# A/B Testing - Fundamentals

# Introduction

A/B test is a hypothesis test to compare two populations

# Introduction

What is a test of hypothesis?

# Test of hypothesis

Statistical procedure

to infer about a population

using random samples

# Test of hypothesis

Statistical procedure

to infer about a population *parameter*

using random samples *values*

# Test of hypothesis

Statistical procedure

to infer about population(s)

using random samples

# Test of hypothesis

Statistical procedure

to infer about population(s) *parameters*

using random samples *values*

# Introduction

Conjecture

- $\mu > \mu_0$ The average number of daily visits exceeds $\mu_0$
- $\mu_1 > \mu_2$ The average number of daily visits of design 1 exceeds that of design 2
- $p > p_0$ The proportion of daily visits exceeds $p_0$
- $p_1 > p_2$ The proportion of daily visits of design 1 exceeds that of design 2

# Introduction

Conjecture

- $\mu > \mu_0$ The average number of daily visits exceeds $\mu_0$
- $\mu_1 > \mu_2$ The average number of daily visits of design 1 exceeds that of design 2
- $p > p_0$ The proportion of daily visits exceeds $p_0$
- $p_1 > p_2$ The proportion of daily visits of design 1 exceeds that of design 2

# Introduction

Two hypotheses

Ho: null hypothesis $\quad = \quad \text{always}$

Ha: alternative hypothesis $\begin{cases} \neq & \text{two-tail test} \\ > & \text{right-tail test} \\ < & \text{left-tail test} \end{cases}$

# Performing a Test of hypothesis

- Define the objective
- Identify the test needed
- Collecting the data (random sampling)
- Statistical analysis
- Conclusion (favoring Ho or Ha)

# Introduction

Conjecture test on means

- $\mu > \mu_0$ The average number of daily visits exceeds $\mu_0$
- $\mu_1 > \mu_2$ The average number of daily visits of design 1 exceeds that of design 2

test on proportions

- $p > p_0$ The proportion of daily visits exceeds $p_0$
- $p_1 > p_2$ The proportion of daily visits of design 1 exceeds that of design 2

# Elements of a Test of hypothesis

- Test statistic (TS)
- Observed test statistic (OTS)
- Regions (RR, AR)
- Critical value (CV)
- p-value

# Elements of a Test of hypothesis

- Test statistic (TS) — r.v. whose distribution is related to the parameter of interest
- Observed test statistic (OTS) — numeric value of the TS found after sampling
- Regions (RR, AR) — RR interval leading to reject Ho; AR interval leading to accept Ho
- Critical value (CV) — boundary between RR and AR
- p-value — Probability P[TS > OTS | Ho true]

# Elements of a Test of hypothesis

p-value     Probability P[TS > OTS | Ho true]

# Elements of a Test of hypothesis

p-value Probability P[TS > OTS | Ho true]

p-value is the probability
of obtaining results
as extreme as the observed results
when Ho is true

# Analyzing the data

- Select the TS
- Use $\alpha$ to find the critical value (quantile)
- Find the OTS
- Find the p-value

# Random variables in R

- *R* identifies a random variable by a label
- It provides four functions for each one
- The functions are identified by a prefix

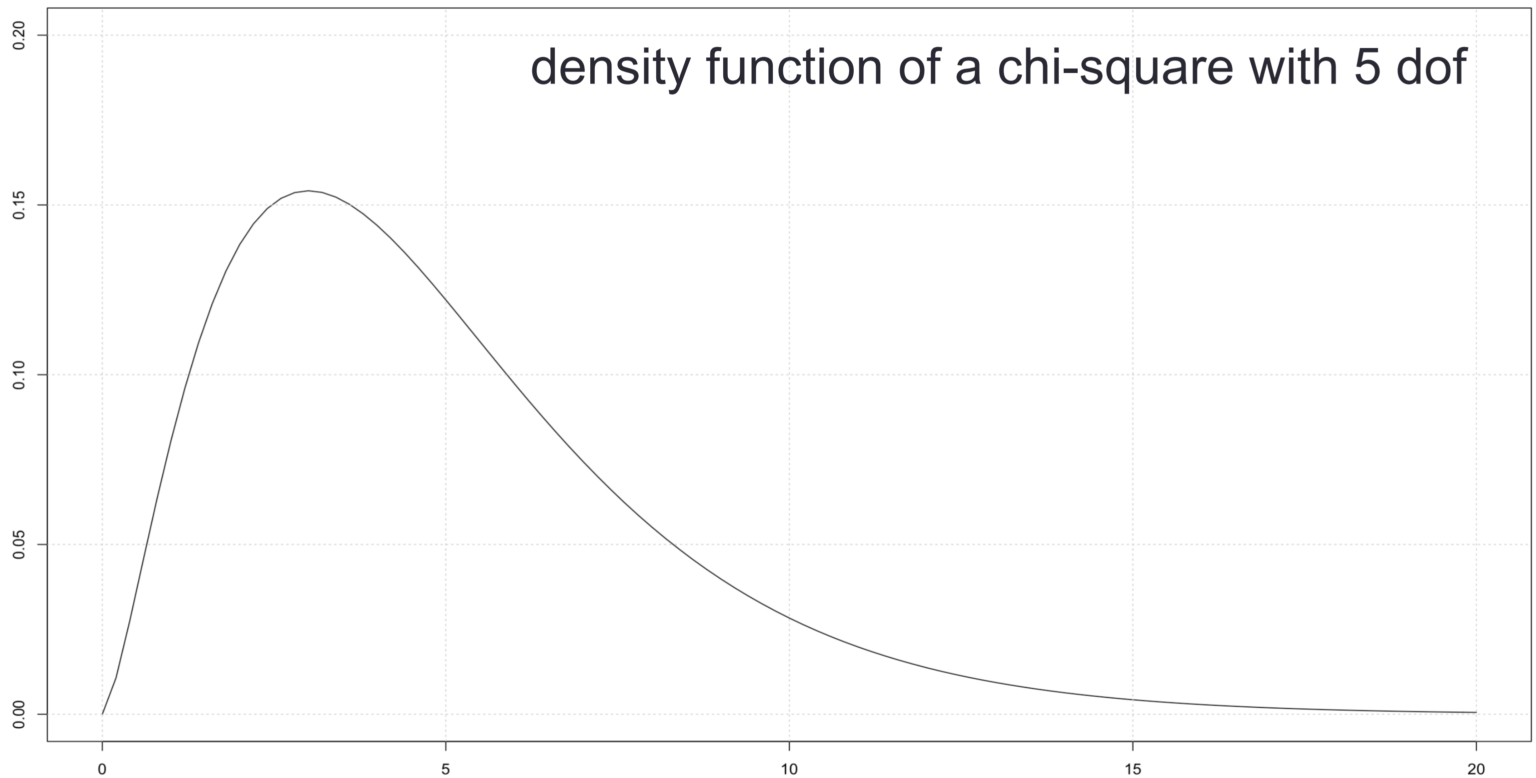
  d probability or density function value

