

[Get started](#)[Open in app](#)488K Followers · [About](#) [Follow](#)

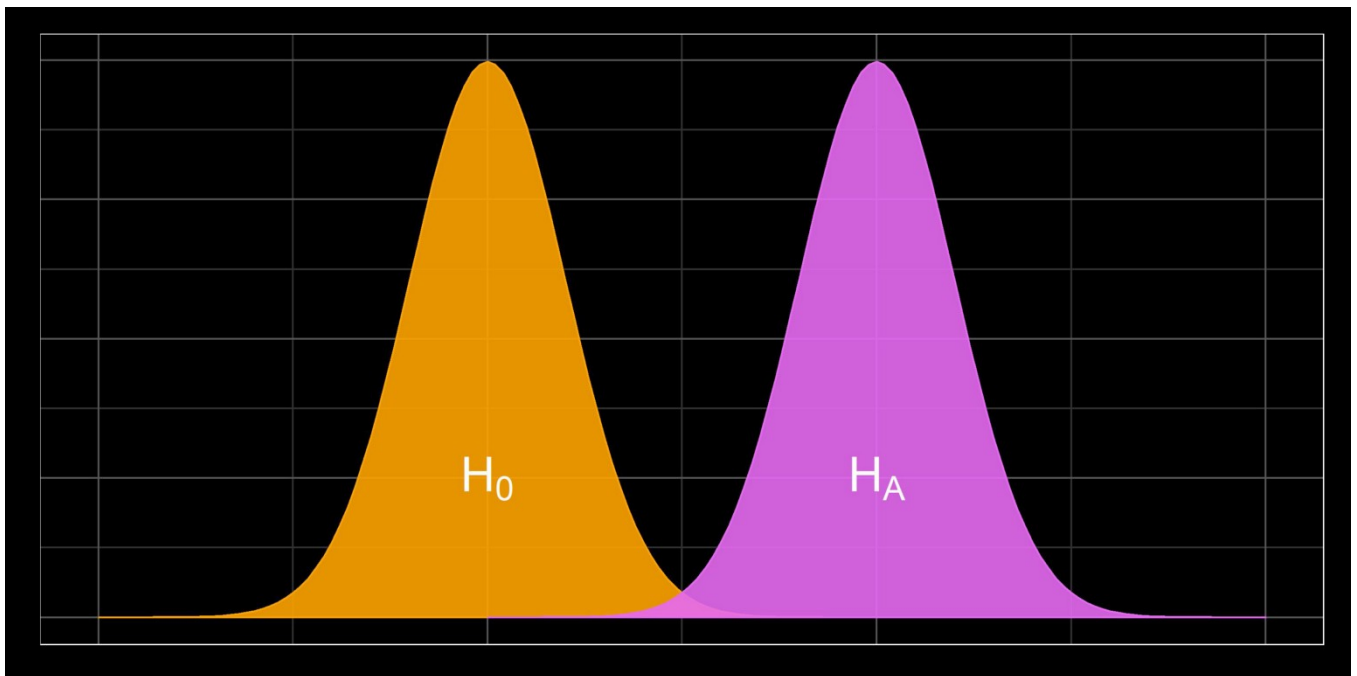
You have **1** free member-only story left this month. [Sign up for Medium and get an extra one](#)

# A/B Test Statistics Made Easy

## Part 2: Proportion Metrics



Rezwan Hoppe-Islam · Apr 16 · 7 min read ★



**T**his article will demonstrate how to evaluate the results of an A/B test when the output metric is a proportion, such as customer conversion rate. Here we exploit the statistical foundations that we built in [Part 1: Continuous Metrics](#), so start there if you need a refresher.

Statistics can be a daunting subject to learn, and it doesn't have to be. The aim of these guides is to make statistical tools more accessible to people. This guide uses simulation

data to demystify the subject and bring key statistical concepts to life. Enjoy!

