Interpret your results

Skip Ahead

Segment your results
Check secondary and monitoring goals
Seasonality and traffic spikes
Broad difference interval

This article will help you:

• Interpret the results of winning, losing, and inconclusive experiments
• Segment by audience for a granular view of customer behaviors
• Evaluate how the results of this test affect other valuable revenue streams
• Gather sufficient data before deciding to stop an experiment

Once you publish an experiment or campaign, you can start checking your Results page. Unlike many testing tools, Optimizely's Stats Engine uses a statistical approach that allows you to peek into results without introducing error.

The Results page is where you'll find value in Optimizely. To run a truly data-driven experimentation program, it's important to take time to review and interpret the data that you collect before deciding to take action.

Your experiment results -- whether winning, losing, or inconclusive -- are an incredibly valuable resource. The data on your Results page helps you learn about your visitors, make data-driven business decisions, and feed the iterative cycle of your experimentation program. Before stopping an experiment, really dig into your data to look for valuable insights beyond which variations won or lost.

This article provides some high-level tactics for investigating the results of your experiment.

A few quick tips:

• Use losing and inconclusive tests to learn more about what visitors expect and how you can provide it
• Use winning variations to learn what changes generated desired outcomes -- and why

• Compare results to qualitative research and your hypothesis to bring your experiment full circle

• Before stopping an experiment, check that you’ve gathered enough data for your business needs

• Document and share takeaways -- they’re valuable resources

Once you’re done analyzing results, decide how to take action on winners, losers, and inconclusive results. See requirements

Materials to prepare

• ◦ Test results
  ◦ Analytics data from other platforms
  ◦ Hypothesis or experiment plan
  ◦ Qualitative data (surveys, customer reports)

People and resources

• ◦ Program manager
  ◦ Analyst

Actions you’ll perform

• ◦ Segment results to look for patterns
  ◦ Check secondary and monitoring goals
  ◦ Consider seasonality or traffic spikes
  ◦ Check the difference interval
  ◦ Use the root cause analysis to evaluate why the test affected visitors’ behaviors

Deliverables

• ◦ Key takeaways from the test (as inputs for your business intelligence report and general insights)

What to watch out for

• ◦ An under-developed hypothesis makes it difficult to interpret results
  ◦ Bias towards certain outcomes can stand in the way of understanding the data
  ◦ Don’t forget to document takeaways and communicate what you’ve learned

• This article is a part of the Optimization Methodology series.

Are you testing a checkout or other type of funnel? Here are a few tips for interpreting patterns in the results of those experiments.

If you’re using Optimizely to test on a checkout page, you might need to configure your site for PCI compliance.
Segment your results

Think of overall results of an experiment as an average across all visitors. Not all visitors behave like your average visitor. Segmenting your results (filtering results for specific audiences or attributes) - is a powerful way to generate insights about your customers.

Different types of visitors have different goals on your site. You may find that a change that doesn’t move the needle for most visitors is a huge hit with a certain subset. Conversely, an experience that lifts conversions across the board might also be very bad for a particular group.

Below, variation_1 is a clear winner for the “add_to_cart” and “browse_products” metrics.

But what if you segment for Mobile Visitors only, and you see that variation_1 is a clear loser for the same two metrics? Moreover, suppose variation_1 is also a statistically significant loss for Mobile Visitors, even though it isn’t statistically significance for visitors overall yet (above).

At this point, you should investigate why variation_1 is a bad experience for Mobile visitors and consider excluding them from the experiment going forward.

ANALYZE

Dig into default segments such as browser type or device type and custom segments that are important to your business.

Here’s what to look for:

• Do any segments of visitors behave differently from visitors overall?

• What do you know about those visitors? Why do you think they respond differently?

• What do your most valuable visitors prefer?

Imagine that you’re testing a streamlined login process on your site. You test a Facebook login and see significant lift across all visitors. But when you segment by browser type, the conversion rate for visitors using Internet Explorer convert is a statistically significant loss. Why?

Assuming that nothing is broken, start by considering what you already know. Maybe Internet Explorer visitors are likely to be older or to come from a professional services environment, compared to Safari visitors (sometimes linked to higher-income or tech-savvy users). Are professional visitors less likely to log in with a personal account? Do older visitors hesitate before connecting through Facebook?
Use these insights to make business decisions about your results. Will you roll out the Facebook login as an option instead of a requirement? Will you personalize it for just the high-converting segments?

**LEARN**

Combine insights from segmenting results with other data, like results from previous experiments, direct data, and indirect data.

In the example above, why did Mobile Visitors respond differently from other visitors. Is the Text CTA difficult to click in mobile? Is the Pop-Up CTA frustrating on a smaller screen?

In your next round of experiments, these insights serve as inputs for your direct data.

Share what you've with your organization. Data-driven insights may benefit other teams, and you'll help increase the impact of your program!

---

**Check secondary and monitoring goals**

Optimizely allows you to set a primary metric to measure success. Stats Engine weighs that primary metric differently, so it always reaches significant as soon as possible. Secondary and monitoring goals are all the goals in the experiment that aren’t the primary goal.

