

## Week 5\_ Data Science Bootcamp - Null Hypothesis \_Solutions

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

Statistical significance is assessed by conducting hypothesis tests. Here's the general process:

- Start with a hypothesis (an educated guess) about an effect or relationship.
- Collect data through experiments or observations.
- Analyze the data using statistical tests, such as t-tests, chi-squared tests, or regression models.
- Calculate a p-value, which tells you the probability of observing the data if there were no real effect (the null hypothesis).
- If the p-value is smaller than a predetermined significance level (often 0.05), you reject the null hypothesis, indicating that the effect is statistically significant.

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

The Central Limit Theorem (CLT) is a fundamental concept in statistics. It states that when you collect a large enough sample of data from any population, the sample means will follow a normal distribution, even if the original population doesn't. This is important because:

- It allows us to make statistical inferences about a population based on sample data, even if we don't know the population's distribution.
- It simplifies hypothesis testing, as we can use the properties of the normal distribution to make accurate predictions.
- It's widely applicable across various fields, making it a crucial concept in statistics.

### 3. What is statistical power?

Statistical power is the probability of correctly rejecting a false null hypothesis in a hypothesis test. In other words, it measures the ability of a statistical test to detect an effect or relationship if it truly exists. A high statistical power means that the test is good at finding real effects and reducing the risk of false negatives (failing to detect an effect that is present).

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

Controlling for biases involves taking steps to minimize the impact of systematic errors in your data or analysis. Here are some methods:

- Randomization: Randomly assign participants or samples to groups or conditions to reduce selection bias.

- Blinding: Keep researchers or participants unaware of certain details to reduce confirmation bias.
- Use of Control Groups: Compare the group you're interested in with a control group to account for external factors.
- Standardized Procedures: Follow standardized data collection and analysis procedures to minimize bias.
- Peer Review: Have independent experts review your work to identify and address potential biases.

#### 5. What are confounding variables?

Confounding variables are factors that can influence both the dependent variable (the outcome you're interested in) and the independent variable (the factor you're manipulating or studying). They can create a misleading association between the two. To control for confounding variables, researchers use techniques like statistical controls, matching, or randomization.

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

A/B testing, also known as split testing, is a method used to compare two versions (A and B) of something, such as a web page, email, or product feature, to determine which one performs better. It involves randomly showing different versions to different users and collecting data on their responses. A/B testing is commonly used in marketing and web design to optimize user experiences and increase desired outcomes.

#### 7. What are confidence intervals?

A confidence interval is a range of values that is likely to contain a population parameter, such as a mean or proportion, with a certain level of confidence. For example, a 95% confidence interval for the mean height of a population might be 160 cm to 170 cm. This means that if you were to take many samples and calculate confidence intervals for each, about 95% of those intervals would contain the true population mean. Confidence intervals provide a measure of the precision and reliability of your estimates.