  p distribution function

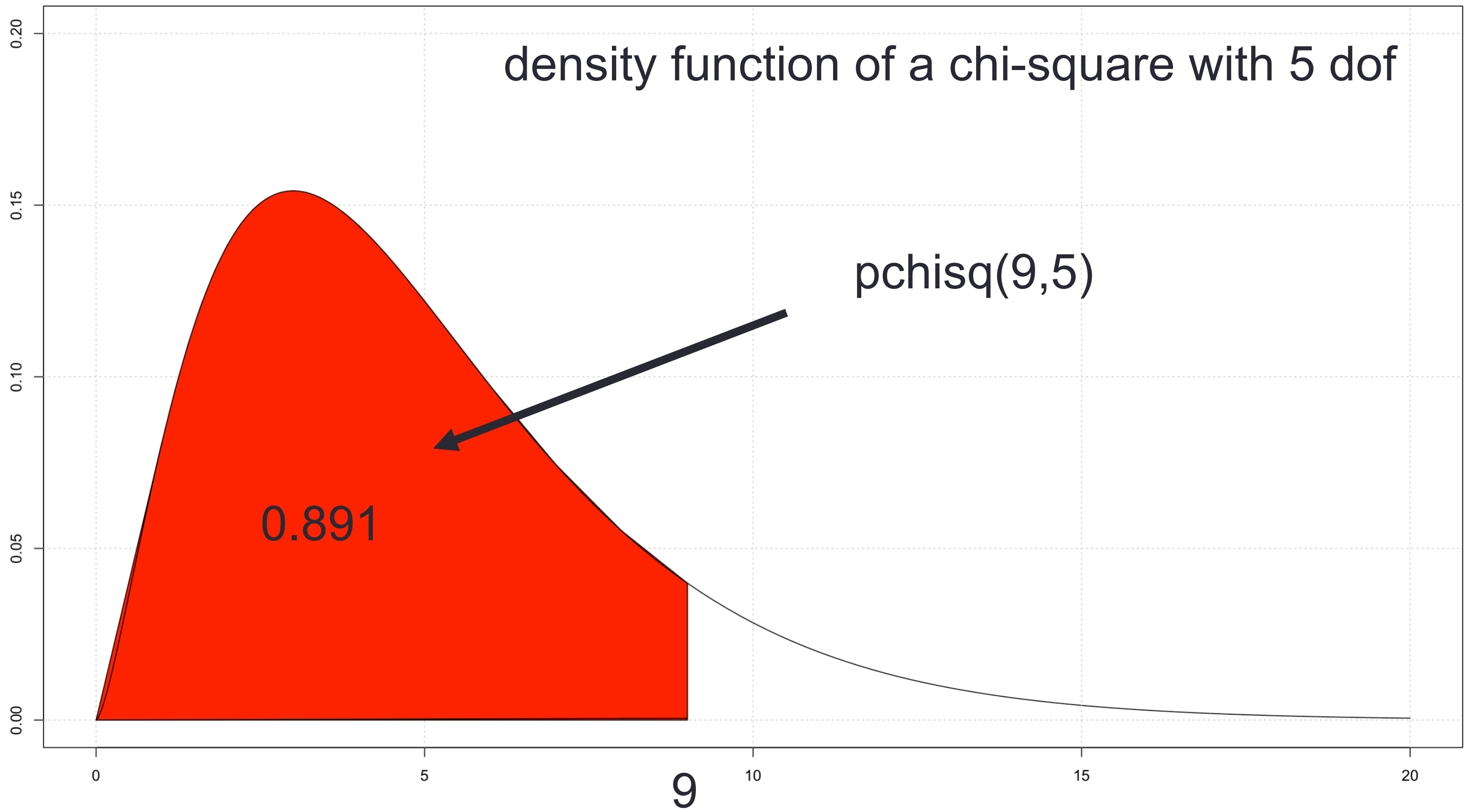
  q quantile

  r random sampling

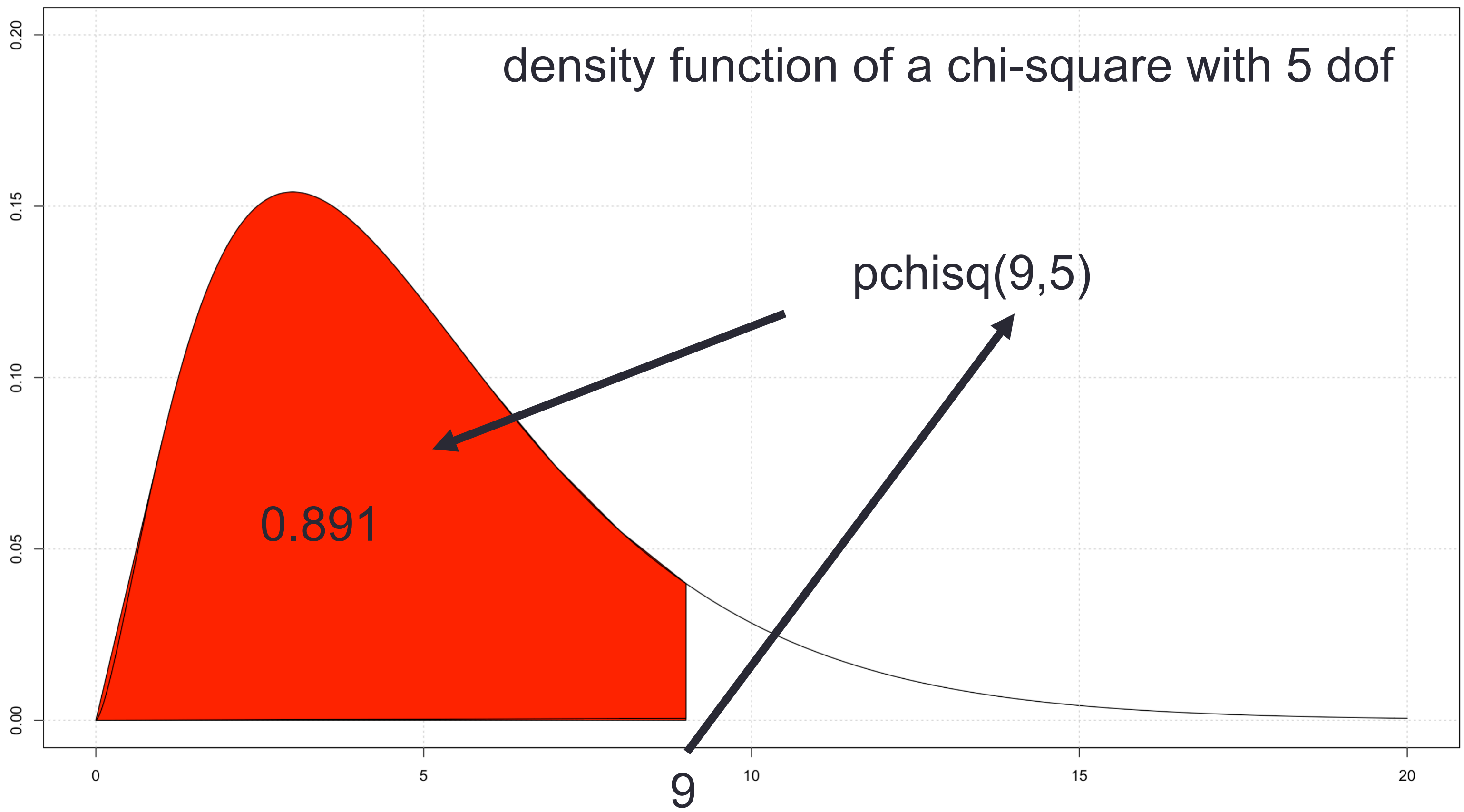
# Chi-square label is chisq( )



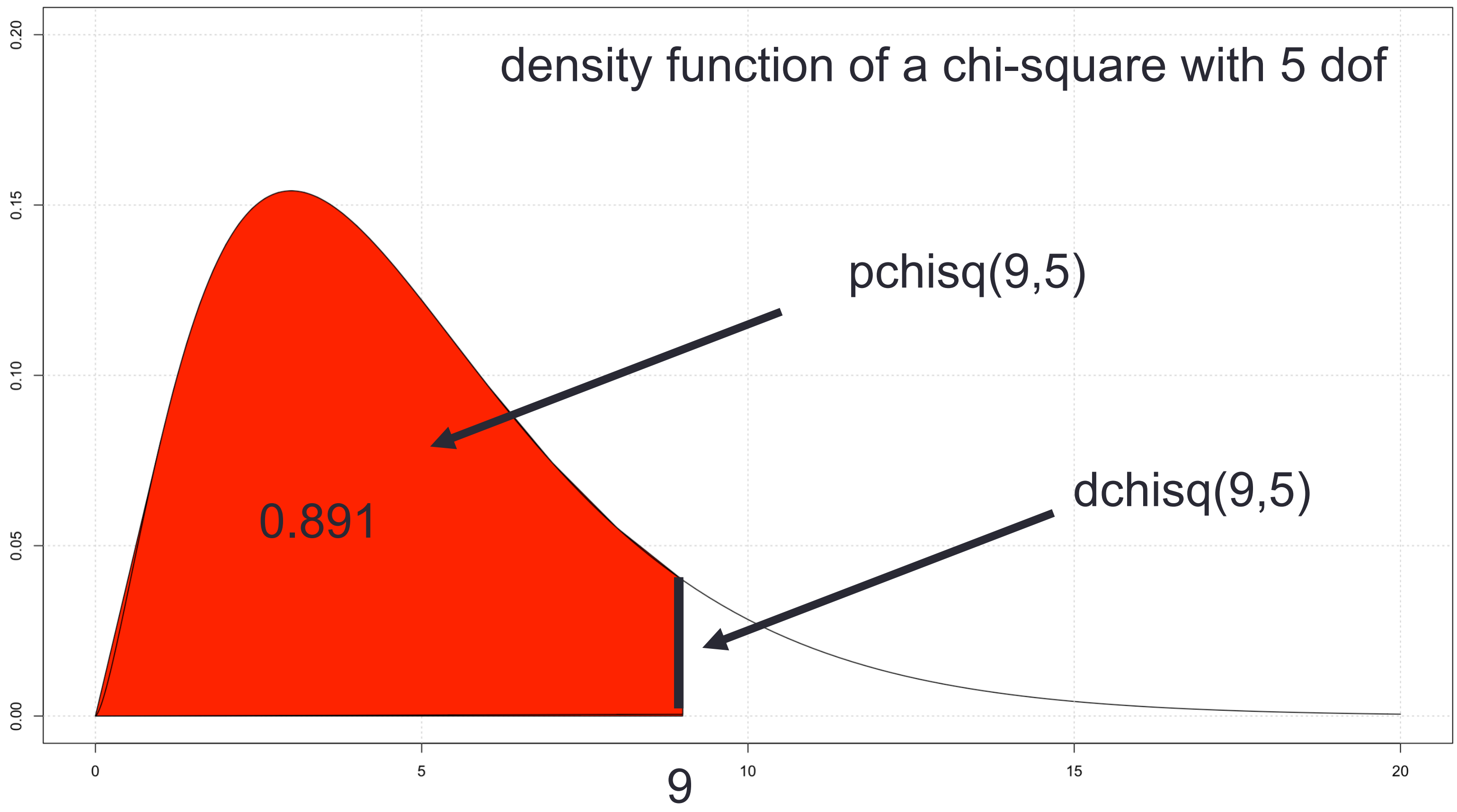
# Chi-square label is chisq( )



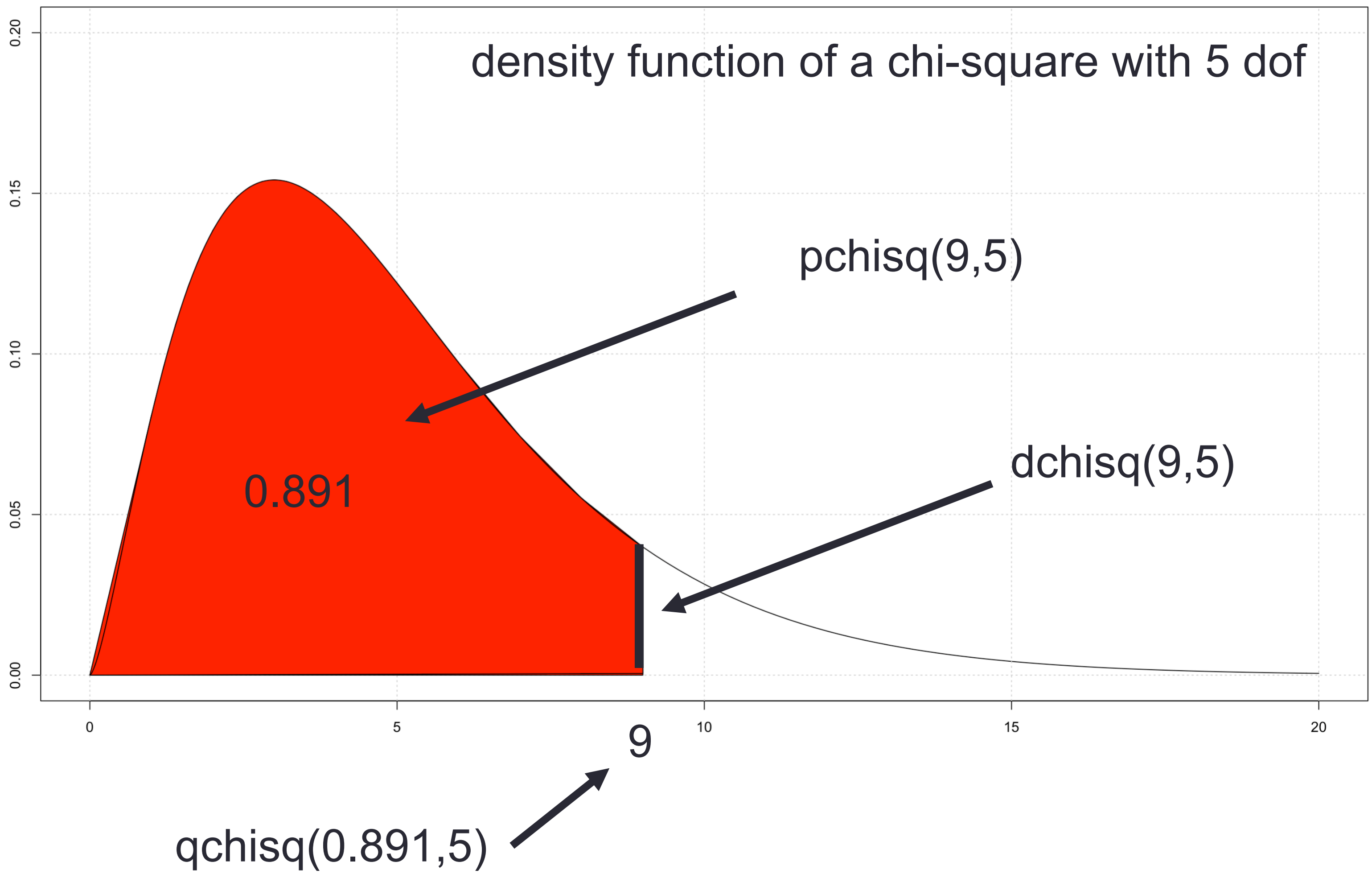
# Chi-square label is chisq( )



