1. For each of the following statements, indicate whether it is True/False. If false, explain why.
2. The sample size of the survey should at least be a fixed percentage of the population size in order to produce representative results.

Ans: - True

1. 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 effect of non-response is that is reduces the sample size. Due to the smaller sample size, the precision of estimators will be smaller. The margins of error will be larger.

A more serious effect of non-response is that it can be selective. This occurs if, due to non-response, specific groups are under- or over-represented in the survey. If these groups behave differently with respect to the survey variables, this causes estimators to be biased that means estimates are significantly too high or too low.

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

Ans: - True

1. *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:
2. The population

Ans: - All readers of the magazine.

1. The parameter of interest

Ans: - The population mean rating the Kodak camera

1. The sampling frame

Ans: - All readers of the magazine

1. The sample size

Ans: - Unknown. Because it is mentioned that more than 8000 but not exact number.

1. The sampling design

Ans: - voluntary response, not random.

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

Ans: - Here we see **voluntary response bias** because the sample members are self-selected volunteers who read the PC magazine.

1. For each of the following statements, indicate whether it is True/False. If false, explain why.
2. 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.

1. 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: - True.

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

Ans: - False.

The 95% confidence interval for *μ* will be applicable even though if the sample data are not exactly normally distributed. But if the sample data are nearly normally distributed then t-interval for *μ* works best.

1. What are the chances that ?
2. ¼
3. ½
4. ¾
5. 1

Ans: - B.

The chances are 0.50 due to assuming the sample size condition is true which is sample size is larger than 10 times the absolute kurtosis.

1. 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.
2. If the sample were based on 2,000 users, could Microsoft conclude that Mozilla has a less than 5% share of the market?

Ans: - No, Microsoft can’t conclude that Mozilla has less than 5% share on the market. Using a 95% confidence interval the actual share could range from 3.62% to 5.58%. In this case Microsoft can **only** conclude that they are 95% certain that Mozilla’s share of the market is between 3.62% and 5.58% of the market.

Standard error = = 0.006.

For 95% confidence level z value is 1.96.

So standard error for mean = .046 ± (1.96) \*(.006) = [0.0362, 0.0558]

1. 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: - No, Microsoft can’t conclude that Mozilla has less than 5% share on the market. Using a 95% confidence interval the actual share could range from 4.59% to 4.61%. In this case Microsoft can **only** conclude that they are 95% certain that Mozilla’s share of the market is between 4.59% and 4.60% of the market.

Assume there are 20M users using internet.

Standard error = = 0.000038.

For 95% confidence level z value is 1.96.

So standard error for mean = .046 ± (1.96) \*(.0000038) = [0.04592, 0.04607]

1. 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?
2. All shipments are between 205 and 295 books.

Ans: - All shipments are not between 205 and 295 because due to 95% Confidence Interval for the size of the shipment but not for all shipments. Hence above statement is not correct.

1. 95% of shipments are between 205 and 295 books.

Ans: - 95% CI = 250 45. So, 95% shipments are between 205 and 295. Hence above statement is correct.

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

Ans: - This statement is not correct.

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

Ans: - This statement is correct. Because the 95% confidence interval mean for any sample is between 205 to 295.

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

Ans: - This statement is not correct as per answer provided in question (B).

1. Which is shorter: a 95% *z*-interval or a 95% *t*-interval for *μ* if we know that σ =s?
2. The z-interval is shorter
3. The t-interval is shorter
4. Both are equal
5. We cannot say

Ans: - C

Actually, it depends on standard deviation. But in case of σ =s, both are equal because the formulae are same. Though t distribution influenced by degrees of freedom, it won’t come into picture when calculate the interval.

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.

1. 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)?
2. 600
3. 400
4. 550
5. 1000

Ans: - A

**Explanation: -**

Because there is no estimate of the proportion given, we use  = 0.50 for a conservative estimate.

For a 95% confidence interval, z = 1.960.