## 1. Simulating the conversion rates of 2 identical websites

Imagine that we own 2 websites. Each website converts visitors into customers at the rate of 30%. Let's see what happens after 1,000 visitors visit each website:

Sample	Website 1: Customers per 1,000 visitors	Website 2: Customers per 1,000 visitors	Difference
1	295	297	2

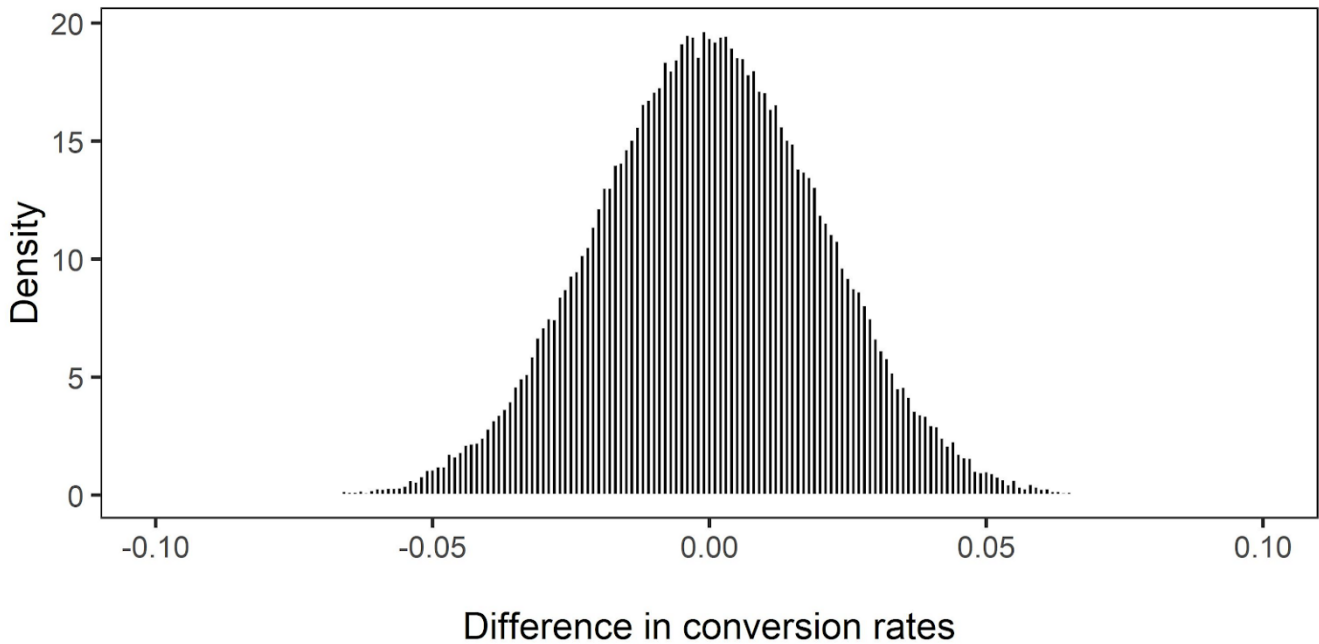
Now we run a simulation that can repeat the above process for a total of 100,000 samples. Here's a peek at the results:

Sample	Website 1: Customers per 1,000 visitors	Website 2: Customers per 1,000 visitors	Difference
1	295	297	2
2	307	306	-1
...	...	...	...
100,000	284	297	13

So even though we know that the true conversion rate is 30%, we see some variation around this number due to random *sampling error*. Now let's convert all the numbers above into proportions:

Sample	Website 1: Customer conversion rate	Website 2: Customer conversion rate	Difference in conversion rates
1	0.295	0.297	0.002
2	0.307	0.306	-0.001
...	...	...	...
100,000	0.284	0.297	0.013

And finally, let's plot the distribution of the difference in conversion rates (the last column):



**Recap:** We have two websites with identical conversion rates. So the difference in conversion rates should equal zero. But due to sampling error, we see a sampling distribution of results around zero.

## 2. Applying simulation to an A/B test

Now imagine that an A/B test yields the following results:

Treatment group	Sample size	Customer conversions	Conversion rate
Test	1,000	335	0.335
Control	1,000	290	0.290
<b>Difference (Test - Control)</b>		<b>45</b>	<b>0.045</b>

