

# THE BUILDING BLOCKS

## Population



In statistics, a population is a set of similar items or events which is of interest for some question or experiment. It's the group we want information about.

When AB testing a webpage or app, the true population is every future individual who will visit that page/app

## Sample



A data sample is a set of data collected and/or selected from a statistical population by a defined procedure. It's a small portion of the larger population.

In product AB testing, the sample is the number of visitors we display our new page variation to in order to collect data and draw inference about the overall population.

## Mean



The mean is the central tendency of a probability distribution.

In product AB testing, the mean is our page's conversion rate with the sample visitors.

## Sampling Variability



This is a measure of error in our population estimate due to differences in samples. Sampling variability will decrease as the sample size increases.

In product AB testing, the sampling variability affects the sample size we need in order to have a chance of deriving statistically significant results.

# STATISTICS FOR A/B TESTING

## Null Hypothesis



Inferential statistics is based on the premise that you cannot prove something to be true but you can disprove something by finding an exception. You decide what you are trying to provide evidence for - which is the alternate hypothesis, then you set up the opposite as the null hypothesis and find evidence to disprove that

In product AB testing, the null hypothesis is that the population conversion rate on the original page and the new page are not different.

## Confidence Level



Confidence level refers to the percentage or probability, or certainty, that the confidence interval would contain the true population parameter when you draw a random sample many times.

In product AB Testing, a 95% confidence level is typically chosen. A 95% confidence level means that the confidence interval around sample mean is expected to include the true mean value 95% of the time.

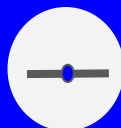
## Margin of error



A margin of error tells you how many percentage points your results will differ from the real population value. For example, a 95% confidence interval with a 4 percent margin of error means that your statistic will be within 4 percentage points of the real population value 95% of the time.

The margin of error is added to and subtracted from the mean to determine the confidence interval.

## Confidence Interval



In statistical inference, we aim to estimate population parameters using observed sample data. A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.

The width of confidence interval depends on 3 things. The variation within the population of interest, the size of the sample and the level of confidence we are seeking.

## Type I Error



A type I error occurs when we incorrectly reject the null hypothesis.

In product AB testing, a type I error would occur if we concluded that population mean of Variation B is different than population mean of Variation A when it reality they were the same, Type I error is avoided by achieving statistically significant results.

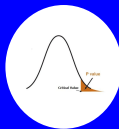
## Type II Error



A type II error occurs when the null hypothesis is false, but we incorrectly fail to reject it.

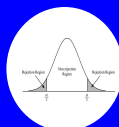
To put this in product AB testing terms, a type II error would occur if we concluded that population mean of Variation B is not different than mean of Variation A when it actually was different. These errors are avoided by running tests with a high statistical power.

## p-value



p-value is the probability of obtaining at least as extreme results as we are seeing, given that the null hypothesis is true. p-value basically tells you whether your evidence makes your null hypothesis look ridiculous.

## Statistical Significance



Statistical significance is attained when the p-value is less than the significance level. The significance level ( $\alpha$ ), is the probability of rejecting the null hypothesis given that it is true.

In AB testing, statistical significance is how we verify that a new page outperforms the original

## Statistical Power



Statistical Power, which as we know is the probability that a test correctly rejects the null hypothesis i.e. the percentage of time the minimal effect will be detected, if it exists