Take Home Assignment Week 7 - Martha McQuillan (mbm7846)

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

You would use a hypothesis test to assess statistical significance. Find the null and alternative hypotheses. Then you calculate a test statistic, may it be z or t or another one depending on the context of the problem. Finally determine the p value, which is the probability of observing the result or something more extreme under the null hypothesis. If the p value is less than or equal to an alpha (number pertaining to significance level), reject the null hypothesis.

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

The Central Limit Theorem states that the sampling distribution of a variable becomes closer and closer to the normal distribution as you increase the sample size. It's important because it means that you can use the confidence intervals and hypothesis tests with z-statistics to understand data that isn't necessarily normally distributed.

3. What is the statistical power?

The statistical power is the probability of correctly rejecting the null hypothesis when it is actually false. Mathematically, power = 1 - beta, where beta is the probability of failing to reject the null when it is actually false.

4. How do you control for biases?

Some precautions you can take to control for biases include random sampling, making sure researchers aren't biased when collecting data, and correctly understanding and depicting confounding variables.

5. What are confounding variables?

Confounding variables are variables that can impact both the independent and dependent variables, leading to a false or uninformed conclusion.

6. What is A/B testing?

A/B testing compares 2 versions of a variable to figure out which one performs better.

7. What are confidence intervals?

A confidence interval is a range of values derived from a data set that is likely to contain the true parameter given a confidence level. The confidence level tells you what percent of times your interval would contain the true population parameter if the experiment was repeated many times.