A/B Testing

Introduction

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

Introduction

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

Extended A/B test can compare more than two populations

Introduction

A/B testing in data science is typically in a web context

Introduction

A/B testing in data science is typically in a web context

Measures of interest are the number of visits, purchases, number of clicks, etc.

Introduction

A single metric must be chosen to decide between two populations

That single metric is a test statistic

Introduction

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

Extended A/B test to compare two populations

What is a test of hypothesis?

Introduction

Hypothesis

a conjecture about

one or more populations

Introduction

Hypothesis

a conjecture about the characteristics of

one or more populations

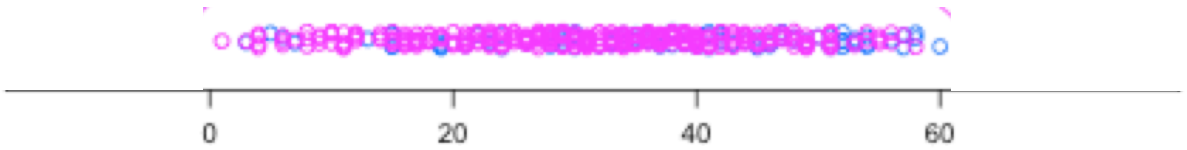
Introduction

Hypothesis

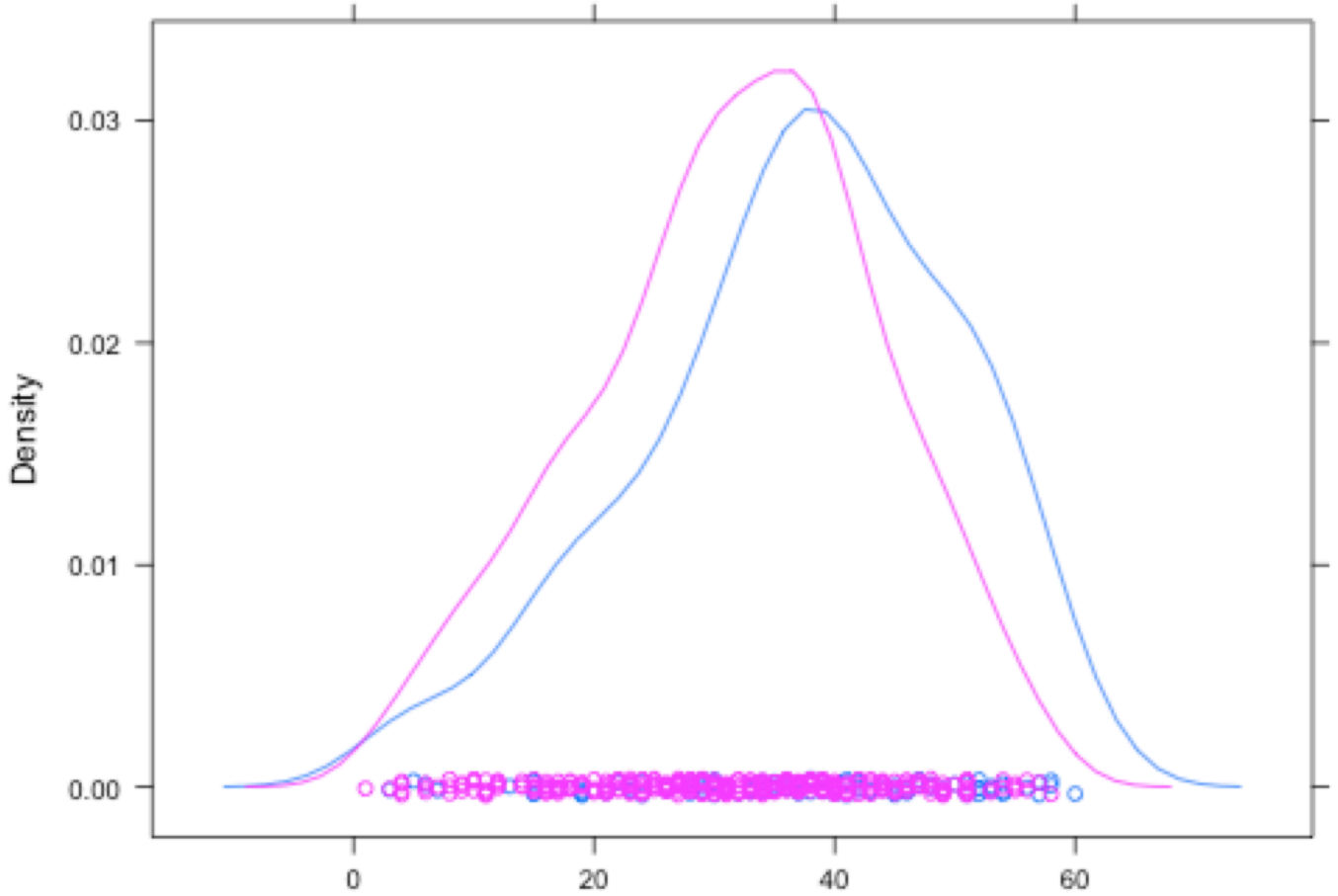
a conjecture about the parameters of

one or more populations

Introduction



Introduction



Conjectures

What can be measured from website visits?

- number of daily visits
- number of transactions
- visit duration
- most used links (from the website)

Conjectures

What can be measured from website visits?

- number of daily visits
- number of transactions
- visit duration
- most used links (from the website)

these are random variables

Conjectures about one website

- average number of daily visits
- average number of transactions
- average visit duration
- fraction of visits that result in a transaction
- fraction of visits that select a new link

Conjectures about two websites

- difference in the average number of daily visits
- difference in the average number of transactions
- difference in the average duration of visits
- difference in the fraction of visits that result in a transaction
- difference in the fraction of visits that select a new link

Conjectures about two websites

- If two populations are concluded to be different, then one should be preferable (better than the other)

