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

You assess statistical significance by conducting a **hypothesis test**, calculating a **p-value**, and comparing it to a **significance level** (typically $\alpha = 0.05$). If the p-value is less than α, the result is statistically significant, meaning it's unlikely to have occurred by random chance.

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

The **Central Limit Theorem** states that the **sampling distribution of the sample mean** approaches a **normal distribution** as the sample size increases, **regardless of the population's original distribution**, provided the sample size is large enough ($n > 30$ is a common rule of thumb).

Importance: It allows us to use **normal distribution-based inference** (like z-tests and confidence intervals) even when the population distribution is unknown.

3. What is statistical power?

Statistical power is the **probability of correctly rejecting the null hypothesis** when it is false (i.e., detecting an effect when there is one). High power (typically ≥ 0.8) reduces the risk of **Type II errors** (false negatives).

4. How do you control for biases?

- **Randomization**: Distributes potential confounding variables evenly.
- **Blinding**: Prevents participant or experimenter bias.
- **Stratification**: Ensures subgroups are evenly represented.
- **Matched sampling**: Matches individuals across groups to control variables.
- **Statistical control**: Using regression or ANCOVA to account for known confounders.

5. What are confounding variables?

Confounders are variables that are **correlated with both the independent and dependent variable**, potentially **distorting the true relationship** between them. They create a **false association** if not controlled for.

6. What is A/B testing?

A/B testing is an experiment where you **compare two versions** (A and B) of a variable (e.g., website design, email subject) to determine **which performs better** based on a predefined metric. It relies on **random assignment** and hypothesis testing.

7. What are confidence intervals?

A **confidence interval** gives a **range of values** that likely contains the population parameter with a certain **confidence level** (e.g., 95%). It reflects both the **estimate** and the **uncertainty** around it.