

A/B TESTING

1. A/B Testing

00:00 - 00:02

Welcome back!

2. Data science workflow

00:02 - 00:06

This chapter covers the final stage, beginning with experiments.

3. What are experiments in data science?

00:06 - 00:19

Experiments help drive decisions and draw conclusions. Generally, they begin with a question and a hypothesis, then data collection, followed by a statistical test and its interpretation.

4. Case study: which is the better blog post title?

00:19 - 00:47

For instance, suppose we want to pick the best blog post title. Our question is: does blog title A or blog title B result in more clicks? We make the hypothesis that both blogs will result in the same amount of clicks. To collect data, we randomly divide our audience into two groups, each seeing a different title. We run the experiment until our sample size is reached - more on that later.

5. Case study: which is the better blog post title?

00:47 - 01:08

Then, we use a statistical test to see whether the difference between the titles' click-through rate is significant. Finally, we interpret results. In our case, we want to choose the better performing title. However, often, it's inconclusive, leading to more questions and experiments.

6. What is A/B Testing?

01:08 - 01:17

This is called A/B Testing, often called Champion/Challenger testing, and it's used to make a choice between two options.

7. Terminology Review

01:17 - 01:43

Sample size is the number of data points used. In data science, you'll hear "Are your results significant?". A statistically significant result means that your result is probably not due to chance given the statistical assumptions made. Statistical tests help determine this and there are many to choose from. Let's take a closer look at A/B testing.

8. A/B Testing Steps

01:43 - 01:56

There are four steps in an A/B test: picking a metric to track, calculating sample size, running the experiment, and checking for significance. Let's examine each step.

9. Pick a metric to track: click-through rate

01:56 - 02:03

Our metric is click-through rate, the percent of people who click on a link after viewing the title.

10. Calculate your sample size

02:03 - 02:32

Next, we'll run the experiment until we reach a sample size large enough to be certain that results aren't due to random chance. The size depends on a "baseline metric" that helps gauge any changes. In our case, it's how often people click on a link to one of our blogs usually. If the rate is much larger or smaller than 50%, then we need a larger sample size. Click-rate is typically under 3%.

11. Calculate your sample size

02:32 - 03:12

The sample size depends on the sensitivity we want. A test's sensitivity tells us how small of a change in our metric we can detect. Larger sample sizes allow us to detect smaller changes. You might think that we want a high sensitivity, but we actually want to optimize for an amount that is meaningful for our question. For example, if the first and second title are clicked on by 5% and 5.01% of viewers respectively, an extra .01% clicks to our blog post won't make a difference to us. However, if this was comparing 100m race time, an .01 difference would be significant.

12. Run your experiment

03:12 - 03:20

We run our experiment until we reach the calculated size. Stopping the experiment before or after could lead to biased results.

13. Check for significance

03:20 - 03:41

Once we reach the target size, we check our metric. We see some difference between the titles, but how do we know if it's meaningful? We check by performing a test of statistical significance. If they are significant, we can be reasonably sure that the difference is not due to random chance, but to an actual difference in preference.

14. What if the results aren't significant?

03:41 - 04:06

What if the results aren't significant? If there are any differences in click rates, they're smaller than the threshold we chose when determining the sensitivity. Running our test longer won't help because it'll only detect a smaller difference and we've decided that smaller differences are irrelevant! There still might be a difference in click rates between the titles: but that difference isn't important to us for making decisions.

15. Let's practice!

04:06 - 04:10

Now, let's practice!