False discovery rate control

Skip Ahead

False discovery rates in Optimizely

THIS ARTICLE WILL HELP YOU:

• Distinguish between false discovery rate control in Optimizely
• Make business decisions based on the results you see

Every experiment has a chance of reporting a false positive—in other words, reporting a conclusive result between two variations when there’s actually no underlying difference in behavior between them. You can calculate the rate of error for a given experiment as 100 - [statistical significance]. This means that higher statistical significance numbers decrease the rate of false positives.

Using traditional statistics, you increase your exposure to false positives as you run experiments on many goals and variations at once (the "multiple comparisons problem"). This happens because traditional statistics controls the false positive rate among all goals and variations. However, this rate of error does not match the chance of making an incorrect business decision or implementing a false positive among conclusive results. Here’s how this risk increases as you add goals and variations:
In this illustration, there are nine truly inconclusive results, and one of those registers as a false winner. This results in an overall false positive rate of about 10%. However, the business decision you'll make is to implement the winning variations, not the inconclusive ones. The rate of error of implementing a false positive from the winning variations is one out of two, or 50%. This is called the proportion of false discoveries.

**Optimizely controls errors, and the risk of incorrect business decisions, by controlling the false discovery rate instead of the false positive rate.** Here's how Optimizely defines error rate:

\[
\text{False Discovery Rate} = \frac{\text{(average number of incorrect winning and losing declarations)}}{\text{(total number of winning and losing declarations)}}
\]

Read more about the distinction between false positive rate and false discovery rate in our blog post.

**Important:**

We do not recommend adding a goal or variation after you've started an experiment. Although it's unlikely to have an effect at first, there's a greater chance that adding a new goal or variation will affect your existing results as you see more and more traffic.

**Note:**

Optimizely makes sure that the goal you choose as your primary goal always has the highest statistical power by treating it differently in our false discovery rate control calculations. Our false discovery rate control protects the integrity of all your goals from the "multiple comparisons problem" of adding several goals and variations to your experiment, without keeping your primary goal from reaching significance in a timely fashion.

Learn more about setting events.

Learn how to optimize your events and goals for achieving significance quickly with Stats Engine [here](#).
False discovery rates in Optimizely

We updated false discovery rates in Optimizely to better match customers’ diverse approaches to running experiments. We explained above how your chance of making an incorrect business decision increases as you add more metrics and variations (the "multiple comparisons problem"). This is true, but it’s not the whole story.

Consider an experiment with seven events: one headline metric that determines success of your experiment; four secondary metrics tracking supplemental information; and two diagnostic metrics used for debugging. These metrics aren’t all equally important. Also, statistical significance isn’t as meaningful for some (the diagnostic metrics) as it is for others (the headline metric).

Optimizely solves this problem by allowing you to rank your metrics. The first ranked metric is still your primary metric. Metrics ranked 2 through 5 are considered secondary. Secondary metrics take longer to reach significance as you add more of them, but they don't impact the primary metric's speed to significance. Finally, any metrics ranked beyond the first five are monitoring metrics. Monitoring metrics take longer to reach significance if there are more of them, but have minimal impact on secondary metrics and no impact on the primary metric.

The result is that your chance of making a mistake on your primary metric is controlled. The false discovery rate of all other metrics is also controlled, all while prioritizing reaching statistical significance quickly on the metrics that matter most.