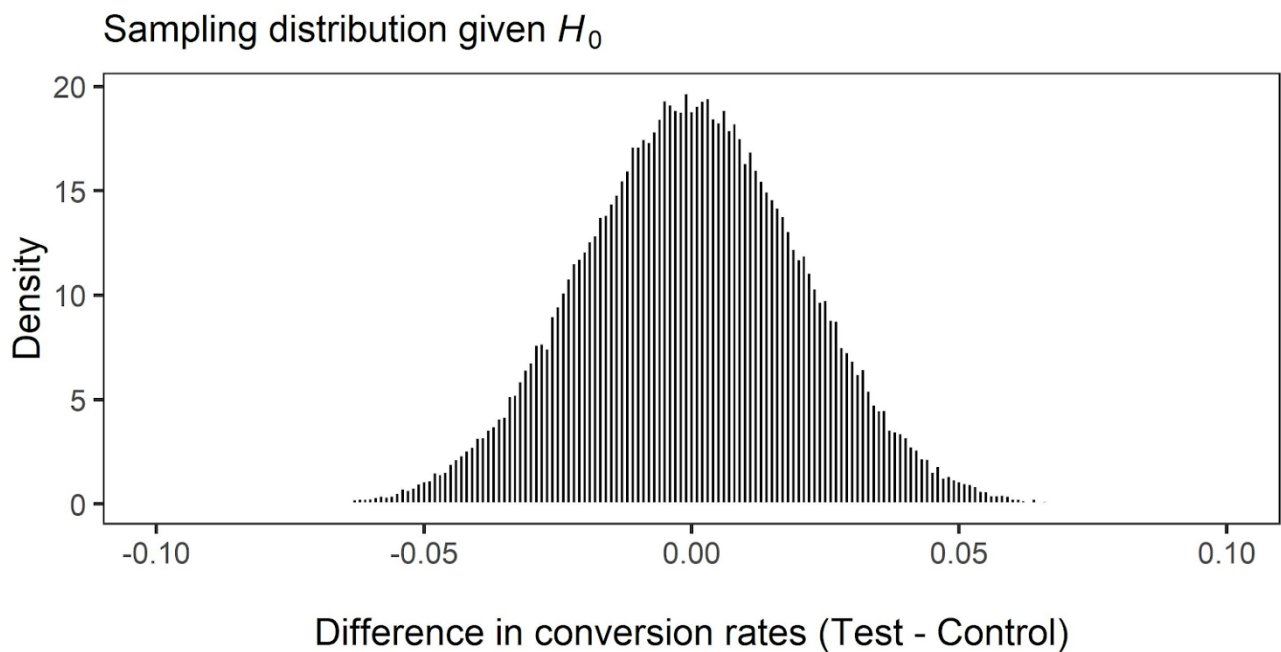
Has the Test treatment really improved the conversion by 45 customers? Or do the Test and Control groups have identical conversion rates? In which case any difference we observe is just a result of random sampling error?

Let's capture these hypotheses more formally:

- **$H_0$  (null hypothesis):** The difference in population conversion rates = 0. Any observed difference is due to sampling error.

- ***HA (alternative hypothesis)***: The difference in population conversion rates  $\neq 0$ . It is unlikely that the observed difference is due to sampling error alone.

So let's begin by determining what the sampling distribution would look like given the null hypothesis. All we need to know is what the true (population) conversion rate might be. Our best guess would be the average of the sample conversion rates:  $(0.335 + 0.290) / 2 = 0.3125$ . And now let's simulate the distribution that we would see if in fact both Test and Control had conversion rates of 0.3125:



**Recap:** We have again used simulation to generate the sampling distribution. Here we assume that both Test and Control populations have the same conversion rate — this is our null hypothesis.

### 3. A faster way to generate the sampling distribution

Simulation is a great tool for generating a sampling distribution to demonstrate the impact of sampling error. If we want to move faster, we can, in fact, *compute* the sampling distribution as follows:

**Step 1: Calculate the pooled sample proportion ( $p$ ).**

We already calculated this to be 0.3125.

