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.

      TRUE; The sample size of a survey should generally be a fixed percentage of the population size to ensure that the results are representative.

   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 Sample Frame is List of an item which responds to the question and not the ones which do not respond to the questions

   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=**ALL READERS OF PC MAG**
   B. The parameter of interest=**RATING ELECTRONICS**
   C. The sampling frame=**READERS WHO RATED PRODUCTS(~9000)**

D. The sample size= **255 READERS**
E. The sampling design = **SUREVEY RESPONSES**
F. Any potential sources of bias or other problems with the survey or sample

Ans =(SELECTION OF READERS,SELECTION OF PRODUCT BRANDS)
The key to random selection is that there is no bias involved in the selection of the sample, but surveys conducted by the magazines often suffer from nonresponse bias and also the source of this data is from readers that read PC magazine vs the whole population that do not read PC magazine yet they use these different brands of electronics.

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

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.
TRUE it indicates 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.

False: The 95% confidence Interval for population mean can be applied to distribution that aren't normal, but they are easy to understand in symmetric distributions.
We can use the central limit theorem to make a normal approximation.
The rule where about 95% of observations are within 2 standard

deviations of the mean is only approximately true. However, it holds very well for the normal distribution.

The mean of data tends to be normally distributed when 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 (This is only an assumption, because if we consider more than 50% for sample mean to be greater than the probability of getting a lower value gets overshadowed because sample mean has an equal chance to be lesser 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?

```
import numpy as np
from scipy import stats
from scipy.stats import norm
# Apply One-Sample One-Tail z-test
z_scores=(0.046-0.05)/(np.sqrt((0.05*(1-0.05))/2000))
z_scores
-0.820782681668124
```

Now, at 5% level of significance the z table gives a critical value of -1.96 for left-tailed test.

      -0.82078>-1.96

hence we accept null hypothesis that Mozilla has greater market share than 5%

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.

This means that the 4.6% is the population percentage. Comparing it with Microsoft's claim that Mozilla has a less than 5% share of the whole market is True.

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

   A. All shipments are between 205 and 295 books.
      Ans: Correct

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

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

      Ans: Correct

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

Ans: Incorrect: as we increase the range +/- 1 sigma the % of confidence increases for a normal distribution it will be 97.5% that the mean will lie in between 160 to 340 (-3 sigma to +3 sigma)

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. OPTION ;A
z-interval is difference between mean of distribution and data points in standard deviation . 95 % confidence interval for mean is shorter for z-interval because t-critical is greater than z-critical value. Yes, z-interval is always shorter because t-critical value cannot be smaller than z-critical value.

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

Confidence Level	z^* Multiplier
90%	1.645
95%	1.960, often rounded to 2
98%	2.327
99%	2.576

Ans : We use p as 0.5

Finding Sample Size for Estimating a Population Proportion

$$n = \left(\frac{z^*}{M}\right)^2 \tilde{p}(1 - \tilde{p})$$

M is the margin of error

$\tilde{p}$ is an estimated value of the proportion

N= $(1.960/.04)^2(0.5)(1-0.5)=600.25$

Hence **A 600** is correct answer

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:**option c**
z value for 98% at confidence ~ 2.32
 P=0.5
Recalculating with the above formula we have n = 848.27
Hence we need to contact minimum 848 employers