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

To assess the statistical significance of an insight, we need to start by establishing a null hypothesis that suggests no effect and an alternative hypothesis that indicates a real effect. Choose a significance level (ie alpha = 0.05), and select an appropriate statistical test based on the data type and task at hand (ie t-test or chi-square). Collect and analyze the data and calculate the p-value, which indicates the probability of observing the results if the null hypothesis were true. A p-value less than or equal to the significance level suggests that the null hypothesis can be rejected in favor of the alternative.

It's important to also evaluate the assumptions underlying the statistical test, the size of the effect, and the confidence intervals to contextualize the statistical significance.

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

The Central Limit Theorem indicates that as you take larger and larger samples from any population, the sample means tend to form a normal distribution around the population mean, even if the population itself is not normally distributed. This principle is important because it justifies using the normal distribution for inferential statistics, like hypothesis testing and confidence intervals, allowing for predictions about population parameters based on sample data, even with non-normal populations.

3. What is statistical power?

Statistical power is the probability that a statistical test will correctly reject a false null hypothesis, which essentially measures a test's capability to identify an actual effect. Power is influenced by the effect size, sample size, significance level, and data variability.

4. How do you control for biases?

To control for biases, we need to employ methods such as randomization to balance groups, blinding to prevent subject and experimenter expectations from influencing results, and use control groups to separate treatment effects from other variables.

5. What are confounding variables?

Confounding variables are extraneous factors that correlate, either positively or negatively, with both the dependent variable and the independent variable in an experimental or non-experimental study. These variables can cause a false association, leading to incorrect conclusions about the relationship between the variables of interest because their effects are not separated from the primary variables being studied.

6. What is A/B testing?

A/B testing is a methodological approach used to compare two versions of a web page or app against each other to determine which one performs better in terms of specific metrics such as conversion rates, click-through rates, or user engagement. In an A/B test, users are randomly assigned to either a control group or a variation group , and their interactions with their respective version are statistically analyzed to ascertain which version is more effective at achieving the predetermined objectives.

7. What are confidence intervals?

Confidence intervals are a range of values, derived from sample statistics, that are believed to contain the true population parameter (ie the mean or proportion) with a certain level of confidence, typically expressed as a percentage (ie 90%, 95%, or 99%). Instead of providing a single point estimate, a confidence interval gives an estimated range, which is calculated from the data, that is likely to include the unknown parameter, assuming that the sampling process is repeated multiple times. The width of this range conveys the precision of the estimate, with the specified confidence level reflecting the degree of certainty in the interval containing the parameter.