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.
   False -- The sample size of a survey does not necessarily need to be a fixed percentage of the population size 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.
   False --The sampling frame is a list of all the items or people within a population that could potentially be selected for the survey. It does not include those who did not respond to the survey

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

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 is all the readers of PC Magazine

   B. The parameter of interest
   the satisfaction level of the readers with a Kodak compact digital camera measured on a scale from 1 to 10

   C. The sampling frame
   list of PC Magazine readers who participated in the survey

   D. The sample size
   225 readers who rated the Kodak compact digital camera

   E. The sampling design
   self-selected sample as the magazine asked all of its readers to participate, and those who responded are the ones included in the sample

   F. Any potential sources of bias or other problems with the survey or sample
   who chose to respond to the survey might not be representative of all PC Magazine readers and Another potential issue is nonresponse bias if the response rate was low

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.
  TRUE -- $100 falls within the given confidence interval

  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.
  FALSE -- Based on the sample data we are 95% confident that the true proportion of moviegoers who purchase concessions falls within this range. It's possible that the true proportion is greater than half of all moviegoers

  III. The 95% Confidence-Interval for μ only applies if the sample data are nearly normally distributed.
  FALSE -- The Central Limit Theorem states that for large sample sizes (typically n ≥ 30), the sampling distribution of the sample mean will be approximately normally distributed, regardless of the distribution of the population This allows for the construction of confidence intervals for the population mean even when the underlying data are not normally distributed as long as the sample size is sufficiently large

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

    A. ¼
    B. ½
    C. ¾
    D. 1
      Ans : B. ½
      There is 50% chance that the sample mean is greater than 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?

```
# H0: p >= 0.05
# H1: p < 0.05
from scipy import stats
z_score=(0.046-0.05)/((0.05*(1-0.05)/2000)**0.5)
z_score
    -0.820782681668124
p_value=stats.norm.cdf(z_score)
 p_value
      0.20588503245107104
```

Microsoft can conclude that Mozilla has a less than 5% share of the market because p-value is less than 0.05

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?

In this case, there would be no need for hypothesis testing, as the sample accurately represents the entire population. Therefore, if WebSideStory's claim is true, Microsoft could indeed conclude that Mozilla has a less than 5% share of the market, based on WebSideStory's data

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.
   Option C is correct

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
   Option A is Correct
   the z-interval uses the known population standard deviation, resulting in a smaller critical value and thus a shorter interval and heavier tails of the t-distribution, which necessitate larger critical values and wider intervals

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

Option A is correct
    confidence=0.95
    z=stats.norm.ppf(0.975) #z value for 95% confidence
    z

      1.959963984540054
    e=0.04 #margin of error
    p=0.5
    n=z**2*p*(1-p)/e**2
    n

      600.2279407334571

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

Option C is correct

confidence=0.98
z=round(stats.norm.ppf(0.99),2) #z value for 99% confidence
z

    2.33
n=z**2*p*(1-p)/e**2
n

    848.265625