Here's the actual formula which is especially useful when the Test and Control groups have different sample sizes:

```
x1 = Test conversions (335)
x2 = Control conversions (290)
n1 = Test sample size (1000)
n2 = Control sample size (1000)

p = (x1 + x2) / (n1 + n2)
  = (335 + 290) / (1000 + 1000)
  = 625 / 2000
  = 0.3125
```

**Step 2: Use  $p$  to calculate the standard error ( $se$ ).**

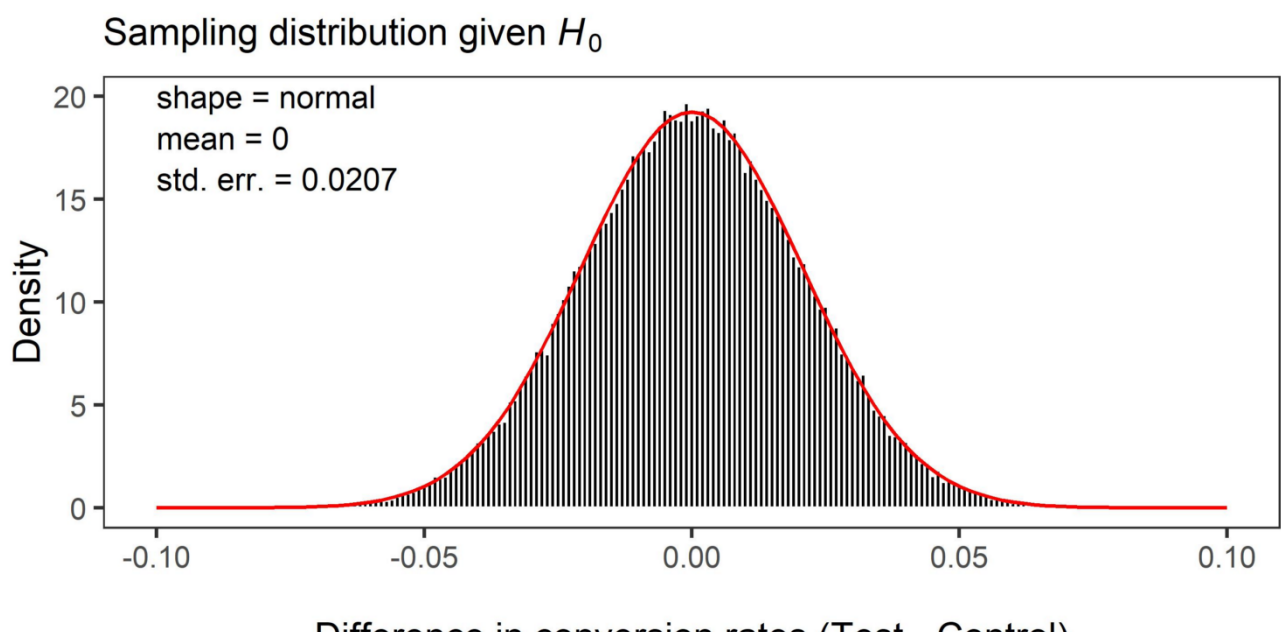
```
se = sqrt(p * (1-p) * [(1/n1) + (1/n2)])
    = sqrt(0.3125 * (1-0.3125) * [(1/1000) + (1/1000)])
    = sqrt(0.3125 * 0.6875 * 2/1000)
    = 0.0207
```

**Step 3: Compute a normal distribution with mean = 0, standard deviation = 0.0207.**

In R we would simply write:

```
rnorm(n = 10000, mean = 0, sd = 0.0207)
```

Let's check to see what this distribution of computed values (red) looks like, compared to the previous distribution that we created through simulation (black):



