval for μ 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

Answer : A. Since t interval tails are longer.

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

Answer : (A)

To calculate the minimum number of samples for the given confidence interval and margin of error, the formula is

$$n = \frac{Z^2 * p * (1 - p)}{E^2}$$

Where n=No of minimum samples required ; Z= Z score associated with desired confidence interval (1.96 for 95% confidence) ; p = 0.5 (we use when we are uncertain) ; E-Margin of Error(0.04-4%).

Here ,

$$n = \frac{1.96^2 * 0.5 * (1 - 0.5)}{0.04^2}$$
$$= 600.25$$

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

Answer : (C)

To calculate the minimum number of samples for the given confidence interval and margin of error, the formula is

$$n = \frac{Z^2 * p * (1 - p)}{E^2}$$

Where n=No of minimum samples required ; Z= Z score associated with desired confidence interval (2.33 for 98% confidence) ; p = 0.5 (we use when we are uncertain) ; E-Margin of Error(0.04-4%).

Here ,

$$n = \frac{2.33^2 * 0.5 * (1 - 0.5)}{0.04^2}$$

=848