Practice Questions Answered

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

Statistical significance is assessed through hypothesis testing using a p-value, which measures the probability of observing the data, or something more extreme, under the null hypothesis. If the p-value is below a predetermined significance level (commonly 0.05), then the result is considered statistically significant, and the null hypothesis is rejected. This approach helps determine whether an observed effect is likely due to chance or some specific cause.

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

Central Limit Theorem (CLT) states that the distribution of the sample means approaches a normal distribution as the sample size gets larger, regardless of the shape of the population distribution. This is crucial because it allows for making inferences about population parameters using sample statistics even when the population distribution is unknown, which is common in practical scenarios.

Importance:

- **Foundation for Inferential Statistics**: CLT allows us to use sample means to make conclusions about population means (or other parameters), which are typically not known.
- **Justifies the use of Confidence Intervals and Hypothesis Tests**: Most statistical tests assume normality of the data, and CLT provides a theoretical foundation for this assumption when dealing with means of samples.

3. What is the statistical power?

Statistical power is the probability that a test correctly rejects the null hypothesis when it is false (i.e., detects an effect when there is one). It is influenced by several factors, including the significance level, sample size, and effect size. A higher power reduces the risk of a Type II error (failing to detect a true effect).

4. How do you control for biases?

To control for biases in statistical analysis:

- **Randomization**: Randomly assigning subjects to different groups helps control for both known and unknown confounders.
- **Blinding**: Keeping study participants and researchers blind to group assignments prevents bias in treatment administration and outcome assessment.

- **Control Groups**: Using control groups allows comparisons that can isolate the effect of the variable of interest.
- **Statistical Controls**: Including covariates in statistical models can help adjust for potential confounders.

5. What are confounding variables?

Confounding variables are extraneous variables that correlate (positively or negatively) with both the dependent variable and the independent variable. These variables can lead to a spurious association between the studied variables, thereby obscuring the true effect of the independent variable on the dependent variable.

6. What is A/B testing?

A/B testing is a basic randomized control experiment comparing two versions (A and B) of a single variable to determine which version more effectively achieves a given outcome. Commonly used in website optimization, marketing strategies, and other fields requiring user engagement analysis.

7. What are confidence intervals?

A confidence interval (CI) is a range of values, derived from the sample data, that is likely to contain the true population parameter with a certain level of confidence (commonly 95%). It provides an estimate of the uncertainty around the sample estimate, offering an interval within which the true parameter value is expected to fall.