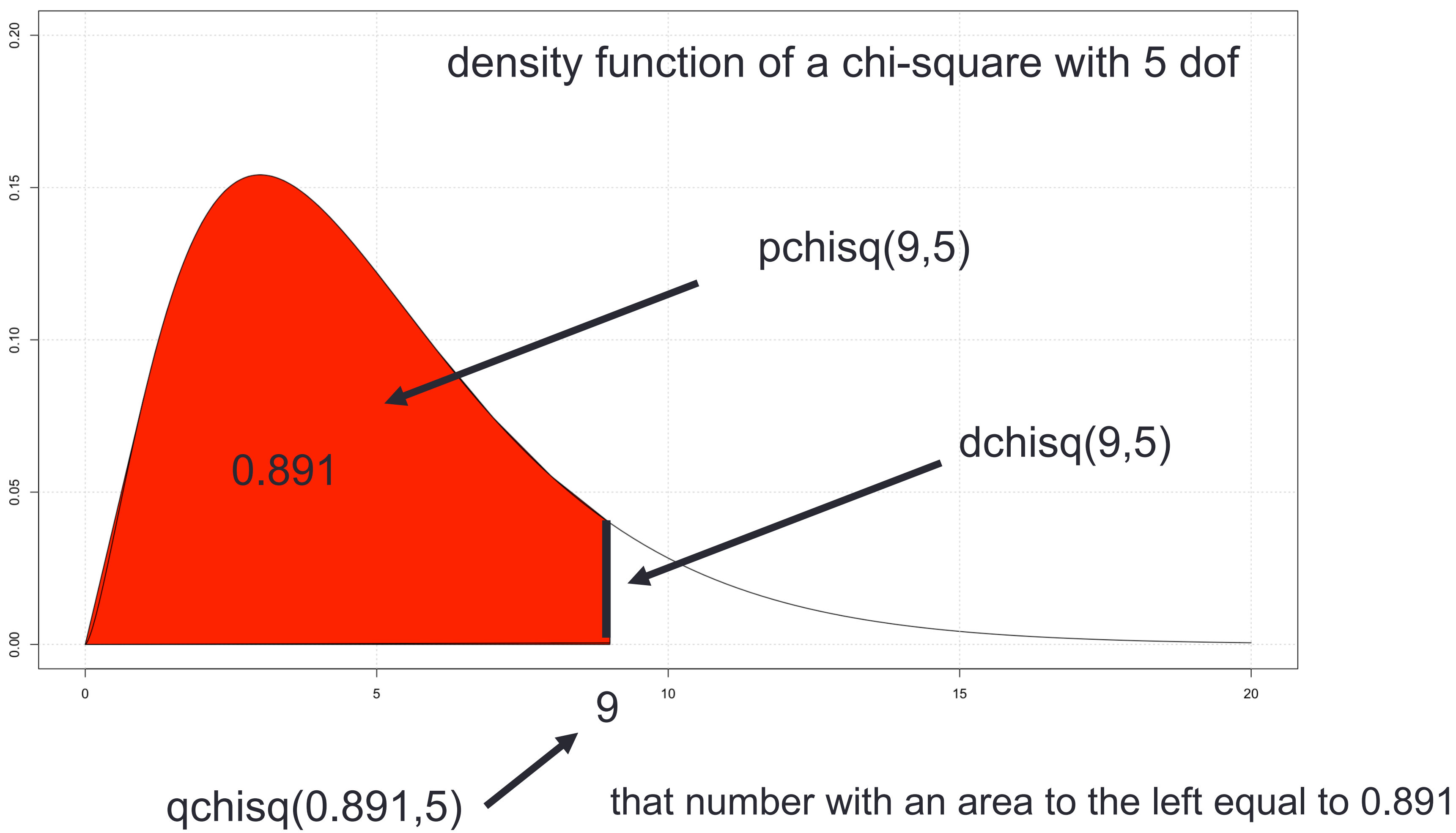
# Chi-square label is chisq( )



# Chi-square label is chisq( )



# Chi-square label is chisq( )



# Chi-square with 5 degrees of freedom

```
> # P[X<9]
> pchisq(9,5)
[1] 0.8909358
> # density height at 9
> dchisq(9,5)
[1] 0.03988664
> # P[X < a] = 0.891
> qchisq(0.891,5)
[1] 9.001609
> # 3 random observations
> rchisq(3,5)
[1] 2.474736 3.504116 2.905611
```

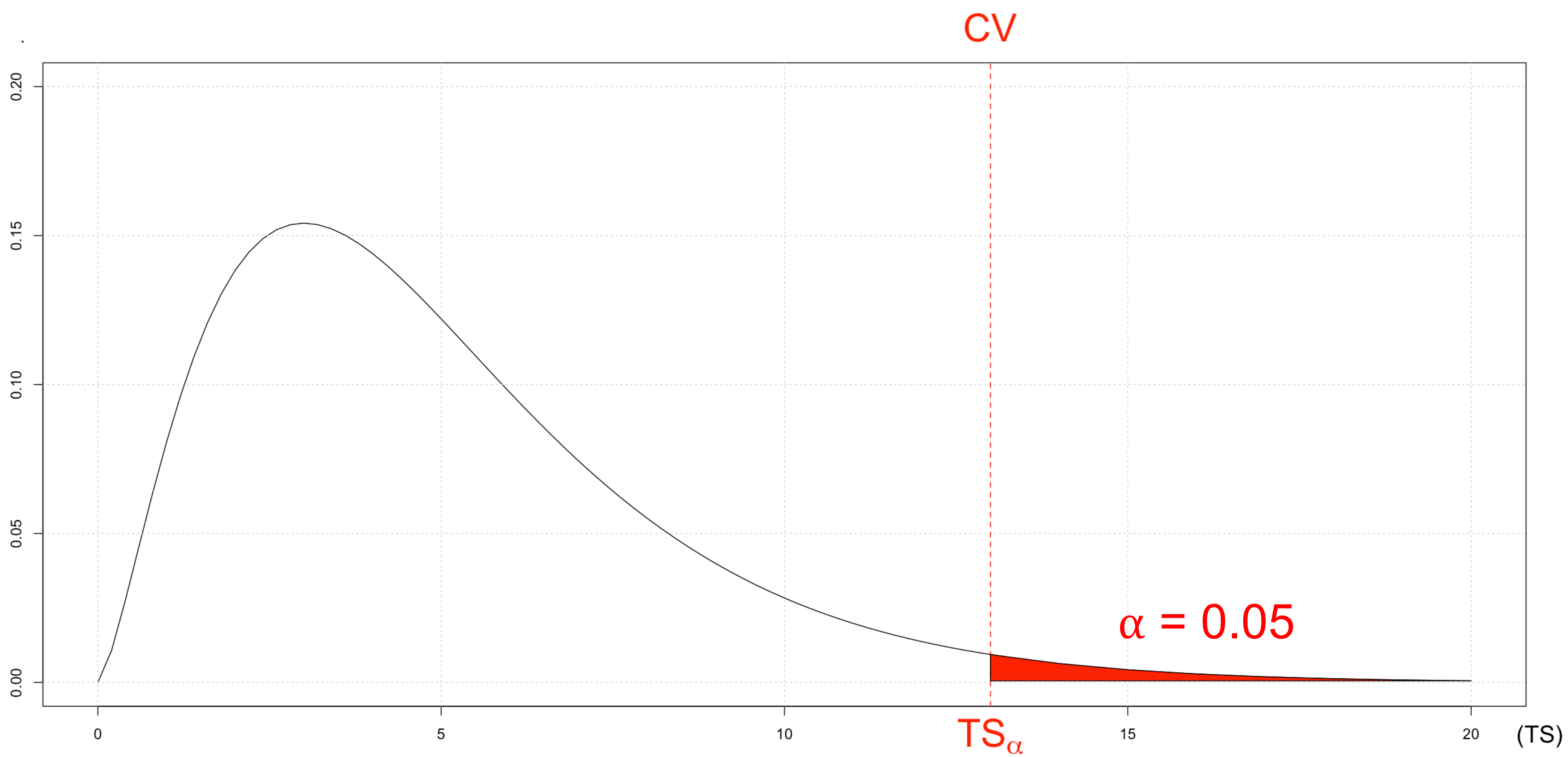
## Analyzing the data

- Select the TS
- Use $\alpha$ to find the critical value (quantile)
- Find the OTS
- Find the p-value

