WEEK 5

1. How do you assess the statistical significance of an insight?
To assess statistical significance, follow these steps:
1. Formulate null and alternative hypotheses
2. Collect relevant data and calculate a test statistic
3. Choose a significance level (alpha)
4. Perform a statistical test and get a p-value
5. Compare p-value to alpha
6. Reject null hypothesis if p-value ≤ alpha

2. What is the Central Limit Theorem? Explain it. Why is it important?
The Central Limit Theorem (CLT) is a critical concept in statistics. It states that, under certain conditions, the sample mean (or other sample statistics) tends to be normally distributed, irrespective of the population distribution. Key takeaways:

- The population must be large enough.
- Samples must be random and independent.
- As the sample size increases, the distribution approaches normal.

The CLT enables statisticians to make inferences about a population based on sample data and use the normal distribution for hypothesis testing and confidence intervals. It plays a crucial role in many statistical methods.

3. What is the statistical power?
Statistical power refers to the likelihood of correctly rejecting a null hypothesis when it is false. It indicates the ability of a statistical test to detect a difference or effect if it truly exists. A test with high statistical power is more likely to identify real effects, while one with low power is more likely to miss them. To increase statistical power, you can take any of the following measures:
- Increase the sample size.
- Use more sensitive measurements.
- Decrease the significance level (alpha).

4. How do you control for biases?
To avoid errors in data collection and analysis, several strategies can be implemented to minimize biases:

- Randomization: Assign subjects randomly to ensure a representative sample
- Blinding: Keep participants unaware of certain aspects to prevent bias
- Matching: Group subjects based on relevant characteristics to balance confounding factors
- Standardized procedures: Consistent data collection and analysis reduce bias
- Sensitivity analysis: Examine biases' impact on results to identify and address issues.

5. What are confounding variables?
Confounding variables are elements that can influence the relationship between the independent variable (the one you're interested in) and the dependent variable (the outcome you're measuring). These variables can lead to bias and incorrect conclusions. To control for confounding variables, you can employ the following methods:

- Randomize subjects or treatments.
- Match groups based on relevant characteristics.
- Incorporate confounding variables as covariates in statistical models.
- Conduct subgroup analysis to explore the impact of confounding factors.

6. What is A/B testing?
A/B testing, which is also known as split testing, is a method used to compare two versions of a web page, app, or any digital content. Its purpose is to determine which version performs better in terms of a predefined metric, such as conversion rate or click-through rate. The process involves randomly assigning users to one of two or more groups (A, B, etc.) and exposing each group to a different version of the content. By analyzing the results, you can determine which version is more effective and make data-driven decisions for optimization.

7. What are confidence intervals?
Confidence intervals are a range of values that are used to estimate a population parameter, such as a mean or proportion, based on a sample of data. Confidence intervals indicate the level of uncertainty in the estimation, and they are useful for expressing the precision of an estimate. A common confidence level is 95%, which means that if you were to collect multiple samples and calculate confidence intervals, you would expect approximately 95% of those intervals to contain the true population parameter. These intervals are often used in inferential statistics to provide a margin of error for the parameter being estimated, and they help to indicate the degree of confidence we can have in our results.