

A/B testing

1. Ask

If different variants of marketing campaigns (V1 vs V2) for both “NewYear” and “BlackFriday” campaigns had significantly better clickthrough rates?

2. Prepare

raw_events table data

adsense_monthly table data

3. Process

SQL query that extracts data from the BigQuery
in Google Spreadsheets

4. Analyze

BlackFriday_V1 vs BlackFriday_V2

Null Hypothesis:

There is **no significant difference** between Control (BlackFriday_V2) and Variant (BlackFriday_V1) clickthrough rates.

Alternative Hypothesis:

There **is significant difference** between Control (BlackFriday_V2) and Variant (BlackFriday_V1) clickthrough rates.

NewYear_V1 vs NewYear_V2

Null Hypothesis:

There **is no significant difference** between Control (NewYear_V2) and Variant (NewYear_V1) clickthrough rates.

Alternative Hypothesis:

There **is significant difference** between Control (NewYear_V2) and Variant (NewYear_V1) clickthrough rates.

4. Analyze

Campaigns: BlackFriday_V1 vs BlackFriday_V2

Test Results

	total	converted	COV%	Standard Error (std)
control	24,276	24	0.10%	0.02%
variant	8,220	8	0.10%	0.03%

95% Confidence Interval (CI) Limits		99% Confidence Interval (CI) Limits	
From	To	From	To
0.06%	0.14%	0.05%	0.15%
0.03%	0.16%	0.01%	0.19%

Distribution of probabilities

Control BlackFriday_V2
Variant BlackFriday_V1

Probability that Variant is better than the control **53.11%**

Results	
Z Score	0.04
P value	0.515399976

Significant at 95 % **NO**
Significant at 99% **NO**

Uplift

Mean Uplift	11.06%
Median Uplift	3.94%
30th percentile	-16.20%

Always aim for a conservative uplift

Campaigns: NewYear_V1 vs NewYear_V2

Test Results

	total	converted	COV%	Standard Error (std)
control	13,039	30	0.23%	0.04%
variant	4,430	57	1.29%	0.17%

95% Confidence Interval (CI) Limits		99% Confidence Interval (CI) Limits	
From	To	From	To
0.15%	0.31%	0.12%	0.34%
0.95%	1.62%	0.85%	1.72%

Distribution of probabilities

Control NewYear_V2
Variant NewYear_V1

Probability that Variant is better than the control **100.00%**

Results	
Z Score	-6.06
P value	0.000000001

Significant at 95 % **YES**
Significant at 99% **YES**

Uplift

Mean Uplift	477.06%
Median Uplift	458.97%
30th percentile	397.40%

Always aim for a conservative uplift

Standard error

is an estimate of **how far the sample mean is likely to be from the population mean.**

In practical terms, **a smaller standard error** indicates that the sample statistic is likely to be closer to the true population parameter, providing **more confidence** in the results of your A/B test.

Standard error decreases as we increase the **sample size**.

Confidence Intervals

is a range within which we can reasonably expect the true population parameter to lie.

For example (NewYear_V2 Control): with **95% of certainty** we are sure that our data falls between range 0.15% and 0.31%. **If we increase confidence level the range increases as well:** with **99% of certainty** we are sure that our data falls between range 0.12% and 0.34%.

The bigger is **standard error** the wider ranges will be. And the wider ranges are the more variant we will have.

Z-score

Z-score is a statistical measure that quantifies **how many standard deviations a data point is from the mean** of a group of data.

In practical terms, a **Z-score helps us determine whether the observed difference between two groups is likely due to a real effect or if it could have occurred by random chance.**

A **higher absolute Z-score** indicates a more extreme deviation from the null hypothesis (no difference), **suggesting a higher level of confidence** in the observed effect.

P-value

P-value is a statistical measure that helps you determine the significance of the observed differences between two groups.

A small p-value (typically less than 0.05) indicates that the observed differences are unlikely to be due to random chance alone, leading to the **rejection of the null hypothesis**.

This suggests that there is evidence in favor of the alternative hypothesis, supporting the idea that **there is a significant difference** between the groups being compared.

5. Share

Campaigns: BlackFriday_V1 vs BlackFriday_V2

Test result

The test result is *not* significant.

The observed difference in conversion rate (-1.56%) isn't big enough to declare a significant winner. There is no real difference in performance between A and B or you need to collect more data.



<https://abtestguide.com/calc/>

Fail to reject the Null Hypothesis

There is **no difference** between A and B.

Because little random things could easily shift the result from Control V2 to Variant V1 then back again. And thus data does not overwhelmingly convince us to reject the Null Hypothesis.

Campaigns: NewYear_V1 vs NewYear_V2

Test result

Significant test result!

Variation B's observed conversion rate (1.29%) was 459.23% higher than variation A's conversion rate (0.23%). You can be 95% confident that this result is a consequence of the changes you made and not a result of random chance.



To reject the Null Hypothesis

There is **difference** between A and B.

Because little random things would not change the results very much if we tested the A and B on a lot of people.

6. Act

Actionable recommendations

Campaigns: **BlackFriday_V1** vs BlackFriday_V2

1. Collect more data - similar sample sizes would be better
2. or have better conversion rates.
3. Choose the one that cost less - **V1**, because they are similar.

Campaign	Impressions	New_clicks	CVR%	Cost
BlackFriday_V1	8220	8	0.10%	90.42
BlackFriday_V2	24276	24	0.10%	364.14

Campaigns: **NewYear_V1** vs NewYear_V2

1. Collect more data in order to have similar sample sizes. Also better conclusions from bigger sample size.
2. Choose **V1** because it is better than V2. V1 cost less than V2 as well.

Campaign	Impressions	New_clicks	CVR%	Cost
NewYear_V1	4430	57	1.29%	44.3
NewYear_V2	13039	30	0.23%	156.465

Thank You