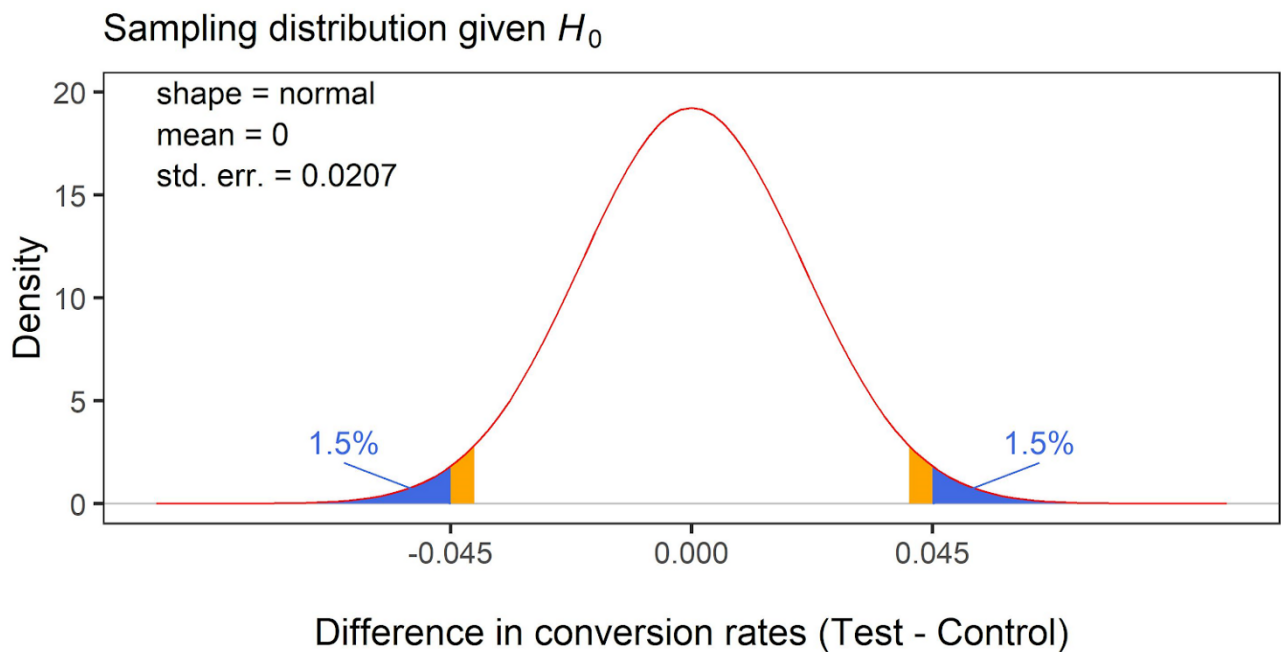
## Difference in conversion rates (Test - Control)

As you can see we have an almost perfect match! In fact, simulation leads to an *estimate* of the sampling distribution. The computed sampling distribution, however, is what we would obtain if we ran the simulation an infinite number of times. Therefore, it is the *exact* sampling distribution.

**Recap: We no longer have to run many simulations to estimate the sampling distribution. We can now compute the precise sampling distribution.**

### 4. Interpreting the A/B test results

So, given the sampling distribution above, what is the probability of obtaining a result as extreme or more as 0.045, under the null hypothesis? Let's re-examine our sampling distribution:



Here our **p-value** (shaded in blue) is 3% and it tells us that if the null hypothesis were true then we would expect to see a result as extreme or more as 0.045 in only 3 out of 100 A/B tests. And since our p-value is less than our **significance level ( $\alpha$ )** of 0.05, we reject the null hypothesis.

**Recap: The chance of seeing an improvement in conversion rate as extreme or more as 0.045 is improbable under the null hypothesis. Specifically, our p-value**

of 0.03 means that we're only likely to see a result as extreme or more as 0.045 in 3 out of 100 experiments. This ratio is less than our significance level ( $\alpha$ ) of 0.05 and, therefore, we reject the null hypothesis.

## 5. A faster way to calculate p-values

We can speed up the process even more by just making use of the *standard normal distribution*. We can transform our distribution into the standard normal distribution by dividing by the standard error. Likewise, we can also transform our test result into a *z score* by dividing by the standard error. So here we get  $0.045 / 0.0207 = 2.17$ .

Using this z-score, we can now quickly determine the p-value via a look-up table, or by calculating it using a programming language such as *R*:

```
pnorm(q = -2.17, mean = 0, sd = 1) * 2
```

In both cases, we arrive at the same p-value as we had before: 0.03.

**Recap:** By converting our observed difference in conversion rates (0.045) into a z score (2.17), we were able to exploit the use of a look-up table against the standard normal distribution to rapidly arrive at the p-value.