# Consider a TH with a TS chi-square with 6 dof

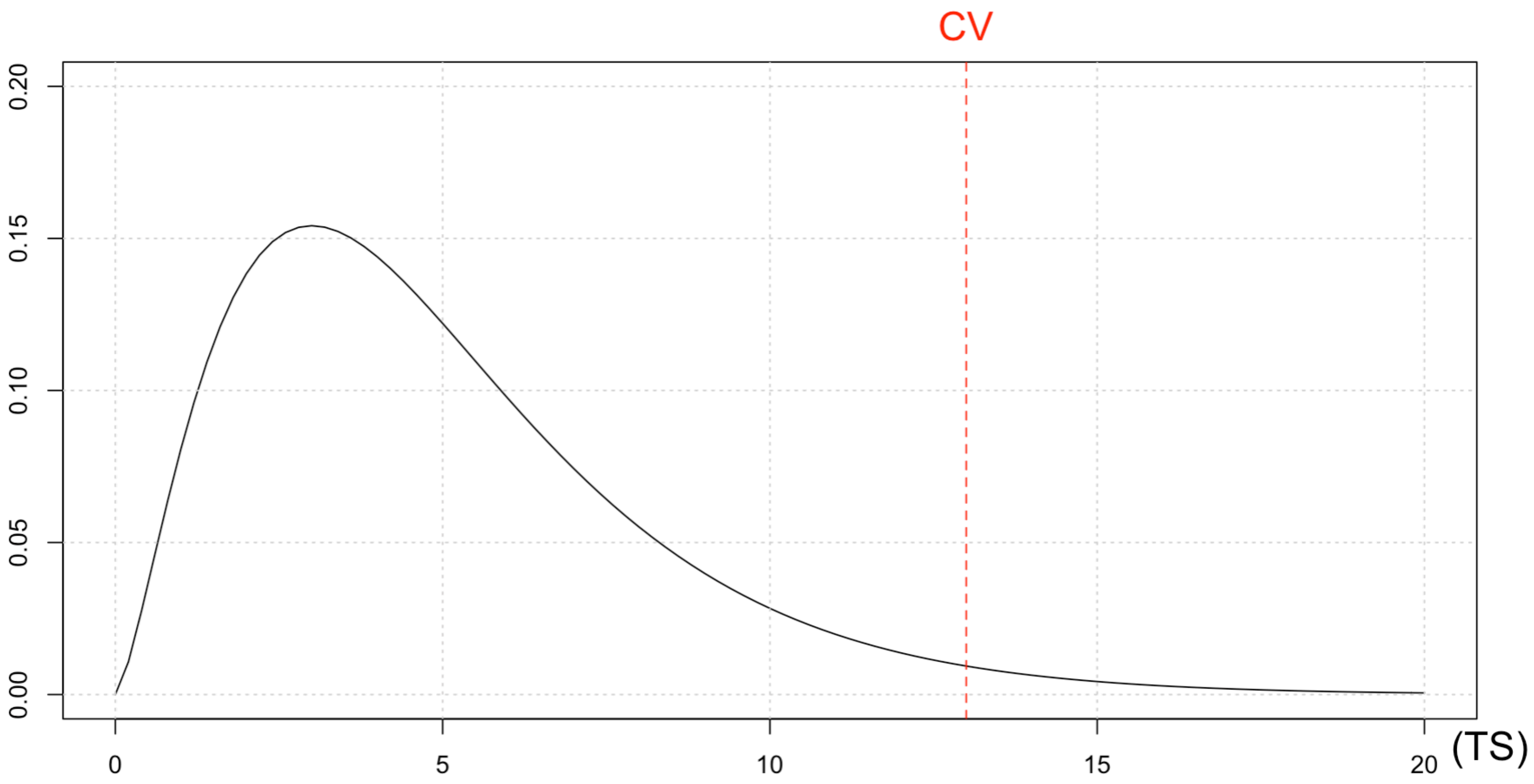
- Select the TS (chi-square with 6 dof)
- Use $\alpha$ to find the critical value (quantile)
- Find the OTS
- Find the p-value

