



The statistical methodology behind drafts and experiments

What method does the drafts and experiments team use to calculate confidence intervals and statistical significance?

Jackknife resampling is applied to bucketed data to calculate the sample variance of the percent change of a metric. Two-tailed significance testing is then run using the 95% confidence interval.

Why bucket the data?

Bucketing the data reduces the effects of minor observation errors. If you'd like to know more about why data bucketing is useful, [here's a good place to start](#).

Even if the data is not normally distributed, the bucketed data will be roughly normally distributed based on the Central Limit Theorem, provided there are enough observations per bucket. In order to account for cases where there are not enough observations per bucket, the Jackknife method is used to calculate the confidence interval.

Why use Jackknife resampling?

Jackknife resampling is the standard at Google because it is a versatile method that provides a high level of coverage. It's also effective for detecting outliers and reducing the bias of the sample estimate. Additionally, it's particularly useful in situations where there isn't enough data to get an accurate estimation using the central limit theorem, so it's used on the bucketed data to further increase the accuracy of our confidence intervals.

You can find a general overview of Jackknife resampling [here](#). If you'd like further explanation of its usefulness, [this paper provides more details](#).

Can external advertisers aggregate performance of multiple experiments after the fact, and recalculate the statistics at the aggregate level?

No, advertisers don't have access to user-level data in order to re-create buckets and run the jackknife

algorithm. At the moment, there are no internal tools to do so on behalf of our clients.

Does targeting affect how the auction share split is applied to the experiment and base campaign?

Targeting does not affect the split. The split is applied to eligible auctions before targeting is applied. For example, a 50:50 split will mean that the experiment and control are entered into the same number of auctions.

What are the conditions to ensure a true A/A test?

An A/A test is one in which the experiment and control are identical for the duration of the test (no difference in campaign ads/ad groups/settings, etc. and no differences in ad approvals). Any changes made during the A/A test would need to be made to both the experiment and control arms at the same time.

What are the expected results of an A/A test?

There should be no statistically significant differences in clicks, impressions, CTR, or CPC.

What is the difference between search-based split and cookie-based split?

These are two different options to decide which treatment a user will receive. With search-based experiment splits, users are randomly placed in either the experiment or original campaign every time a search occurs. It's possible that the same user could see both the experiment and your original campaign if they search multiple times. With cookie-based experiment splits, users may see only one version of your campaign, regardless of how many times they search. This can help ensure that other factors don't impact your results.

How many buckets are used?

Twenty buckets are used in the control arm and twenty buckets are used in the treatment arm. If there are too many buckets, then it might take too long to get statistically significant results. If there are too few buckets, then the confidence interval calculations may not be accurate. This strikes a good balance between practical requirements and statistical power.

Yes

No

Need more help?

Try these next steps:



Ask the Help Community












Get answers from community experts



Contact us

Tell us more and we'll help you get there

Test your campaigns

-  About campaign drafts and experiments
 -  About campaign drafts and experiments statuses
 -  Set up a campaign draft
 -  Find and edit your campaign drafts
 -  Apply your draft's changes to the original campaign
 -  Set up a campaign experiment
 -  Find and edit your campaign experiments
 -  Monitor your campaign experiments
 -  Apply a campaign experiment
 -  Apply your ad variation
 -  **The statistical methodology behind drafts and experiments**
-

English



English

