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

1. For each of the following statements, indicate whether it is True/False. If false, explain why.
   1. 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 representation of the survey results should have a sample size. The sample size must be a fixed percentage of the total population size of the survey.

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

1. False: The sampling frame refers to a list of an item which responds to the question and not the ones which do not respond to the questions.
   1. Larger surveys convey a more accurate impression of the population than smaller surveys.
2. True: The larger conveys a more accurate impression of the population as larger surveys involve large sample size which reduces the chances of error.
3. *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:
4. The population

9000

1. The parameter of interest

Customer rating

1. The sampling frame

SRS

1. The sample size
2. 225
3. The sampling design

Rating from 1 to 10

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

Review result will be different with different samples

Small sampling size may lead to misinterpretation

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?

Sample **mean**of  Description: https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%3D%20200.

Sample **standard deviation** of Description: https://tex.z-dn.net/?f=s%20%3D%2030.

Sample **size**of Description: https://tex.z-dn.net/?f=n%20%3D%202000.

We use t distribution here.

The **interval** is:

Description: https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%5Cpm%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D

In which **t** is the critical value for the two-tailed confidence interval.

Considering a **94%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 1.8916**, hence:

Description: https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20-%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20-%201.8916%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20198.73

Description: https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%2B%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20%2B%201.8916%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20201.27

The **94%** confidence interval is **(198.73, 201.27).**

Considering a **96%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.0673**, hence:

Description: https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20-%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20-%202.0673%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20198.61

Description: https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%2B%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20%2B%202.0673%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20201.39

The **96%** confidence interval is **(198.61, 201.39).**

Considering a **98%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.3452**, hence:

Description: https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20-%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20-%202.3452%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20198.43

Description: https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%2B%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20%2B%202.3452%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20201.57

The **98%** confidence interval is **(198.43, 201.57).**

The confidence intervals are,

* The **94%** confidence interval is **(198.73, 201.27).**
* The **96%** confidence interval is **(198.61, 201.39).**
* The **98%** confidence interval is **(198.43, 201.57).**

1. What are the chances that

*X*  ** ?

1. ¼
2. ½
3. ¾
4. 1

Ans: Option D

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.
3. 95% of shipments are between 205 and 295 books.
4. The procedure that produced this interval generates ranges that hold the population mean for 95% of samples.
5. If we get another sample, then we can be 95% sure that the mean of this second sample is between 205 and 295.
6. We can be 95% confident that the range 160 to 340 holds the population mean.
7. Which is shorter: a 95% *z*-interval or a 95% *t*-interval for *μ* if we know that σ =s?
8. The z-interval is shorter
9. The t-interval is shorter
10. Both are equal
11. We cannot say

Ans: Option A





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)?

A. 600

B. 400

C. 550

D. 1000

Ans:-A. Because there is no estimate of the proportion given, we use for a conservative estimate. This is the minimum sample size, therefore we should round up to **600**. In order to construct a 95% confidence interval with a margin of error of 4%, we should obtain a sample of at least 600.

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?

A. 1000

B. 757

C. 848

D. 543

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

C

1. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the

spacing of adjacent data values.)

D

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

A

1. Have outliers on both sides of the center?

B



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.

TRUE. In this case, at least 30 sample packages must be selected and weighed everyday. Based on the central limit theorem, the sampling distribution of the sample mean approach normal distribution as the sample size become bigger (over 30).

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

TRUE. Standard error equal to standard deviation divided by square root of sample size = 5/sqrt(25) =1



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

**Hints:**

1. Business Problem
   1. Objective
   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:

* 1. Be thorough with the concepts of Probability, Central Limit Theorem and Perform the calculation stepwise
  2. For True/False Questions, explanation is must.
  3. R & Python code for Univariate Analysis (histogram, box plot, bar plots etc.) for data distribution to be attached

1. All the codes (executable programs) should execute without errors