

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.

Sol:- The survey should have a specific sample size, a fixed percentage of population that helps to analyse the data well give more results. The statement is TRUE

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

Sol:- FALSE Because if the survey without no response adds no value to the analysis and just it increase the size of the data.

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

Sol:- TRUE the survey is data is larger then accuracy also more compared to smaller surveys.

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

Sol:- The total population of the data of readers is more then 9000.

B. The parameter of interest

Sol:- The mean of the data is 7.5

C. The sampling frame

Sol:- The sampling frame is total readers are more than 9000 readers.

D. The sample size

Sol:- The sample size of the data is 225.

E. The sampling design

Sol:- The selection of sample is from a survey that is filtered is KODAK COMPACT DIGITAL CAMERA.

F. Any potential sources of bias or other problems with the survey or sample

Sol:- The Potential sources of the bias is from PC magazine.

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.**

Sol: True

- 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.

Sol: True

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

Sol:- FALSE because the central limit theorem implies a normal sampling distribution regardless of the distribution of the data.

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

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

Sol:- $P(A) = \frac{1}{4}$

$P(B) = \frac{1}{2}$

$P(C) = \frac{3}{4}$

$P(D) = 1$

THEN $P(a) = 1 - P(A) = \frac{1}{2}$

$P(b) = 1 - P(B) = \frac{2}{3}$

$P(c) = 1 - P(C) = \frac{3}{4}$

$P(d) = 1 - P(D) = 0$

$P(\text{Problem will be solved}) = 1 - P(a) P(b) P(c) P(d)$

$1 - \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times 0 = 1$

Then the answer is D

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.

Sol:

The screenshot shows the Spyder Python IDE interface. The main editor displays a Python script for a z-test. The script includes a docstring with metadata, imports for scipy, numpy, and scipy.stats, and a calculation of a z-score and p-value based on a sample mean of 0.046, a population mean of 0.05, a standard deviation of 0.05, and a sample size of 2000. The IPython console on the right shows the execution of the script, displaying the calculated z-score and p-value.

```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Thu Dec 29 15:28:22 2022
4
5 @author: vaishnav
6 """
7
8 from scipy import stats
9 import numpy as np
10 from scipy.stats import norm
11
12 #applying one sample one tail z test
13
14 z_score = (0.046-0.05)/(np.sqrt((0.05*(1-0.05))/2000))
15 z_score
16
17 #then probability assuming null hypothesis, as to compare with Type-1 error  $\alpha = 0.05$ 
18
19 p_value=1-stats.norm.cdf(abs(z_score))
20 p_value
21
22
```

Python 3.9.12 (main, Apr 4 2022, 05:22:27) [MSC v.1916 64 bit (AMD64)]
Type "copyright", "credits" or "license()" for more information.
IPython 8.2.0 -- An enhanced Interactive Python.

In [1]:
....: from scipy import stats
....: import numpy as np
....: from scipy.stats import norm

In [2]: z_score = (0.046-0.05)/(np.sqrt((0.05*(1-0.05))/2000))
....: z_score
Out[2]: -0.828782681668124

In [3]:
....: p_value=1-stats.norm.cdf(abs(z_score))
....: p_value
Out[3]: 0.28588583245107104

In [4]:

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

Sol: Mozilla has more than 5 percent or has equal to 5% share in 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?

Sol: Yes, 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 ± 45 books. Which, if any, of the following interpretations of this interval are correct?

A. All shipments are between 205 and 295 books.

Sol: Incorrect

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

Sol: correct

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

Sol: correct

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

Sol: correct

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

Sol: Incorrect

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

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.

Sol:- The option is A because it difference between mean of distribution and data points in standard deviation.

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**

Sol: 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**

Sol: D