1. To assess the statistical significance of an insight, one has to perform the ‘Null Hypothesis Significance Testing’. This allows us to calculate the confidence value of our initial insight. This is a very effective way of checking the statistical accuracy of an insight.
2. The Central Limit Theorem (CLT) is very important in calculating statistics and probability as this allows us to create the shape of the distribution of a large number of samples regardless of how they are distributed. A lot of data in real world is not normalized and CLT is able to normalize even the most widely spread data. Once we are able to build the normally distribute curve, this allows us to then analyze the statistics and probability of the data set.
3. Statistical power is defined as the ability to correctly identify a false Null Hypothesis. This is important when analyzing large data sets and each hypothesis has to be tested. Rejecting these false Null Hypothesis allows us to save a lot of time.
4. Assume every hypothesis and insight that you have for the data is false and you are testing its significance for the first time. This ensures that you set the confidence level for each hypothesis at the same value and hence there is no confirmation bias or any other bias.
5. Confounding Variables are described as the known and unknown relationship in the data collected which may not be true. These variables look like they are true on first glance but are not intrinsically true. Randomizing the data allows us to eliminate this.
6. A/B testing, also known as split testing, compares two sets of data/hypothesis and based on a set metric and determines which one performs better.
7. Confidence intervals give an estimated range where you can estimate the population parameter to lie. It is also provides a level of confidence that the population parameter is within the given range.