Week 7- Homework

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

We check whether in the data for real or just random noise. We set up a null hypothesis — like “there’s no effect” — and then use a test (like a t-test) to get a p-value. If that p-value is small enough (usually less than 0.05), we say, this result is statistically significant meaning it’s unlikely to have happened by chance.

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

The Central Limit Theorem (CLT) says that if you take a lot of samples from any population, no matter how the original data looks, the averages of those samples will start to look like a normal (bell curve) distribution — as long as the sample size is big enough. This is super useful because it lets us use normal distribution tools (like confidence intervals and hypothesis tests) even when the data isn’t normally distributed.

3. What is statistical power?

Power is like the test’s ability to catch a real effect. If there is actually something going on (like a difference between two groups), power tells us how likely we are to detect it. A high power (like 80% or more) means we have a good shot at not missing the truth. If power is low, even real effects can go unnoticed.

4. How do you control for biases?

We control for biases by using randomization, blinding, and control groups in experimental design. In observational studies, techniques such as matching, stratification, and regression adjustment are used to account for confounding variables and reduce bias.

5. What are confounding variables?

Confounding variables are extraneous variables that correlate (positively or negatively) with both the dependent and independent variable. They can distort the perceived relationship between the variables being studied, leading to incorrect conclusions.

6. What is A/B testing?

A/B testing is a method of comparing two versions of a variable to determine which one performs better. It involves randomly assigning users to either group A (control) or group B (variant) and measuring differences in outcome to infer which version is more effective, often using hypothesis testing. It's like a mini experiment — you show one group version A (like the current website layout) and another group version B (a new one), then see which one performs better. It’s common in marketing and product design. It’s all about making data-driven decisions instead of guessing.

7. What are confidence intervals?

A confidence interval gives a range of values that is likely to contain a population parameter with a specified level of confidence (e.g., 95%). It reflects the uncertainty or variability in the sample estimate and is used in inferential statistics to assess the precision of estimates.