# Introduction




CV = qchisq(0.95,6) = 12.6

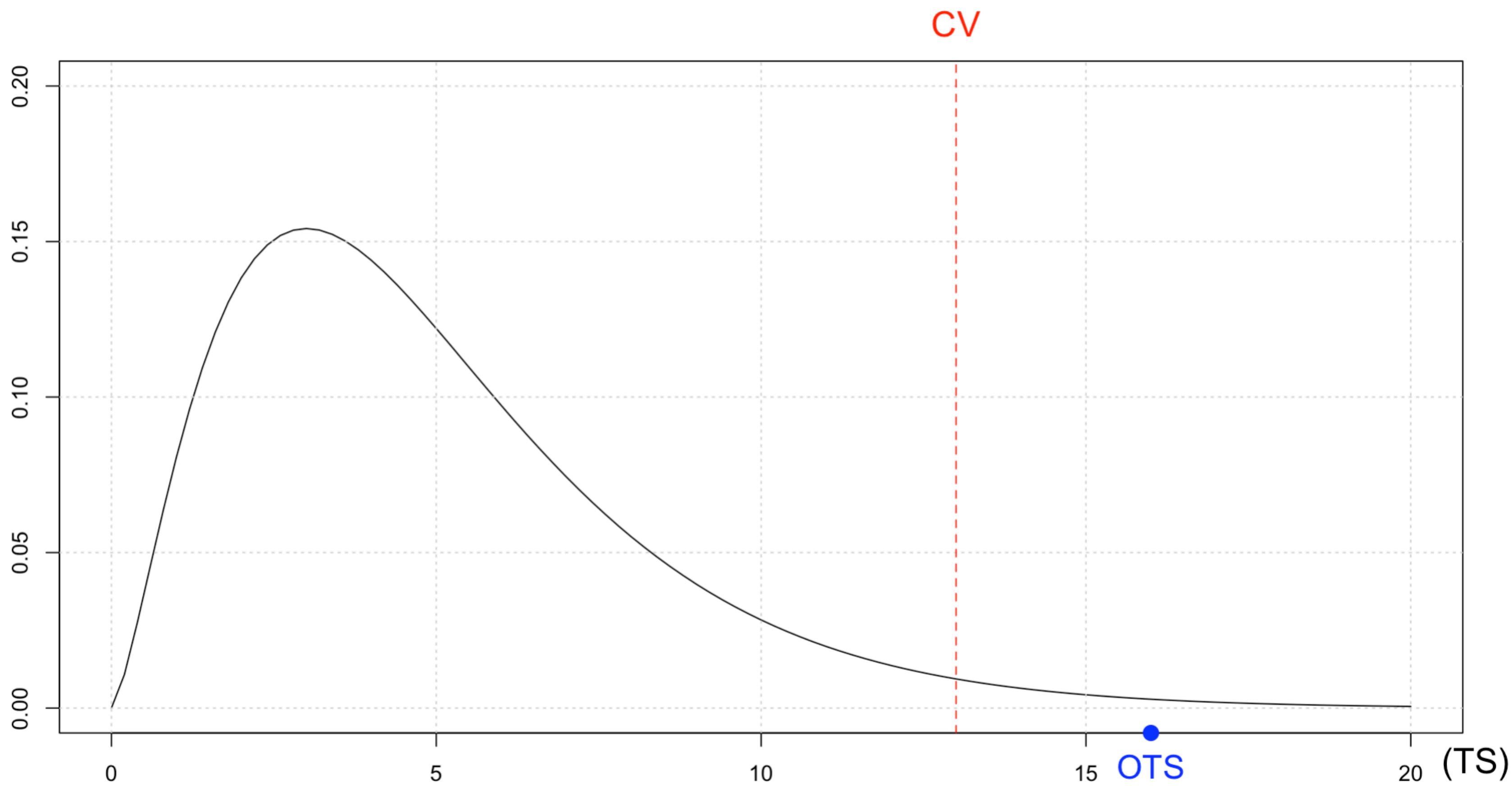
# Introduction




Acceptance Region

Rejection Region

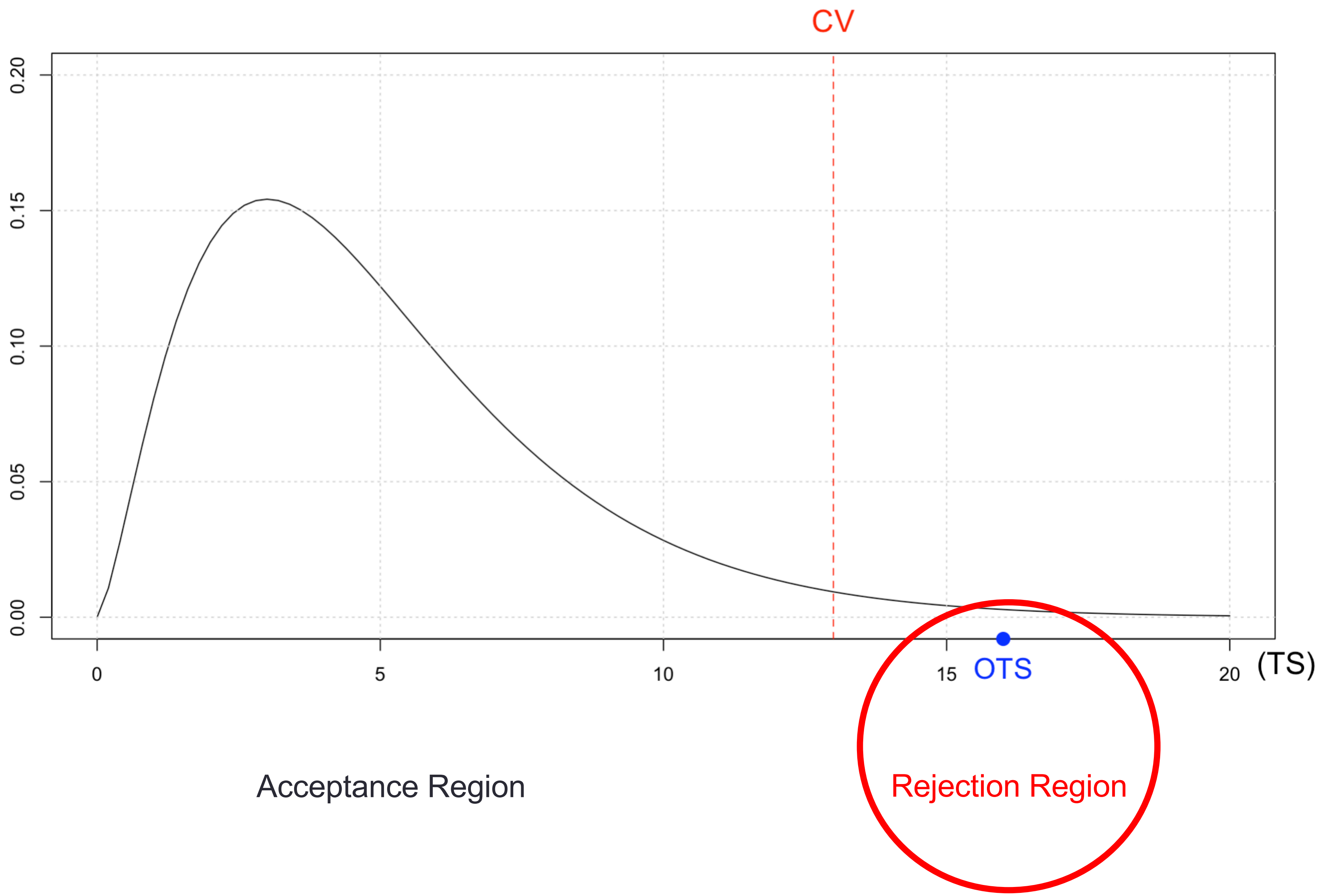
# Introduction



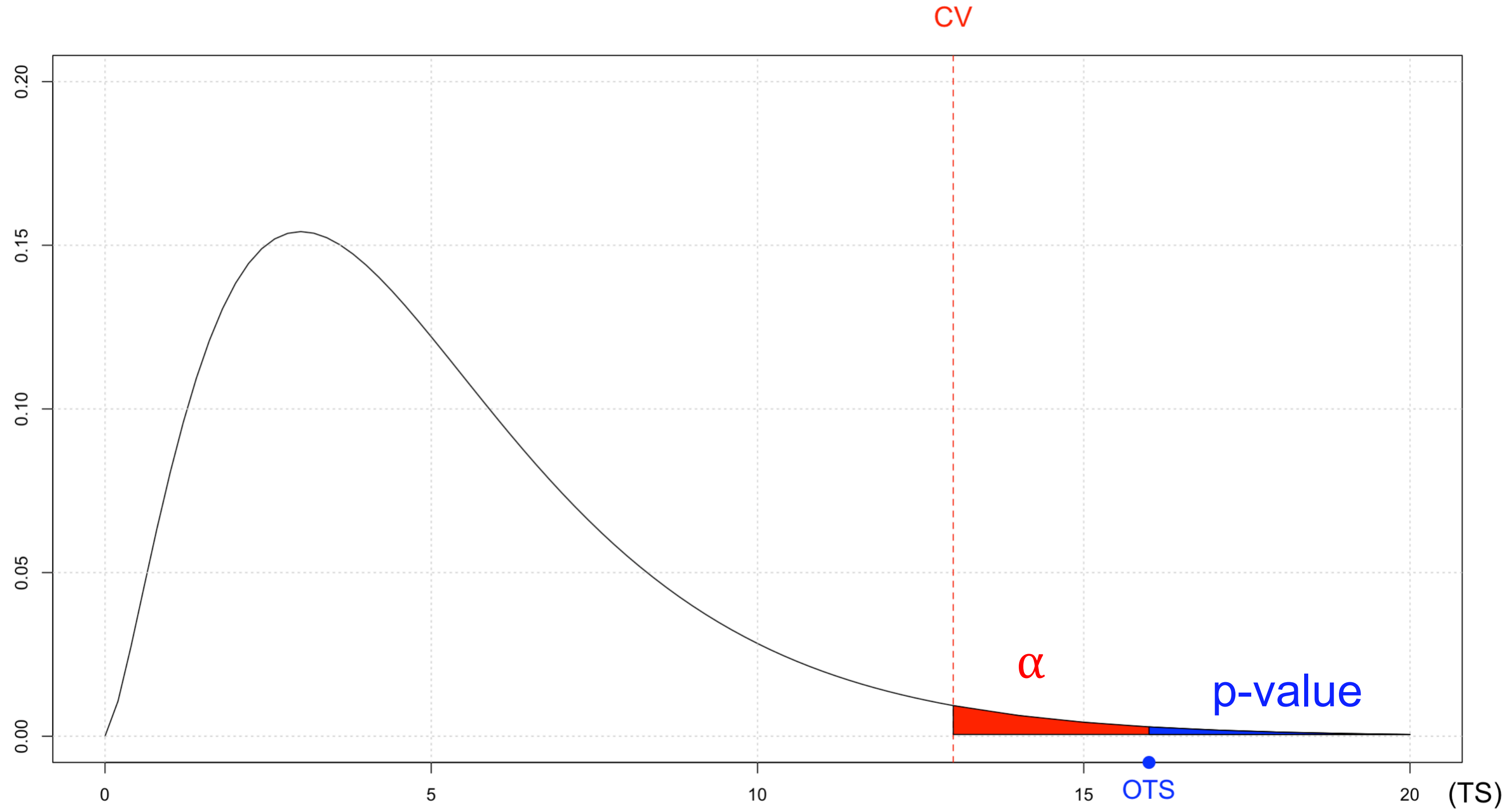

Acceptance Region

Rejection Region

# Introduction



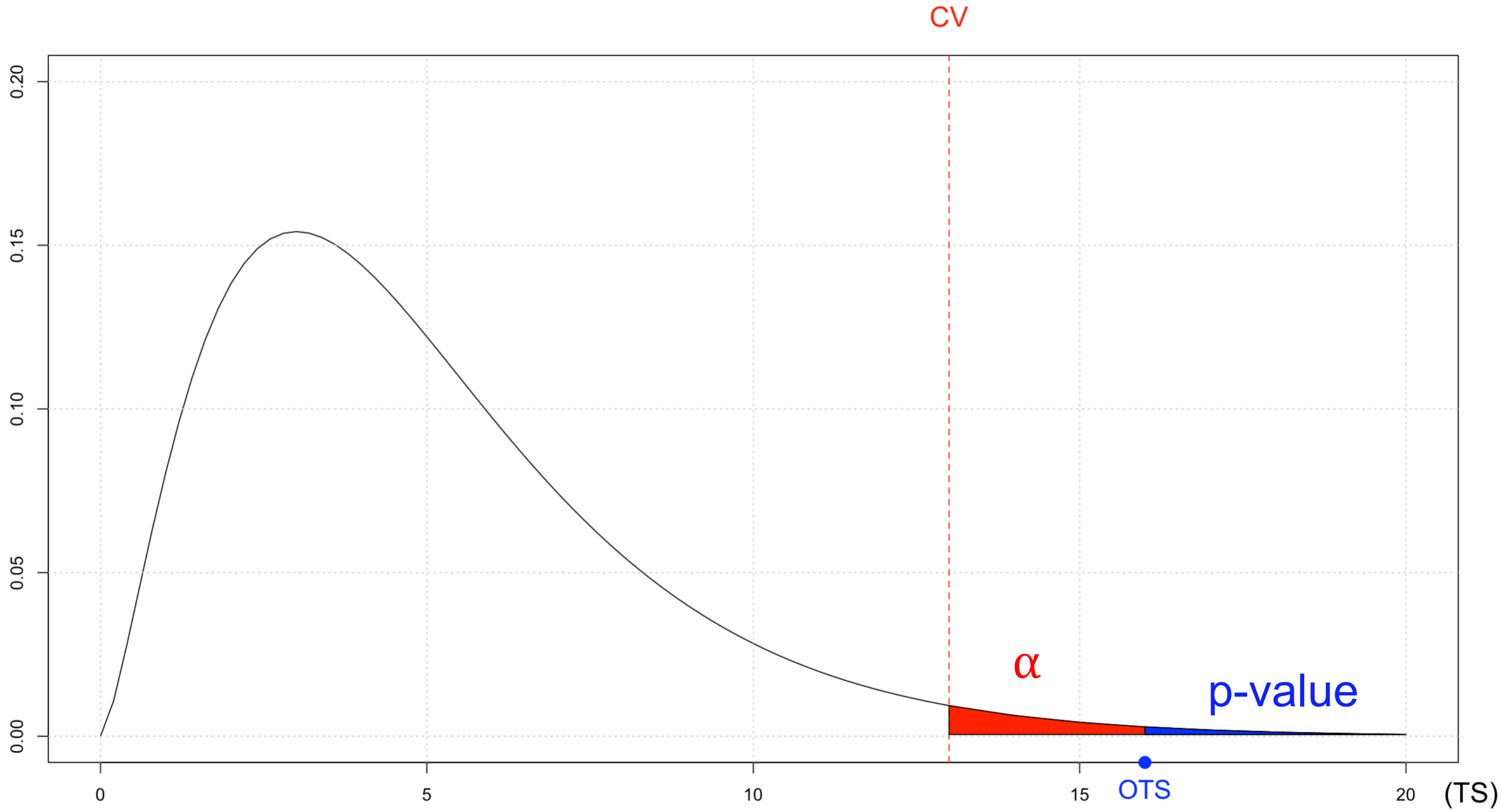
# Introduction




Acceptance Region

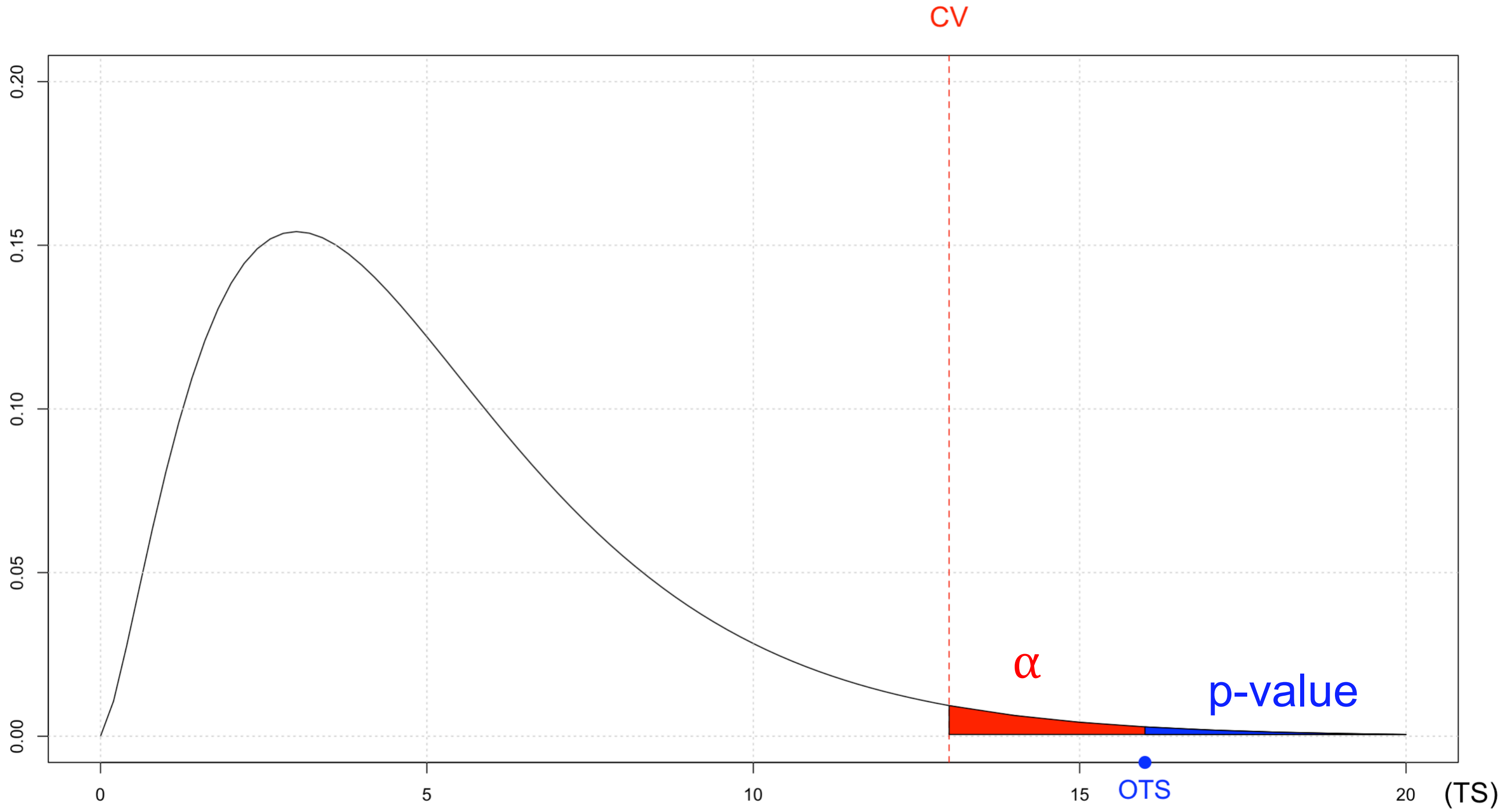
Rejection Region

# Introduction




If p-value < α, reject Ho

# Introduction




If p-value < α,  OTS must be in the RR

# Elements of Statistics

- Binomial and Bernoulli variables
- Normal random variable
- Central Limit Theorem
- Chi-square Theorem
- Confidence Interval