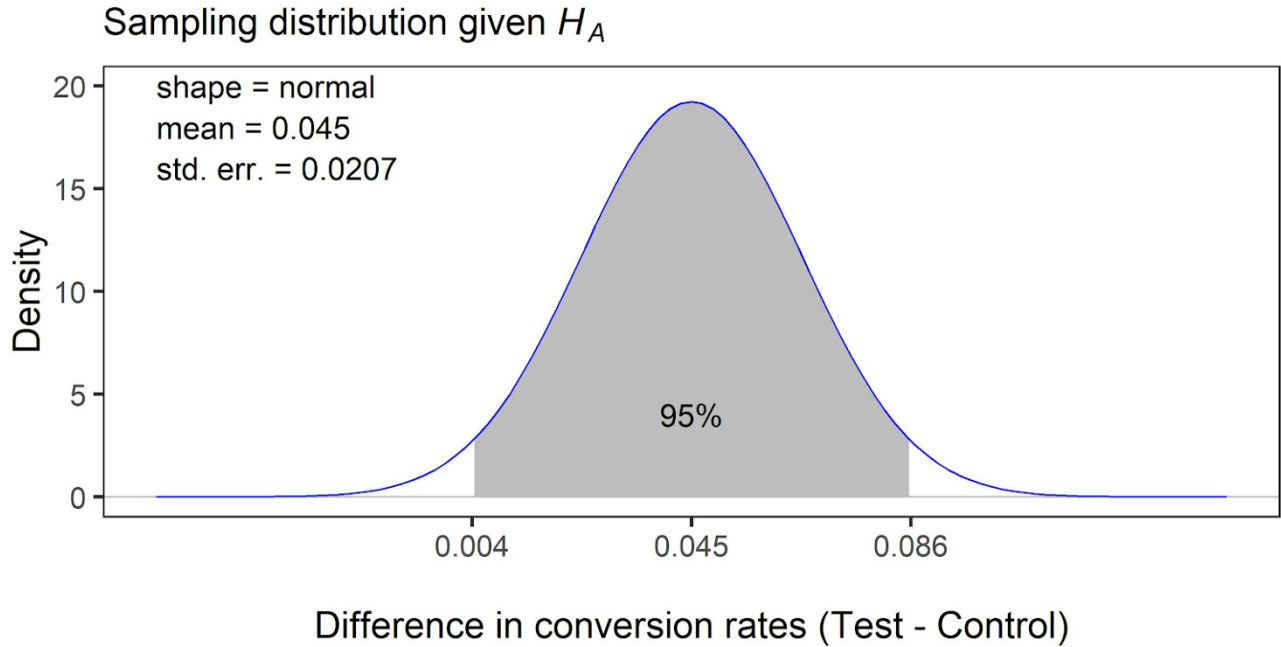
## 6. The confidence interval

To recap:

- We have *rejected* the null hypothesis (***H0***) that the difference in population conversion rates = 0.
- Therefore, we must *accept* the alternative hypothesis (***HA***) that the difference in population conversion rates  $\neq 0$ .

If we accept that difference in population conversion rates *does not* equal zero then what *does* it equal? What is the range of possible values given the alternative

hypothesis? Let's refer to the sampling distribution given the alternative hypothesis:



The area shaded in grey represents the middle 95% of the chart which corresponds to a confidence interval of (0.004, 0.086) or  $0.045 \pm 0.041$ . Every confidence interval has an associated confidence level which is predetermined by the significance level ( $\alpha$ ) and is equal to  $1 - \alpha = 1 - 0.05 = 95\%$ . The confidence interval tells us that if we were to repeat the entire A/B test many times then, in the long run, 95% of our confidence intervals would capture the true difference between population conversion rates.

**Recap: A confidence interval (at an associated confidence level) shows the range of possible values for our A/B test result. Here, we report a confidence interval of  $0.045 \pm 0.041$  at a confidence level of 95%.**

## 7. Statistical analyses in R

By this point, you should feel fairly comfortable with the overall framework of a statistical analysis. Now let's see how we can reduce all the above-mentioned steps into just a few lines of code. In R we would write:

This produces:





## 8. Summary

Here we explored the steps involved in analyzing an A/B test, together with a framework that ties all the steps together. Hopefully, this has left you with a greater sense of comfort around A/B test statistics. In this scenario, we had sufficient evidence (in terms of a p-value) to reject the null hypothesis. But what if we hadn't collected a large enough sample to yield a p-value  $< 0.05$ ? In the [next article](#), we'll demonstrate how you can deduce ahead of time your required sample size.

---

### Sign up for The Daily Pick

By Towards Data Science

Hands-on real-world examples, research, tutorials, and cutting-edge techniques delivered Monday to Thursday. Make learning your daily ritual. [Take a look](#)

Your email

---

Get this newsletter

By signing up, you will create a Medium account if you don't already have one. Review our [Privacy Policy](#) for more information about our privacy practices.

[About](#) [Help](#) [Legal](#)

Get the Medium app

