

Code ▾

A/B Testing

Let us consider two ad-copies that you are testing and they produced values as below

Hide

```
set.seed(42)
# simulate data
test <- rbinom(n=100000,size = 1,prob = 0.02)
control <- rbinom(n=90000,size = 1, prob = 0.033)
```

Hide

```
get_ctr <- function(vec){
  return (sum(vec)*100/length(vec))
}
test_ctr <- get_ctr(test)
control_ctr <- get_ctr(control)
print(test_ctr)
```

```
[1] 2.032
```

Hide

```
print(control_ctr)
```

```
[1] 3.365556
```

t test for two population:

we know sample mean of each experiment is the CTR we need compare. According to CTR, sample mean is limiting normal distributed. Then we have

$$\bar{X}_1 \sim N(p_1, p_1(1 - p_1)/n_1)$$

$$\bar{X}_2 \sim N(p_2, p_2(1 - p_2)/n_2)$$

$$\bar{X}_1 - \bar{X}_2 \sim N(p_1 - p_2, \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2})$$

Hypothesis

$$H_0 : \bar{X}_1 - \bar{X}_2 = 0, H_a : \bar{X}_2 - \bar{X}_1 > 0$$

$$t = \frac{\bar{X}_1 - \bar{X}_2 - 0}{SE} = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Before running the t-test, first, ensure that your data pass a test of homoscedasticity—are the variances homogenous? We do this in R with a Fisher's F-test, `var.test(x, y)`.

Hide

```
var.test(test, control)
```

F test to compare two variances

```
data: test and control
F = 0.61209, num df = 99999, denom df = 89999, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6043475 0.6199389
sample estimates:
ratio of variances
 0.6120949
```

If your $p > 0.05$, then you can assume that the variances of both samples are homogenous. In this case, we run a classic Student's two-sample t-test by setting the parameter `var.equal = TRUE`.

Hide

```
agep <- t.test(test, control, var.equal = TRUE, alternative = "less")
agep
```

Two Sample t-test

```
data: test and control
t = -18.04, df = 190000, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -0.01211967
sample estimates:
 mean of x mean of y
0.02032000 0.03365556
```

For a one-sided test, p-value is less than 5%. Hence, we reject the null hypothesis that test is not inferior to control.

Using Chi-squared test

Chi-square test can be used for two purposes:

- Goodness of fit test. When you have a categorical variable from a single population, test whether sample data are consistent with a hypothesis.
- Test for independence. When you have two categorical variables from a single population, test whether there is a significant association between the two variables.

Hypothesis:

H0: Variable Ad type and variable Whether click are independent

Ha: Variable Ad type and variable Whether click are not independent

The chi-squared test is essentially always a one-sided test. Here is a loose way to think about it: the chi-squared test is basically a 'goodness of fit' test. Sometimes it is explicitly referred to as such, but even when it's not, it is still often in essence a goodness of fit. For example, the chi-squared test of independence on a 2 x 2 frequency table is (sort of) a test of goodness of fit of the first row (column) to the distribution specified by the second row (column), and vice versa, simultaneously. Thus, when the realized chi-squared value is way out on the right tail of it's distribution, it indicates a poor fit, and if it is far enough, relative to some pre-specified threshold, we might conclude that it is so poor that we don't believe the data are from that reference distribution.

To perform the chi-squared test, you need the number of successes and failures for each variation. This makes what's called a contingency table, something like this:

[Hide](#)

```
library(knitr)
Clicks <- c(sum(test), sum(control))
Impressions <- c(length(test), length(control))
NonClicks <- Impressions - Clicks
ad <- data.frame(Clicks, NonClicks, Impressions)
row.names(ad) <- c('Test', 'Control')
kable(ad, align = 'l')
```

	Clicks	NonClicks	Impressions
Test	2032	97968	1e+05
Control	3029	86971	9e+04

[Hide](#)

```
print(ad[1:2, 1:2])
```

	Clicks <int>	NonClicks <int>
Test	2032	97968
Control	3029	86971
2 rows		

[Hide](#)

```
# method 2: use chisq.test directly
chisq.test(ad[1:2, 1:2], correct = FALSE)
```

Pearson's Chi-squared test

```
data: ad[1:2, 1:2]
X-squared = 324.9, df = 1, p-value < 2.2e-16
```

Again, p-value is less than 0.05, so, we reject the null hypothesis. That is, variable Whether click and Ad type has no relationship is false, which means we can say that test group is drawn from a different distribution than the control group. Chi-square test got the consistent result with t-test.

Yates's correction for continuity

Using the chi-squared distribution to interpret Pearson's chi-squared statistic requires one to assume that the discrete probability of observed binomial frequencies in the table can be approximated by the continuous chi-squared distribution. This assumption is not quite correct and introduces some error.

To reduce the error in approximation, Frank Yates suggested a correction for continuity that adjusts the formula for Pearson's chi-squared test by subtracting 0.5 from the absolute difference between each observed value and its expected value in a 2×2 contingency table. This reduces the chi-squared value obtained and thus increases its p-value.

More reading <https://www.bounteous.com/insights/2014/07/01/statistical-significance-test/>
(<https://www.bounteous.com/insights/2014/07/01/statistical-significance-test/>)