# Bernoulli random variable

- **Trial** is a random experiment with two possible outcomes (success or failure)
- Bernoulli random variable is defined on a trial

$$X = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$$

# Bernoulli random variable

- **Trial** is a random experiment with two possible outcomes (success or failure)
- Bernoulli random variable is defined on a trial

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & 1-p \end{cases}$$

# Bernoulli random variable

- **Trial** is a random experiment with two possible outcomes (success or failure)
- Bernoulli random variable is defined on a trial

$$X \sim BERN(p) \quad \begin{cases} E[X] & = & p \\ Var[X] = & & p(1-p) \end{cases}$$

# Binomial random variable

- Binomial random variable is defined on a sequence of *n* Bernoulli trials
- *X*: number of successes in *n* trials

$$X \sim BINO(n,p) \quad \begin{cases} E[X] &= np \\ Var[X] &= np(1-p) \end{cases}$$

# Binomial random variable

If $(X_1, X_2, \ldots X_n)$ are independent BERN(p)

then

$$\sum_{i=1}^{n} X \quad = \quad X_1 + X_2 + \ldots X_n \qquad \sim BINO(n, p)$$

# Normal random variable

- If *X* is a normal variable with mean μ and variance $\sigma^2$
- then Z is standard normal if

$$Z = \frac{X - \mu}{\sigma} \qquad \begin{cases} E[Z] & = & 0 \\ Var[Z] & = & 1 \end{cases}$$

# Normal random variables

If *($X_1$, $X_2$, … $X_n$)* are independent normal variables with mean μ and variance $\sigma^2$

$$\sum_{i=1}^{n} X \quad = \quad X_1 + X_2 + \ldots X_n \qquad \sim N[n\mu, n\,\sigma^2]$$

$$\overline{X} \quad = \quad \frac{1}{n}[X_1 + X_2 + \ldots X_n] \sim N[\mu, \frac{\sigma^2}{n}]$$

# Central Limit Theorem

The sum of *many* random variables

is approximately

a normal random variable

# Central Limit Theorem

The sum of *many* (independent) random variables

is approximately

a normal random variable

# Central Limit Theorem

If $(X_1, X_2, \ldots X_n)$ are independent from any distribution with mean μ and variance $\sigma^2$,

and *n* is large,

$$\sum_{i=1}^{n} X \quad = \quad X_1 + X_2 + \ldots X_n \qquad \sim N[n\mu, n\,\sigma^2]$$

$$\overline{X} \quad = \quad \frac{1}{n}[X_1 + X_2 + \ldots X_n] \sim N[\mu, \frac{\sigma^2}{n}]$$

# Test on *p*

Testing one population

# Test on *p* Example

A web designer found a job at a consumer products company and was asked to design a new website. After a month he finished the new design, claiming that at least 50% of the clients visiting the new website would make a purchase.

During the following 2 weeks the number of visitors is 1100 with 500 purchases registered.

Is the web designer claim true?

# Test on *p*

- Let $X$ be the number of successes in $n$ trials

$$X \sim BINO(n,p) \quad \begin{cases} E[X] &= np \\ Var[X] &= np(1-p) \end{cases}$$

# Test on $p$

- Let $X$ be the number of successes in $n$ trials

$$X \sim BINO(n,p) \quad \begin{cases} E[X] &= np \\ Var[X] &= np(1-p) \end{cases}$$

$$\hat{p} = \frac{X}{n} \qquad \begin{cases} E[\hat{p}] &= p \\ Var[\hat{p}] &= \frac{p(1-p)}{n} \end{cases}$$

and if $n$ is large

$$\hat{p} \sim N\left[p, \frac{p(1-p)}{n}\right]$$

# Test on *p*

- Let $X$ be the number of successes in $n$ trials

$$X \sim BINO(n,p)$$

$$\hat{p} = \frac{X}{n} \qquad \begin{cases} E[\hat{p}] &= p \\ Var[\hat{p}] &= \frac{p(1-p)}{n} \end{cases}$$

and if $n$ is large

$$\hat{p} \sim N\left[p, \frac{p(1-p)}{n}\right]$$

To test

$$H_0 : p = p_0$$

$$H_a : p > p_0$$

use

Test Statistic (TS) $\quad Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$

Critical Value (CV) $\quad Z_\alpha$

Obs. Test Statistic (OTS) $\quad z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

p-value $\quad P[Z > z_0]$

# Test on *p*

- 1000 flips of a coin results in 530 heads.
- Test if the coin is fair

# Test on *p*

- 1000 flips of a coin results in 530 heads.
- Test if the coin is fair

$$H_0 : p = 0.50$$

$$H_a : p > 0.50$$

# Test on *p*

$$H_0 : p = 0.50$$

$$H_a : p > 0.50$$

Test Statistic (TS) $\quad Z = \dfrac{\hat{p} - 0.5}{\sqrt{\dfrac{p(1-p)}{n}}}$

Critical Value (CV) $\quad Z_{0.05} = 1.645$

Observed fraction $\quad \hat{p} = \frac{x}{n} = \frac{530}{1000} = 0.53$

Obs. Test Statistic (OTS)

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.53 - 0.50}{\sqrt{\dfrac{0.5(1-0.50)}{1000}}} = 1.897$$

p-value $\quad P[Z > 1.897] = 0.029$

# Test on *p*

$H_0 : p = 0.50$

$H_a : p > 0.50$

Test Statistic (TS) $$Z = \frac{\hat{p} - 0.5}{\sqrt{\frac{p(1-p)}{n}}}$$

Critical Value (CV) $$Z_{0.05} = 1.645$$

Observed fraction $$\hat{p} = \frac{x}{n} = \frac{530}{1000} = 0.53$$

Obs. Test Statistic (OTS) $$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.53 - 0.50}{\sqrt{\frac{0.5(1-0.50)}{1000}}} = 1.897$$

p-value $$P[Z > 1.897] = 0.029$$

Reject Ho

# Test on *p*

```
> 1-pnorm(1.897)
[1] 0.02891397
>
>
> binom.test(530,1000,0.50,'greater')

        Exact binomial test

data:  530 and 1000
number of successes = 530, number of trials = 1000, p-value = 0.03101
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.5034989 1.0000000
sample estimates:
probability of success
                  0.53
```

# Test on *p*

```
> 1-pnorm(1.897)
[1] 0.02891397
>
>
> binom.test(530,1000,0.50,'greater')

        Exact binomial test

data:  530 and 1000
number of successes = 530, number of trials = 1000, p-value = 0.03101
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.5034989 1.0000000
sample estimates:
probability of success
                  0.53
```

# Test on *p*

```
> 1-pnorm(1.897)
[1] 0.02891397
>
>
> binom.test(530,1000,0.50,'greater')

        Exact binomial test

data:  530 and 1000
number of successes = 530, number of trials = 1000, p-value = 0.03101
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.5034989 1.0000000
sample estimates:
probability of success
                  0.53
```

Reject Ho since p-value = 0.0289 < 0.05, coin is not fair

# Confidence Interval (CI)

A CI is a random interval that may include the parameter of interest with probability $1-\alpha$

# Confidence Interval (CI)

A random interval example

$P[3 < X < 5] \quad =$

# Confidence Interval (CI)

## A random interval example

$$
\begin{aligned}
P[3 < X < 5] &= P[0 < X - 3 < 2] \\
&= P[0 > 3 - X > -2] \\
&= P[X > 3 > X - 2] \\
&= P[X - 2 < 3 < X]
\end{aligned}
$$

# Confidence Interval (CI)

A random interval that may include the parameter of interest with probability 1-$\alpha$

What is $\alpha$ ?

$\alpha$ is the probability that the CI does not include the parameter of interest

$\alpha$ is the probability of an incorrect conclusion

# CI on $p$

for any random variable $$1 - \alpha = P[-x_{\alpha/2} < X < x_{\alpha/2}]$$

# CI on *p*

for any random variable $$1-\alpha = P[-x_{\alpha/2} < X < x_{\alpha/2}]$$

let us use it for Z $$1-\alpha = P[-z_{\alpha/2} < Z < z_{\alpha/2}]$$

# CI on *p*

for any random variable $$1-\alpha = P[-x_{\alpha/2} < X < x_{\alpha/2}]$$

let us use it for Z $$1-\alpha = P[-z_{\alpha/2} < Z < z_{\alpha/2}]$$

CI is given by this expression

where $$Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

# CI on *p*

for any random variable $$1 - \alpha = P[-x_{\alpha/2} < X < x_{\alpha/2}]$$

let us use it for Z $$1 - \alpha = P[-z_{\alpha/2} < Z < z_{\alpha/2}]$$

CI is given by this expression

instead use TS $$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

# CI on *p*

$$1-\alpha = P\left[-z_{\alpha/2} < \frac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}} < z_{\alpha/2}\right]$$

and solve the inequality for *p*

# CI on *p*

$$1 - \alpha = P\left[ -z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < z_{\alpha/2} \right]$$

and solve the inequality for *p*

# CI on *p*

$$1-\alpha = P\left[-z_{\alpha/2} < \frac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}} < z_{\alpha/2}\right]$$

$$1-\alpha = P\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

# CI on *p* - Example

Use Monte Carlo simulation

to understand the meaning of

the confidence coefficient $\alpha$

in a Confidence Interval

# CI on *p* - Example

- Consider that $p$ = 30% of customers that visit a new website make a purchase.
- Suppose that $p$ is unknown (the website is new)
- To estimate $p$, the website is monitored over 50 days
- Each day the first 50 visits are recorded and a CI on $p$ is constructed
- Use Monte Carlo simulation to find how many CIs cover $p$

# CI on *p* - Example

```r
p = 0.3
N = 1000
population = rbinom(N,1,p)
table(population)
```

```
## population
##    0   1
## 696 304
```

