

## Assignment – 5

### 1. How do you assess the statistical significance of an insight?

The steps to assess the statistical significance of an insight are as follows:

- Defining the hypothesis – Null hypothesis and the alternate hypothesis
- Choosing a significance level – alpha – usually 5%
- Collecting the data
- Choosing the correct statistical test
- Calculating the test statistic
- Finding the p-value
- Considering the test power
- Calculating confidence intervals
- Check assumptions
- Report results
- Interpretation of results
- Replication

### 2. What is the Central Limit Theorem? Explain it. Why is it important?

The central limit theorem states that the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough, that is, as the sample size becomes larger the distribution of the sample mean will approach a normal distribution.

CLT is important for the following reasons:

- It allows us to use normal probability models to compute probabilities about sample means or sums, even when the original variable is not normally distributed.
- It is foundational for many statistical methods, including confidence intervals and hypothesis tests. These procedures assume the normality of the sample mean, which the CLT justifies when the sample size is large.
- In industrial processes and quality control, the CLT can explain why many processes have a normal distribution of attributes, despite the underlying process not being normally distributed.

### 3. What is the statistical power?

Statistical power is a concept in hypothesis testing that describes the likelihood of correctly rejecting the null hypothesis when it is false.

### 4. How do you control for biases?

The strategies to control biases are as follows:

- Randomly assigning subjects to different groups helps ensure that each group is similar, which minimizes selection bias.
- Double-blind studies help reduce observer and measurement biases.
- Using standardized protocols for data collection can reduce measurement bias and ensure consistency.

- Ensuring that the sample adequately represents the population of interest can help prevent sampling biases.

### **5. What are confounding variables?**

A confounder is a third variable that influences both the independent variable and the dependent variable, leading to a spurious association. The identification and control of confounding variables are essential for the validity of a study's conclusions. Failure to control for confounders can lead to erroneous interpretations of the data and incorrect conclusions about the nature of the relationship between variables of interest.

### **6. What is A/B testing?**

A/B testing is a method in marketing and other disciplines to compare two versions of a webpage, email, app, or other marketing asset with just one varying element to determine which version performs better in terms of a predefined outcome.

### **7. What are confidence intervals?**

Confidence intervals are a range of values, derived from the statistics of observed data, that is likely to contain the true value of an unknown population parameter. They are used to indicate the reliability of an estimate. Confidence intervals provide more information than a simple point estimate, such as a sample mean. They give a range of plausible values for the population parameter and convey the precision of the estimate. A wide interval may indicate that more data should be collected before making inferences or decisions, while a narrow interval suggests more precision.