1. Assessing Statistical Significance: The statistical significance of an insight is typically determined through hypothesis testing. This involves calculating a p-value, which indicates the probability of observing the data if the null hypothesis (usually stating no effect) were true. A significance level, commonly 0.05, is chosen, and if the p-value is less than this threshold, the result is considered statistically significant.
2. Central Limit Theorem (CLT): The CLT states that the sampling distribution of the sample means approaches a normal distribution as the sample size increases, regardless of the shape of the population distribution. It's important because it allows us to make inferences about population parameters using sample statistics, forming the basis for many statistical methods like hypothesis testing and confidence intervals.
3. Statistical Power: Statistical power is the probability of correctly rejecting a false null hypothesis, i.e., the probability of detecting a true effect. It depends on factors like sample size, effect size, and significance level. High statistical power is desirable as it reduces the likelihood of Type II errors (failing to detect a true effect).
4. Controlling for Biases: Biases can be controlled through various techniques such as randomization, blinding, and careful experimental design. Biases can be mitigated by identifying and minimizing sources of systematic error in data collection, analysis, and interpretation.
5. Confounding Variables: Confounding variables correlate with the independent and dependent variables in a study, leading to incorrect conclusions about the relationship between them. Controlling for confounding variables through study design or statistical methods helps isolate the true effect of interest.
6. A/B Testing: A/B testing (or split testing) is a method used to compare two versions of something (like a webpage, email, or advertisement) to determine which one performs better. Participants are randomly assigned to either the control group (A) or the experimental group (B), and their responses are compared to assess the impact of changes made between the two versions.
7. Confidence Intervals: Confidence intervals provide a range of values within which the true population parameter is likely to fall, along with a degree of confidence (typically 95%). They help quantify the uncertainty associated with estimating population parameters from sample data. A wider confidence interval indicates greater uncertainty, while a narrower one suggests more precision in estimation.