# CI on *p* - Example

```
p = 0.3
N = 1000
population = rbinom(N,1,p)
table(population)
```

rbern(N,p)

```
## population
##   0   1
## 696 304
```

# CI on *p* - Example

- Select 50 rows from vector `population`
- Find `phat` and the std deviation of `phat`
- Construct the CI on `p`

# CI on *p* - Example

```r
# collect sample
n = 50
id = sample(1:N,n)
obs = population[id]
```

```r
table(obs)
```

```
## obs
##  0  1
## 30 20
```

```r
ob_success = table(obs)[2]
phat = ob_success/n
phat
```

```
##   1
## 0.4
```

```r
sdev = sqrt(phat*(1-phat)/n)
sdev
```

```
##          1
## 0.06928203
```

# CI on *p* - Example

```
alpha = 0.05
z_alpha = qnorm(1-alpha/2)
z_alpha
```

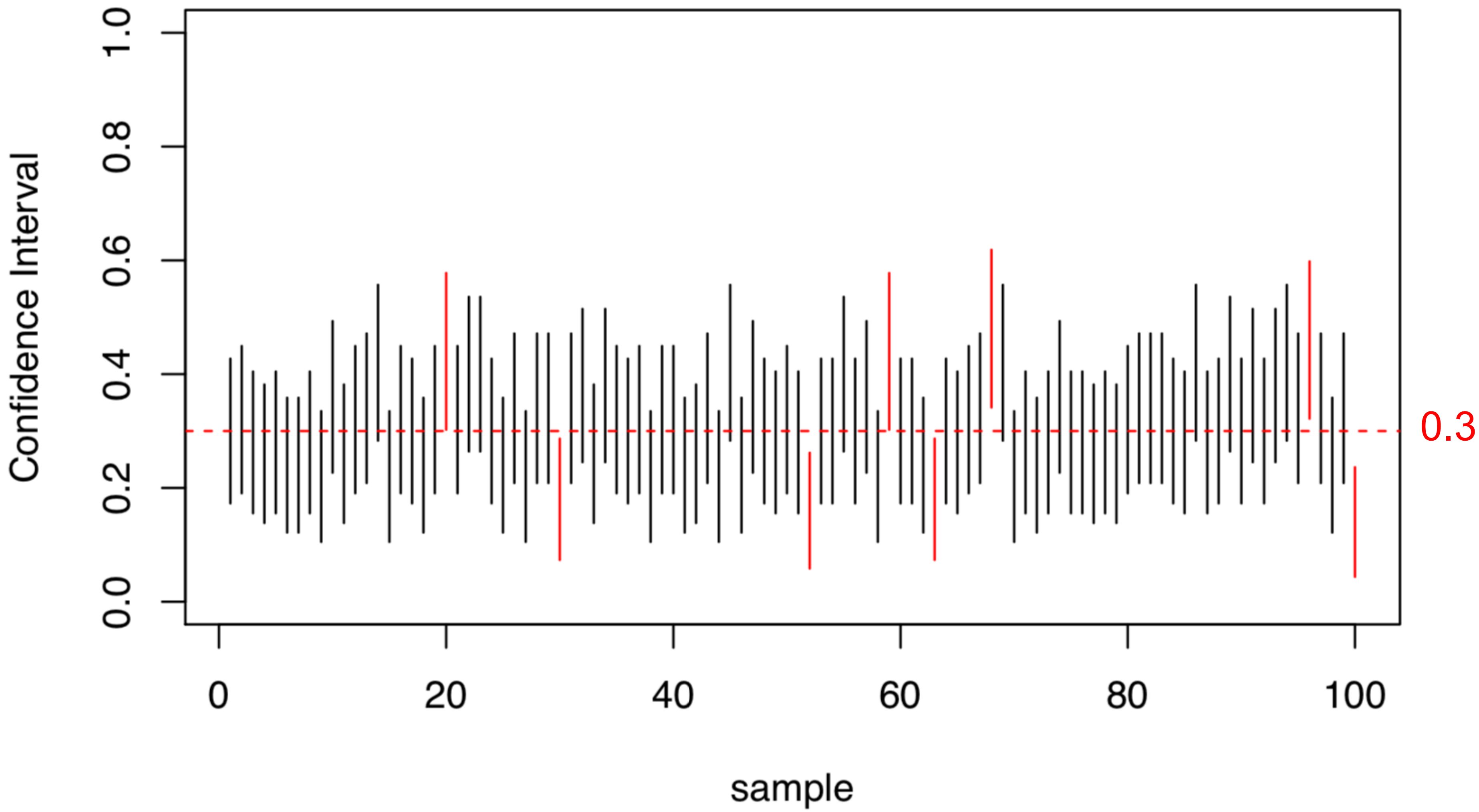
```
# find Confidence interval
lb = phat - z_alpha*sdev
ub = phat + z_alpha*sdev
c(lb,ub)
```

```
##         1         1
## 0.2642097 0.5357903
```

# CI on *p* - Example

- Select 50 rows from vector population
- Find phat and the std deviation of phat
- Construct the CI on p
- Repeat 100 times
- Plot all CIs

# CI on *p* - Example



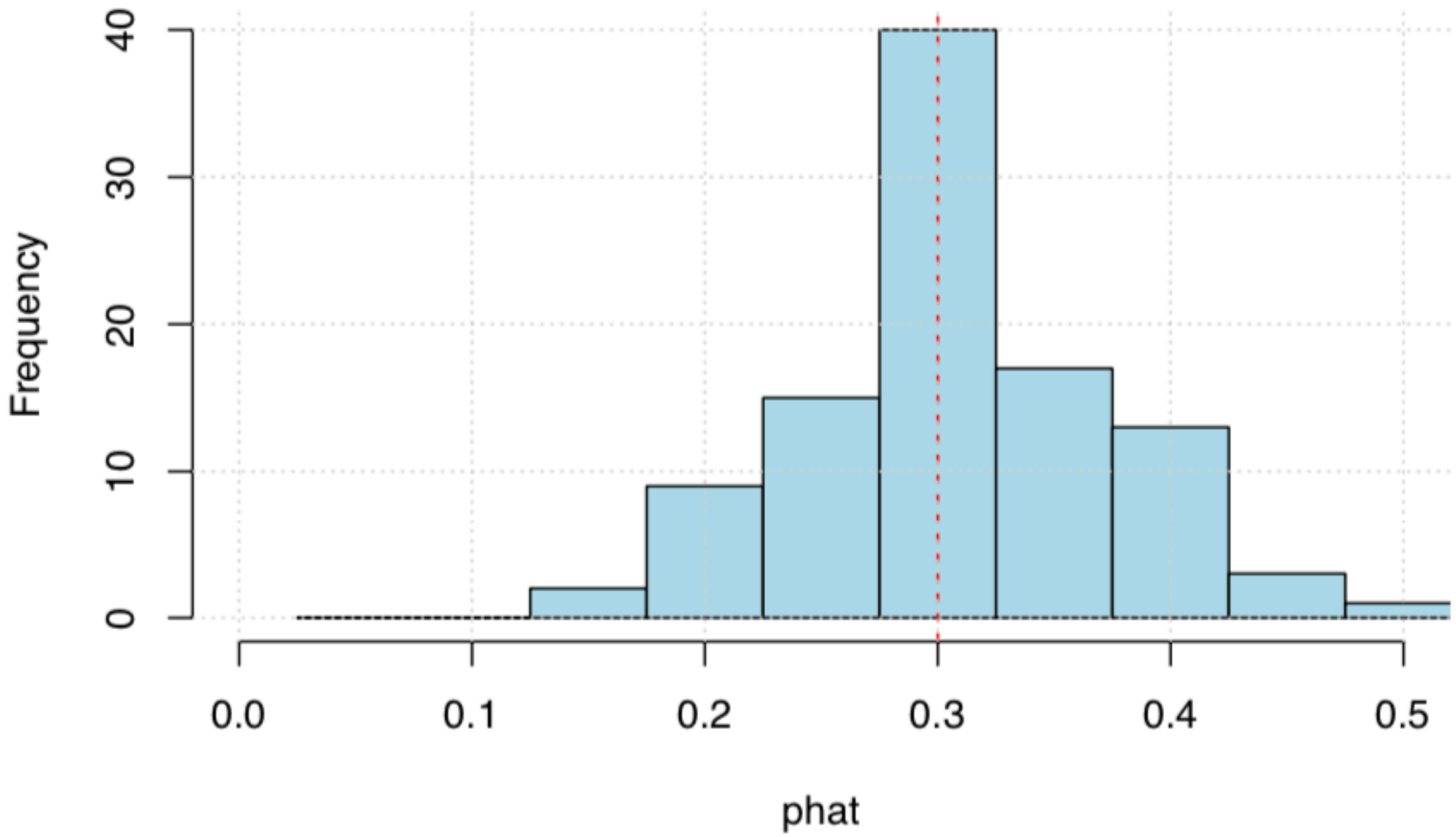
# CI on *p* - Example

- This experiment shows 8 intervals that do not include p = 0.30
- In practice, we just observe one and hope that it covers p
- We are 95% confident that, that CI, includes p = 0.30