As a best practice, we recommend that you set secondary to track conversions down the funnel. Monitoring goals help you answer: where am I optimizing this experience, and where (if anywhere) am I worsening it?

Here are a few questions to help you evaluate secondary goals:

- Where in your funnel do you see improvement or loss? Does a pattern emerge?
- Is the exit rate higher on any step in the funnel, compared to the original?
- How does a significant lift or loss at a certain step correspond to changes you’ve made?

To learn more about interpreting your secondary goals, read this article on the five most common patterns in results.

Here are a few questions to help you evaluate monitoring goals:

- How does my test affect this monitoring goal?
- Are there multiple monitoring goals? What story do these goals tell together?
- How valuable is my primary goal, compared to the metrics tracked by this monitoring goal?

Here's an example:

Imagine you’re testing a more attention-grabbing CTA on your Homepage. Your primary goal is clicks to the Submit button. But you wonder how this change affects browsing behavior on the Product Categories page. If you track click
events on the search button and pageviews on the Product Categories page, you can evaluate how your sign-up experiment affects purchase behavior. Did visitors sign up, then exit the site? Consider how this tradeoff affects key company metrics and the bottom line.

Evaluate all monitoring goals to look for warnings that you’re cannibalizing another revenue path.

Secondary and monitoring goals provide broad context for immediate lifts and losses. They help you guide your program towards a **global maximum** so you don’t end up refining small parts of your site in isolation. Keep your program focused on providing long-term value to your business.

If your test is taking longer to reach significance, take a look at your primary goal. Is it a high-signal goal?

A **high-signal goal** measures a behavior that’s directly affected by the changes in your variation. A **low-signal goal** is not directly impacted by your test. For example, if you add a value proposition such as free shipping on your product details page, the Add-to-Cart click might be a high-signal goal. Clicks to navigation links or revenue at the end of the checkout funnel are low-signal goals; they aren’t the strongest indicators that your new offer works.

Stats Engine calculates your primary goal independently from secondary and monitoring goals; the primary goal will reach significance faster than if it were pooled with those goals. To ensure that your test reaches significance as quickly as possible, **use your primary goal to measure a high-signal goal**.

If you need to change your primary goal in the middle of your experiment, you can. We don’t recommend making this a regular practice. Stats Engine will recalculate your test with all previous data, as if the new goal was always the primary goal. The old primary goal will be pooled with the secondary goals, so it’ll take longer to reach significance than otherwise.

Adding too many low-signal monitoring goals can also slow down your experiment. So, take stock of what you need to know for the results of your test and long-term planning, and set your goals accordingly! To learn more about setting different types of goals, check out this article.

---

**Seasonality and traffic spikes**

Before you stop the test, check that you’ve captured all the data you need.

If external events or traffic spikes are influencing your results, or if the difference interval of your statistically significant experiment is too large, consider letting your experiment run longer for a more comprehensive test.

Sometimes, optimization teams focus experiments on high-traffic periods or seasons when they make the most money. Testing during traffic surges can help speed up optimization.

But there are a couple of things to watch out for. If you’re testing promising experiences that are likely to generate lift - for instance, seasonal messages during the winter holidays - it might be more effective to translate those experiments into personalization campaigns. By focusing all testing on high-traffic or high-profit periods, you also risk missing part of the conversion cycle; your data will provide an incomplete picture.

For example, imagine you run tests on the weekend because most of your visitors make purchases on Saturday and
Sunday. If you limit your experiment window to the weekend, you assume that visitors encounter your variation and convert within the same period. But it can take multiple visits for a customer to convert.

To capture data from first interaction to final conversion, run your experiment on weekdays as well as the weekend. Design your experiment to optimize the entire conversion cycle.

In general, we recommend testing across your full conversion cycle and through peaks and troughs in traffic.

**Broad difference interval**

Sometimes, when your goal reaches statistical significance, the difference interval may still be relatively large. The *difference interval* is a range of values where the difference between the original and variation actually lies; it tells you what you can expect your conversion rate to be if you run the test again.

For example, if a variation “wins” with a difference interval of 0.1 to 10%, the lift you can expect if you run that variation is within that broad range of values. Decide whether the level of uncertainty is acceptable for your business before you decide to stop the test.

If your primary goal is revenue-generating, a narrower difference interval can help you project the impact of this change more precisely. In other words, you’d be able to predict whether your improvement is worth $1,000 or $1 million. If you're making a business case such as asking for more developer resources to push changes live to your site, it can help to be more specific.

**Segment your results** to see if a certain subset of visitors are moving the needle. Create a new experiment that is targeted specifically to those visitors to see if you can recreate that lift. If this subset of visitors displays consistent behavior over time, your results will show improvement with a smaller difference interval.

If the primary goal is engagement or user acquisition instead of revenue, a large difference interval and more nebulous result may serve your purposes just as well -- a more precise prediction may not make a difference. Since you know that the change led to a better experience in terms of overall conversions, you can feel comfortable pushing the changes live to your site.

Once you’ve analyzed your results and documented what you’ve learned, you’re ready to decide how to take action.