1. How do you control for biases?

Controlling for biases in research and data analysis involves using strategies to minimize the impact of these biases on study results. Some common methods include:

- Randomization: Assigning subjects randomly to different groups or treatments helps ensure that each group is similar on average, reducing the impact of confounding variables.
- Blinding: Keeping study participants, researchers, or both unaware of the allocation to treatment or control groups to prevent their expectations from affecting the results.
- Matching: Pairing subjects in treatment and control groups based on similar characteristics (e.g., age, gender) to control for these variables.
- Statistical Control: Using statistical techniques like regression to adjust for the influence of variables that might bias the results.
- Replication: Repeating studies to see if the results are consistent across different samples and settings can help to identify and correct for biases.

2. What are confounding variables?

Confounding variables are factors other than the independent variable that might affect the dependent variable in a study. They can create a false impression of a relationship between studied variables by either hiding or exaggerating an association. Properly identifying and adjusting for confounding variables is crucial for establishing accurate and reliable research findings.

3. What is A/B testing?

A/B testing is a method in statistics used to compare two versions of a variable to determine which performs better in a controlled environment. For example, it is often used in web development to compare two webpage designs by showing them to different visitors at the same time and measuring the effectiveness of each design on user behavior.

4. When will you use Welch t-test?

The Welch t-test is used when comparing the means of two groups that may have different variances and potentially different sample sizes. It is an adaptation of the Student's t-test and is more reliable when the assumption of equal variances (homoscedasticity) in the two groups is not met.

5. A company claims that the average time its customer service representatives spend on the phone per call is 6 minutes. You believe that the average time is actually higher. You collect a random sample of 50 calls and find that the average time spent on the phone per call in your sample is 6.5 minutes, with a standard deviation of 1.2 minutes. Test whether there is sufficient evidence to support your claim at a significance level of 0.05.

- Null Hypothesis (H0): The average time μ is 6 minutes.
- Alternative Hypothesis (H1): The average time μ is more than 6 minutes.
- Significance Level: $\alpha=0.05$

Calculations:

- Mean of sample = 6.5 minutes
- Standard deviation (s) = 1.2 minutes
- Sample size (n) = 50
- Standard Error (SE) = s/sqrt (n)
- Test Statistic (t) = $(\text{mean}-\mu 0)/SE$
- Test Statistic (t): 2.95
- p-value: 0.0025

Since the p-value (0.0025) is less than the significance level (0.05), we reject the null hypothesis. There is sufficient evidence to support the claim that the average time spent on the phone per call is more than 6 minutes.

6. A researcher wants to determine whether there is a difference in the mean scores of two groups of students on a math test. Group A consists of 25 students who received traditional teaching methods, while Group B consists of 30 students who received a new teaching method. The average score for Group A is 75, with a standard deviation of 8, and the average score for Group B is 78, with a

standard deviation of 7. Test whether there is a significant difference in the mean scores of the two groups at a significance level

- Null Hypothesis (H0): There is no difference in mean scores between the groups.
- Alternative Hypothesis (H1): There is a difference in mean scores.
- Significance Level: $\alpha = 0.05$

Calculations:

- Mean of Group A = 75
- Standard deviation of Group A= 8
- Mean of Group B = 78
- Standard deviation of Group B= 7
- Sizes: Na=25, Nb=30
- Test Statistic (t) = $(\text{mean} - \mu 0)/SE$
- Test Statistic (t): -1.18
- p-value: 0.244

The p-value (0.244) is greater than the significance level (0.05), indicating that we fail to reject the null hypothesis. Therefore, there is not enough evidence to conclude that there is a significant difference in the mean scores of the two groups on the math test based on the teaching methods used.