The formula for estimating population proportion is n = (z/M)2 (1-) = (1.96/0.04)2 0.5(1-0.5) = 600.25, here M = Margin of error.

1. Suppose we want the above margin of error to be based on a 98% confidence level. What sample size (minimum) must we now use?
2. 1000
3. 757
4. 848
5. 543

Ans: - C

**Explanation: -**

For a 98% confidence interval, z = 2.326.

The formula for estimating population proportion is n = (z/M)2 \* (1-) = (2.326/0.04)2 \* 0.5(1-0.5) = 845.355 here M = Margin of error.

1. Examine the following normal Quantile plots carefully. Which of these plots indicates that the data?
2. Are nearly normal?

Ans: - The plot given (C) is nearly normal because the plotted points lie in and around the diagonal reference line. Remaining plots are not normal.

1. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.)

Ans: - The plot given (D) is bimodal because there is a gap in the adjacent data points at the middle.

1. Are skewed (i.e. not symmetric)?

Ans: - The plot given (B) is skewed because the data points are drifting away from the diagonal reference line.

1. Have outliers on both sides of the center?

Ans: - The plot given (A) has outliers both the sides. Because the data points are drifting away from the diagonal reference line and there is a gap in extreme end points at both sides.



1. For each of the following statements, indicate whether it is True/False. If false, explain why.

The manager of a warehouse monitors the volume of shipments made by the delivery team. The automated tracking system tracks every package as it moves through the facility. A sample of 25 packages is selected and weighed every day. Based on current contracts with customers, the weights should have μ = 22 lbs. and σ = 5 lbs.

1. Before using a normal model for the sampling distribution of the average package weights, the manager must confirm that weights of individual packages are normally distributed.

Ans: - True.

1. The standard error of the daily average SE() = 1.

Ans: - True.

Standard error = = = = 1

1. Auditors at a small community bank randomly sample 100 withdrawal transactions made during the week at an ATM machine located near the bank’s main branch. Over the past 2 years, the average withdrawal amount has been $50 with a standard deviation of $40. Since audit investigations are typically expensive, the auditors decide to not initiate further investigations if the mean transaction amount of the sample is between $45 and $55. What is the probability that in any given week, there will be an investigation?
2. 1.25%
3. 2.5%
4. 10.55%
5. 21.1%
6. 50%

Ans: - D

μ = 50, σ = 40, n = 100

Probability of mean of sample = P (45 < < 55) = P ( < Z < )

= P ( -1.25 < Z < 1.25)

= P (Z < 1.25) – P (Z < - 1.25)

= 0.8944 – 0.1056

= 0.7888

So, the probability of investigation for the given week = 1 – 0.7888 = 0.2112 i.e. 21.1%.

1. The auditors from the above example would like to maintain the probability of investigation to 5%. Which of the following represents the minimum number transactions that they should sample if they do not want to change the thresholds of 45 and 55? Assume that the sample statistics remain unchanged.

1. 144
2. 150
3. 196
4. 250
5. Not enough information

Ans: - D

If we calculate the number of withdrawal transactions with above formula we will get around 250.

1. An educational startup that helps MBA aspirants write their essays is targeting individuals who have taken GMAT in 2012 and have expressed interest in applying to FT top 20 b-schools. There are 40000 such individuals with an average GMAT score of 720 and a standard deviation of 120. The scores are distributed between 650 and 790 with a very long and thin tail towards the higher end resulting in substantial skewness. Which of the following is likely to be true for randomly chosen samples of aspirants?
2. The standard deviation of the scores within any sample will be 120.
3. The standard deviation of the mean of across several samples will be 120.
4. The mean score in any sample will be 720.
5. The average of the mean across several samples will be 720.
6. The standard deviation of the mean across several samples will be 0.60

Ans: - A

Population N =40000.

With an average GMAT score of 720, μ =720

And a standard deviation of 120, σ = 120

Scores distributed between 650 to 790 and rightly skewed.

For randomly chosen samples of aspirants’ true statement is given by option A.

Which is “The standard deviation of the scores within any sample will be 120.