

## Short Answer Questions:

Q1. What is the central limit theorem? How is the theorem applied to analytics?

### **Central Limit Theorem:**

The Central Limit Theorem in statistics states that, regardless of the distribution of the variable in the population, the sampling distribution of the mean will approximate a normal distribution with sufficiently large sample size.

### **Application of Central Limit Theorem in Analytics:**

- The Central Limit Theorem is at the core of what every data scientist does daily. It makes statistical inferences about data.
- The theorem gives us the ability to quantify the likelihood that our sample will deviate from the population without having to take any new sample to compare it with. We don't need the characteristics of the whole population to understand the likelihood of our sample being representative of it.
- The concepts of the confidence interval and hypothesis testing are based on the CLT. By knowing that our sample mean will fit somewhere in a normal distribution, we know that 68 percent of the observations lie within one standard deviation from the population mean, 95 percent will lie within two standard deviations, and so on.

Q2. What is the role of the sampling distribution of the same mean and proportion?

### **Sampling distribution of the proportion:**

Sampling distribution of the proportion gives you information about the proportions of a population. You would select samples from the population and get the sample proportion. The mean of all the sample proportions that you calculate from each sample group would become the proportion of the entire population.

### **Importance of Using a Sampling Distribution:**

Since populations are typically large in size, it is important to use a sampling distribution so that you can randomly select a subset of the entire population. Doing so helps eliminate variability when you are doing research or gathering statistical data.