

Basic Statistics (Module – 4 (Part – 2))

Name: Nukala Ayyappa Bharthwaj

Batch Id: DSWDMCOS 21012022

Topic: Basic Statistics (Module – 4 (Part – 2))

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.

Ans: True. Sample size be more than or equal to 30 for a medium type of data set, if the sample size is not appropriate then the interpretations on the will not be more accurate which might result in the failure of the analysis.

- II. 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. If the responses are not given in the survey that will not be taken as consideration for the analysis of the data in the sampling.

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

Ans: True. By conducting larger surveys we can get the better results of the output so that there will be good accuracy in the data but the bigger disadvantage with the larger surveys is that it is a time consumption task.

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: **Ans :** 9000
- B. The parameter of interest - **Ans:** Ratings
- C. The sampling frame – **Ans:** 2004 Survey
- D. The sample size – **Ans:** 225
- E. The sampling design – **Ans:** Finding probability using Z table
- F. Any potential sources of bias or other problems with the survey or sample
Ans: Problem might be with the sample because only for Kodak higher rank is identified their might be a other product which will be having higher ratings.

Q3) Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%, 98%, 96% confidence interval?

Ans: Confidence Interval for 94% is 200 ± 1.17 .

Confidence Interval for 98% is 200 ± 1.56 .

Confidence Interval for 96% is 200 ± 1.39 .

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

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

Ans: A $\frac{1}{4}$. The chances of getting that $X > \mu$ should be less so that there should be less error in the analysis.

5 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?

Ans: Correct statement for the above confidence interval is option "C" & "D".

- 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.

- 6 Which is shorter: a 95% z-interval or a 95% t -interval for μ if we know that $\sigma = s$?

Ans: B. the t -interval will be shorter.

- 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.

- 7 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)?

Ans: C (550)

- A. 600
- B. 400
- C. 550
- D. 1000

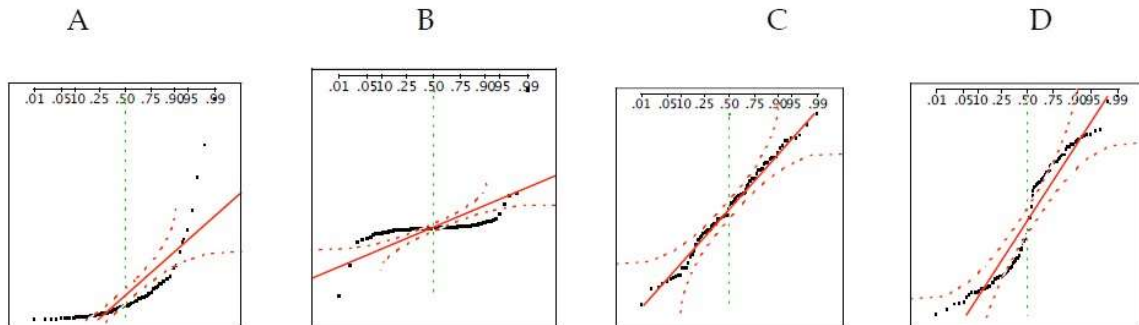
- 8 Suppose we want the above margin of error to be based on a 98% confidence level. What sample size (minimum) must we now use?

Ans: D (543)

- A. 1000
- B. 757
- C. 848
- D. 543

9 Examine the following normal Quantile plots carefully. Which of these plots indicates that the data?

1. Are nearly normal? **Ans: C**
2. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.) **Ans: D**
3. Are skewed (i.e. not symmetric)? **Ans: A**
4. Have outliers on both sides of the center? **Ans: B**



10 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 $\mu = 22$ lbs. and $\sigma = 5$ lbs.

- i) 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.

- ii) The standard error of the daily average $SE(\bar{x}) = 1$

Ans: True. Formulae is $\sigma/\sqrt{n} = 5/\sqrt{25} = 1$.

- 11 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?

Ans: D

- A. The standard deviation of the scores within any sample will be 120.
- B. The standard deviation of the mean of across several samples will be 120.
- C. The mean score in any sample will be 720.
- D. The average of the mean across several samples will be 720.
- E. The standard deviation of the mean across several samples will be 0.60

Hints:

1. Business Problem
 - 1.1. Objective
 - 1.2. Constraints (if any)
2. For each assignment the solution should be submitted in the below format
3. Research and Perform all possible steps for obtaining solution
4. For Basic Statistics explanation of the solutions should be documented in black and white along with the codes.

One must follow these guidelines as well:

- 4.1. Be thorough with the concepts of Probability, Central Limit Theorem and Perform the calculation stepwise
 - 4.2. For True/False Questions, explanation is must.
 - 4.3. R & Python code for Univariate Analysis (histogram, box plot, bar plots etc.) for data distribution to be attached
5. All the codes (executable programs) should execute without errors