# 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.

Ans – False

Reason: It is very important to use a correct sample size. When your sample is too big, this will lead to unnecessary waste of money and time. On the other hand, when it’s too small, your results will not be statistically [significant](http://en.wikipedia.org/wiki/Significance_level) and you will not come to reliable conclusions. In order to have confidence that your survey results are representative, it is critically important that you have a large number of randomly-selected participants in each group you survey. So, what exactly is "a large number?" For a 95% confidence level, which means that there is only a 5% chance of your sample results differing from the true population average. But not to be a fix percentage.

* 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- Ture

Reason: the population is generic and the sampling frame is a specific list of all items in the population, Hence the sampling frame include those that did not respond to question.

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

Ans – true

Reason: Smaller surveys convey a more accurate impression of smaller populations and larger surveys convey a more accurate impression of larger populations. A larger sample size should hypothetically lead to more accurate or representative results.

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:
   * 1. The population - All PC Magazine readers
     2. The parameter of interest - 1 to 10
     3. The sampling frame - Around 9000
     4. The sample size - 255
     5. The sampling design - Z distribution
     6. Any potential sources of bias or other problems with the survey or sample

- Selection of the reader, selection of the issue which will contain the survey

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?

1. What are the chances that

*X*  ** ?

1. ¼
2. ½
3. ¾
4. 1

Ans – D, Sample mean is always greater than population mean.

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.

Ans – D

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 - D, we cannot say that which is shorter because it is also depend on sample size .

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. 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

Ans- C. 848

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

Ans – C

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

spacing of adjacent data values.)

Ans –D and B

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

Ans \_ B

1. Have outliers on both sides of the center?

Ans -A



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, In this case, at least 30 sample packages must be selected and weighed every day. 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

Ans- 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

Ans – D.

**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