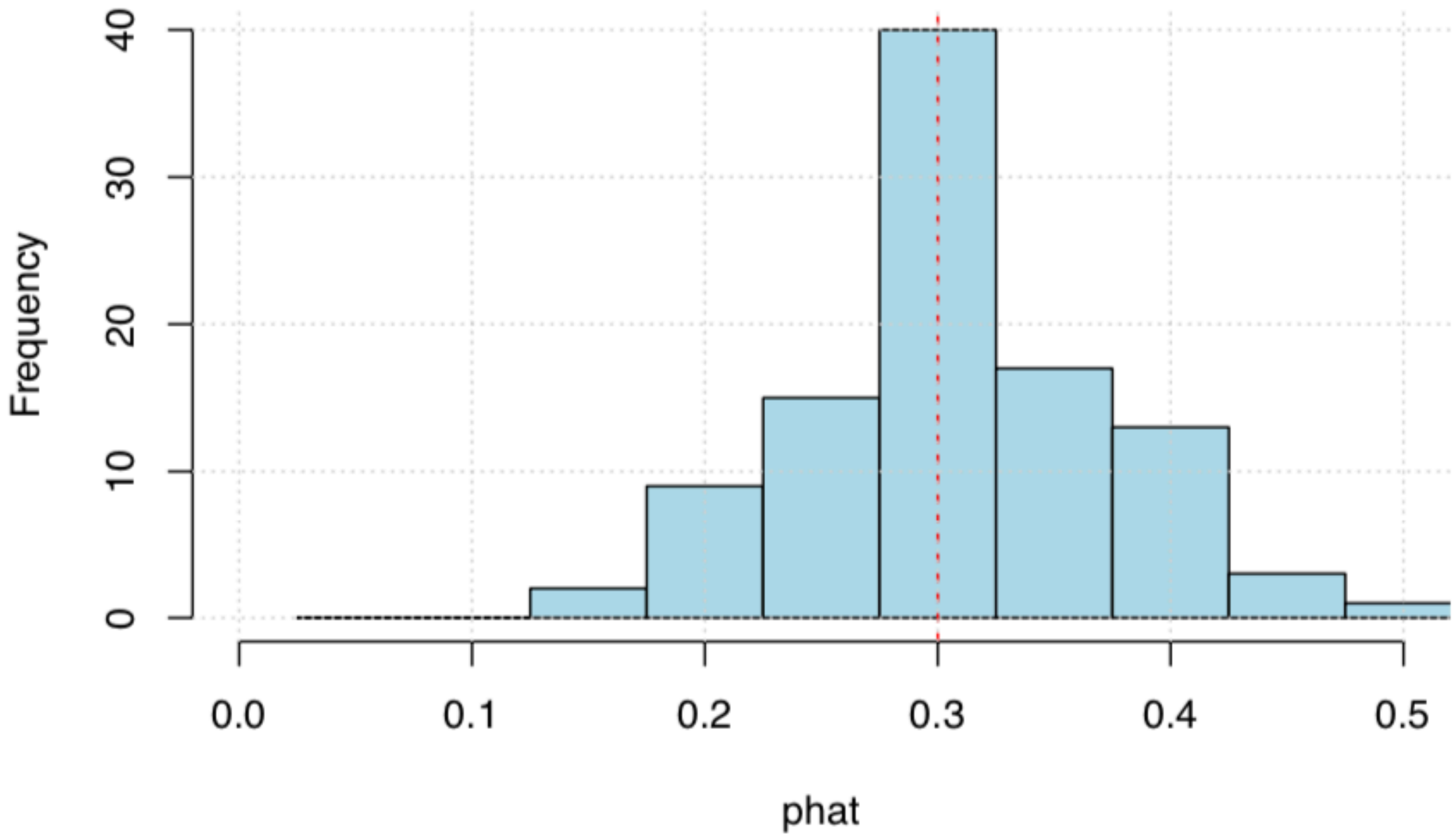
# CI on *p* - Example

Construct a histogram of the simulated fractions of customers that make a purchase

# CI on *p* - Example



# CI on *p* - Example




On average we expect that the fraction of purchases is equal to 0.3

# CI vs TH on *p*

- If $H_0 : p = 0.50$ is not rejected

  then the CI contains 0.50

- If the CI does not contain 0.50

  then $H_0 : p = 0.50$ is rejected

# CI vs TH on *p*

- If $H_0 : p = 0.50$ is not rejected

  then the CI contains 0.50

- If the CI does not contain 0.50

  then $H_0 : p = 0.50$ is rejected

- This relation applies to *all* CI and TH on the same dataset

# Comparing two populations

Which one is preferable?

# Testing $p_1 - p_2$ Example

Compare two headlines A and B

| | A | B |
|---|---|---|
| Click | 405 | 380 |
| No click | 495 | 570 |
| | 900 | 950 |

Does headline A have a higher rate over headline B

# Testing $p_1 - p_2$ from two populations

$X_1$ the number of successes in $n_1$ trials from population 1 $\quad X_1 \sim BINO(n_1, p_1)$

$X_2$ the number of successes in $n_2$ trials from population 2 $\quad X_2 \sim BINO(n_2, p_2)$

# Testing $p_1 - p_2$ from two populations

$X_1$ the number of successes in $n_1$ trials from population 1 $\quad X_1 \sim BINO(n_1, p_1)$

$X_2$ the number of successes in $n_2$ trials from population 2 $\quad X_2 \sim BINO(n_2, p_2)$

$$\hat{p}_1 = \frac{X_1}{n_1} \qquad \hat{p}_1 \sim N\left[p_1, \frac{p_1(1-p_1)}{n_1}\right]$$

$$\hat{p}_2 = \frac{X_2}{n_2} \qquad \hat{p}_2 \sim N\left[p_2, \frac{p_2(1-p_2)}{n_2}\right] \qquad \hat{p}_1 - \hat{p}_2 \sim N\left[p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right]$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

# Testing $p_1 - p_2$ from two populations

To test

$$H_0 : p_1 = p_2$$

$$H_a : p_1 > p_2$$

or

$$H_0 : p_1 - p_2 = 0$$

$$H_a : p_1 - p_2 > 0$$

use

| | |
|---|---|
| Test Statistic (TS) | $Z = \dfrac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$ |
| Critical Value (CV) | $Z_\alpha$ |
| Obs. Test Statistic (OTS) | $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$ |
| p-value | P[Z> $z_0$] |

# Testing $p_1 - p_2$ from two populations

To test

$$H_0 : p_1 = p_2$$

$$H_a : p_1 > p_2$$

or

$$H_0 : p_1 - p_2 = 0$$

$$H_a : p_1 - p_2 > 0$$

use

Test Statistic (TS) $\quad Z = \dfrac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$

Critical Value (CV) $\quad Z_\alpha$

Obs. Test Statistic (OTS) $\quad z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$

p-value $\quad$ P[Z> $z_0$]

# Testing $p_1 - p_2$ from two populations

.

To test

$$H_0 : p_1 = p_2$$

$$H_a : p_1 > p_2$$

or

$$H_0 : p_1 - p_2 = 0$$

$$H_a : p_1 - p_2 > 0$$

use

Test Statistic (TS) $$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$$

Critical Value (CV) $Z_\alpha$

Obs. Test Statistic (OTS) $$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}_p(1-\hat{p}_p)}{n_1} + \dfrac{\hat{p}_p(1-\hat{p}_p)}{n_2}}}$$

p-value P[Z> $z_0$]

# Testing $p_1 - p_2$ from two populations

use the *pooled* fraction of successes

$$
\begin{aligned}
\hat{p}_p &= \frac{x_1 + x_2}{n_1 + n_2} \\
&= \frac{x_1}{n_1 + n_2} + \frac{x_2}{n_1 + n_2} \\
&= \frac{n_1}{n_1 + n_2} \left(\frac{x_1}{n_1}\right) + \frac{n_2}{n_1 + n_2} \left(\frac{x_2}{n_2}\right) \\
&= \frac{n_1}{n_1 + n_2}\ \hat{p}_1 + \frac{n_2}{n_1 + n_2}\ \hat{p}_2
\end{aligned}
$$

# Testing $p_1 - p_2$ Example

Compare two headlines A and B

| | A | B |
|---|---|---|
| Click | 405 | 380 |
| No click | 495 | 570 |
| | 900 | 950 |

Does headline A have a higher rate over headline B

# Testing $p_1 - p_2$ Example

Compare two headlines A and B

| | A | B |
|---|---|---|
| Click | 405 | 380 |
| No click | 495 | 570 |
| | 900 | 950 |

Does headline A have a higher rate over headline B

# Testing $p_1 - p_2$ Example

$$H_0 : p_A = p_B \quad n_A = 900 \quad \hat{p}_A = 0.45$$

$$H_a : p_A > p_B \quad n_B = 950 \quad \hat{p}_B = 0.40 \quad Z_\alpha = 1.645$$

*pooled* fraction of successes

| | A | B |
|---|---|---|
| Click | 405 | 380 |
| No click | 495 | 570 |
| | 900 | 950 |

$$\hat{p}_p = \frac{405 + 380}{900 + 950} = 0.42432$$

# Testing $p_1 - p_2$ Example

the observed test statistic

$$\begin{aligned} z_0 &= \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\frac{\hat{p}_p(1-\hat{p}_p)}{n_1} + \frac{\hat{p}_p(1-\hat{p}_p)}{n_2}}} \\ &= \frac{0.45 - 0.40}{\sqrt{0.24427\left(\frac{1}{900} + \frac{1}{950}\right)}} \\ &= 2.17486 \end{aligned}$$

# Testing $p_1 - p_2$ Example

the observed test statistic

$$\begin{aligned} z_0 &= \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\frac{\hat{p}_p(1-\hat{p}_p)}{n_1} + \frac{\hat{p}_p(1-\hat{p}_p)}{n_2}}} \\ &= \frac{0.45 - 0.40}{\sqrt{0.24427\left(\frac{1}{900} + \frac{1}{950}\right)}} \\ &= 2.17486 \end{aligned}$$

$$\begin{aligned} \text{p-value} &= P[Z > 2.17486] \\ &= 1 - \texttt{pnorm(2.17486)} \\ &= 0.01482 \end{aligned}$$

# Testing $p_1 - p_2$ Example

the observed test statistic

$$
\begin{aligned}
z_0 &= \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\frac{\hat{p}_p(1-\hat{p}_p)}{n_1} + \frac{\hat{p}_p(1-\hat{p}_p)}{n_2}}} \\
&= \frac{0.45 - 0.40}{\sqrt{0.24427\left(\frac{1}{900} + \frac{1}{950}\right)}} \\
&= 2.17486
\end{aligned}
$$

$$
\begin{aligned}
\text{p-value} &= P[Z > 2.17486] \\
&= 1 - \texttt{pnorm(2.17486)} \\
&= 0.01482
\end{aligned}
$$

# Testing $p_1 - p_2$ Example

Compare two headlines A and B

| | A | B |
|---|---|---|
| Click | 405 | 380 |
| No click | 495 | 570 |
| | 900 | 950 |

Does headline A have a higher rate over headline B

Yes, it does

# CI on $p_1 - p_2$

CI is given by this expression $1-\alpha = P[-z_{\alpha/2} < Z < z_{\alpha/2}]$

where
$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$$

instead use TS
$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

to get
$$1-\alpha = P\left[-z_{\alpha/2} < \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} < z_{\alpha/2}\right]$$

and solve the inequality for $p_1 - p_2$

# CI on $p_1 - p_2$

$$1-\alpha = P\left[\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} < p_1 - p_2 < \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right]$$

or simply put

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

# CI on $p_A - p_B$ Example

Compare two headlines A and B

| | A | B |
|---|---|---|
| Click | 405 | 380 |
| No click | 495 | 570 |
| | 900 | 950 |

Construct a CI on the difference of click rates

between headline A and headline B

# CI on $p_A - p_B$ Example

$$\begin{aligned}
&\hat{p}_A - \hat{p}_B \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_1} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_2}} \\
&= 0.05 \pm 1.645\sqrt{\frac{0.45\,(1-0.45)}{900} + \frac{0.40\,(1-0.40)}{950}} \\
&= 0.05 \pm 0.0378 \\
&= (0.0122, 0.0878)
\end{aligned}$$

Since this interval does not include 0, we conclude that the click-rate of headline A is larger than that of headline B.

More specific, the click rate of headline A is larger than that of headline B by somewhere between 1.22% and 8.78%.