**1. How do you control for biases?**

Randomization: Randomly assigning subjects to treatment groups or sampling data to ensure representativeness.

Stratification: Subdividing the population or dataset into homogeneous groups based on relevant characteristics to control for potential biases.

Blinding: Concealing information about the treatment or outcome from participants or analysts to reduce bias due to expectations or preconceptions.

Sensitivity analysis: Assessing the robustness of results to different assumptions, parameter values, or analytical approaches to identify and mitigate biases.

Validation: Cross-validating findings using independent datasets or alternative methodologies to confirm the reliability and generalizability of results.

**2. What are confounding variables?**

Confounding variables are extraneous factors that are associated with both the independent and dependent variables in a study, making it difficult to determine the true relationship between them. Confounding variables can lead to spurious correlations or biased estimates if not properly controlled for.

A confounding variable is a variable that (a) alters the risk of the disease or condition under study independent of the exposure or characteristic of primary interest, and (b) is associated with the exposure or characteristic of primary interest in the study population, but (c) is not a consequence of that exposure

**3. What is A/B testing?**

A/B testing, also known as split testing, is a controlled experiment used to compare two or more versions of a webpage, app feature, marketing campaign, or other elements to determine which one performs better in terms of a predefined outcome (e.g., conversion rate, click-through rate, engagement). In an A/B test, participants are randomly assigned to different variants (A, B, etc.), and their responses or behavior are measured and compared to assess the impact of changes or interventions. A/B testing helps businesses and organizations make data-driven decisions and optimize their strategies to improve performance and user experience.

**4. When will you use Welch t-test?**

The Welch t-test, also known as the unequal variances t-test, is used to compare the means of two independent samples when the assumption of equal variances is violated. You would use the Welch t-test in situations where the variances of the two groups being compared are significantly different, which can occur when the sample sizes are unequal or when the populations being sampled have different variances. The Welch t-test provides a more robust and reliable comparison of means compared to the standard Student's t-test when the assumption of equal variances cannot be 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.

Given:

Population mean (μ): 6 minutes (claimed by the company)

Sample size (n): 50 calls

Sample mean (x̄): 6.5 minutes

Sample standard deviation (s): 1.2 minutes

Significance level (α): 0.05 (typically)

The null and alternative hypotheses are:

Null hypothesis (H0): μ = 6 (The average time spent on the phone per call is 6 minutes.)

Alternative hypothesis (H1): μ > 6 (The average time spent on the phone per call is greater than 6 minutes.)

We will conduct a one-tailed t-test and compare the test statistic to the critical t-value at the specified significance level.

t\_statistic = (sample\_mean - population\_mean) / (sample\_std / (n \*\* 0.5)) = 2.946

critical\_t\_value = 1.677

t\_statistic > critical\_t\_value

Reject the null hypothesis. There is sufficient evidence to support your claim.

**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 of 0.05.

To test whether there is a significant difference in the mean scores of two groups of students (Group A and Group B) on a math test, we can use an independent samples t-test. Given the following information:

Group A:

Sample size (n1) = 25

Sample mean (x̄1) = 75

Sample standard deviation (s1) = 8

Group B:

Sample size (n2) = 30

Sample mean (x̄2) = 78

Sample standard deviation (s2) = 7

We will conduct a two-tailed t-test to determine if there is a significant difference between the means of the two groups. The null and alternative hypotheses are:

Null hypothesis (H0): μ1 = μ2 (There is no difference in the mean scores of the two groups.)

Alternative hypothesis (H1): μ1 ≠ μ2 (There is a difference in the mean scores of the two groups.)

sp = ((n1 - 1) \* s1\*\*2 + (n2 - 1) \* s2\*\*2) / (n1 + n2 - 2)

t\_statistic = (x1 - x2) / ((sp \* ((1/n1) + (1/n2)))\*\*0.5) = -1.483

df = n1 + n2 – 2 = 25 + 30 – 2 = 53

critical\_t\_value = 2.006

|t\_statistic| < critical\_t\_value

Fail to reject the null hypothesis. There is no significant difference in the mean scores of the two groups.

**Note**: Calculations can be shown by hand.