Inferential Analysis

4 min

A/B tests are a popular business tool that data scientists use to optimize websites and other online platforms. A/B tests are a type of **inferential analysis**. Inferential analysis lets us test a hypothesis on a sample of a population and then extend our conclusions to the whole population.

Imagine we want to know if a blue or a green button will get more clicks on a website. We hypothesize that the green button will be more successful and we run an A/B test on a sample of people that visit our site. Half the sample sees the blue button (option A) and half see the green button (option B). At the end of the test, 90% of people that saw the green button clicked it, whereas 60% of the people that saw the blue button clicked it.

Now we need to ask, "If color wasn't related to click rate, how likely was a 30% difference just by chance?" We can use

Preview: Docs Statistics is the science that is concerned with methods for collecting, organizing, analyzing, and interpreting data.

[statistics](#)

 to calculate that probability. If there is less than a 5% probability that our results happened by chance, we have good evidence that green buttons get more clicks! We can extend these results to everyone that visits our site (the whole population), so it makes sense to use green buttons on our website.

This is a powerful thing to be able to do! But since it's so powerful, there are some rules about how to do it:

- Sample size must be big enough compared to the total population size (10% is a good rule-of-thumb).

- Our sample must be randomly selected and representative of the total population.

- We can only test one hypothesis at a time.

Instructions

Based on what we know about A/B tests, consider the following scenarios:

1. If the A/B test was being run for a site with 1 million visitors per day, would a sample size of 100 be big enough to extend conclusions from the sample to the whole population?

2. The site in the A/B test has visitors from all over the planet. If you ran the A/B test for 12 hours when only one-half of the planet was awake, would your sample be representative of the whole population?

3. When assigning people to see either the blue or green button, would it be better to randomly assign people or only assign people to the green group if they say their favorite color is green?

4. If instead of testing a green button versus a blue button, the test was of a blue button with small font vs. a green button with large font, would we be able to conclude that the color of the button was influencing whether visitors clicked the button?

Next up, causal analysis!