Example – Time spent in website designs

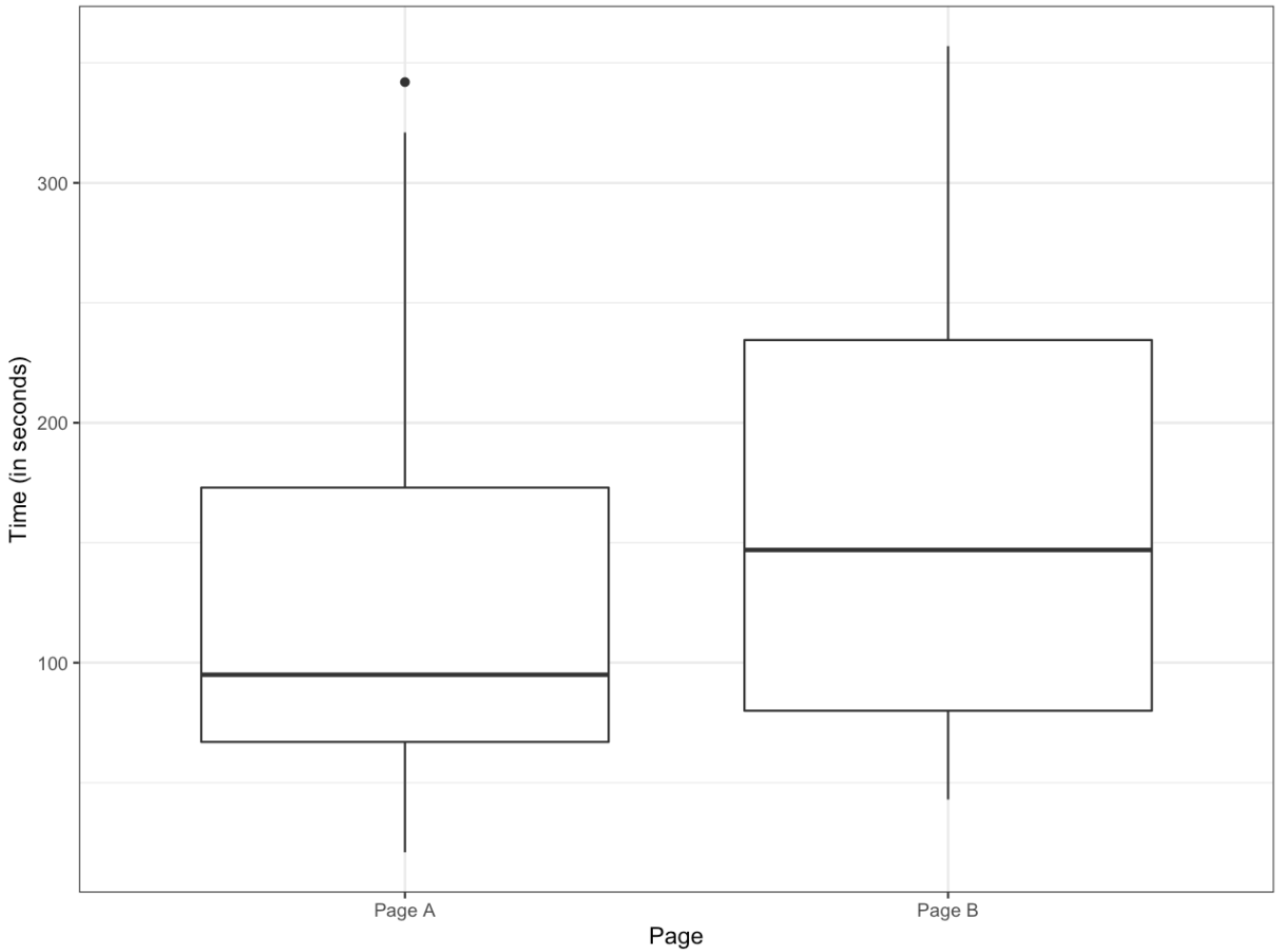
- Consider two webpages A and B
- The session times from randomly chosen users has been recorded (in seconds)
- Which webpage do the users spend more time?

Example – Time spent in website designs

```
> head(times1)
     Page Time
1 Page A   21
2 Page B  253
3 Page A   35
4 Page B   71
5 Page A   67
6 Page B   85
> tapply(times1$Time,times1$Page,mean)
  Page A   Page B
126.3333 162.0000
```

It seems that Page B is more attractive to the users

Example – Mean time spent in website designs



Example – Mean time spent in website designs

- The observed difference in average time spent by randomly chosen users is

```
> obs_diff = mean_b - mean_a
> obs_diff
[1] 35.66667
```

- This observed difference is random
- If the true difference is greater than zero, then we conclude that webpage B is more attractive

Example – Time spent in website designs

- What is the TRUE difference?

Example – Time spent in website designs

- What is the TRUE difference?
- If all possible sets of randomly chosen users are collected and we record for each webpage
    - time spent by them
    - average time spent
    - the difference in the average time spent is positive
- then we conclude that one page is preferred

Example – Time spent in website designs

What is the TRUE difference?

Unfortunately
finding the TRUE difference
is not possible

Example – Time spent in website designs

- Consider file web_page_data.csv

- It contains the time spent by 36 users
- From this set of users,
    - 21 visited website A, and
    - 15 others visited website
- We recorded the time spent in each website

Example – Time spent in website designs

Permutation Test

- Combine all 36 observed times in a single set
- Randomly choose 21 and find their average
- Find the average of the remaining 15 times
- Find the difference in the two averages
- Repeat

Example – Time spent in website designs

Permutation Test

- Combine all 36 observed times in a single set
- Randomly choose 21 and find their average
- Find the average of the remaining 15 times
- Find the difference in the two averages
- Repeat

```
> head(times1)
     Page Time
1 Page A   21
2 Page B  253
3 Page A   35
4 Page B   71
5 Page A   67
6 Page B   85
```

Example – Time spent in website designs

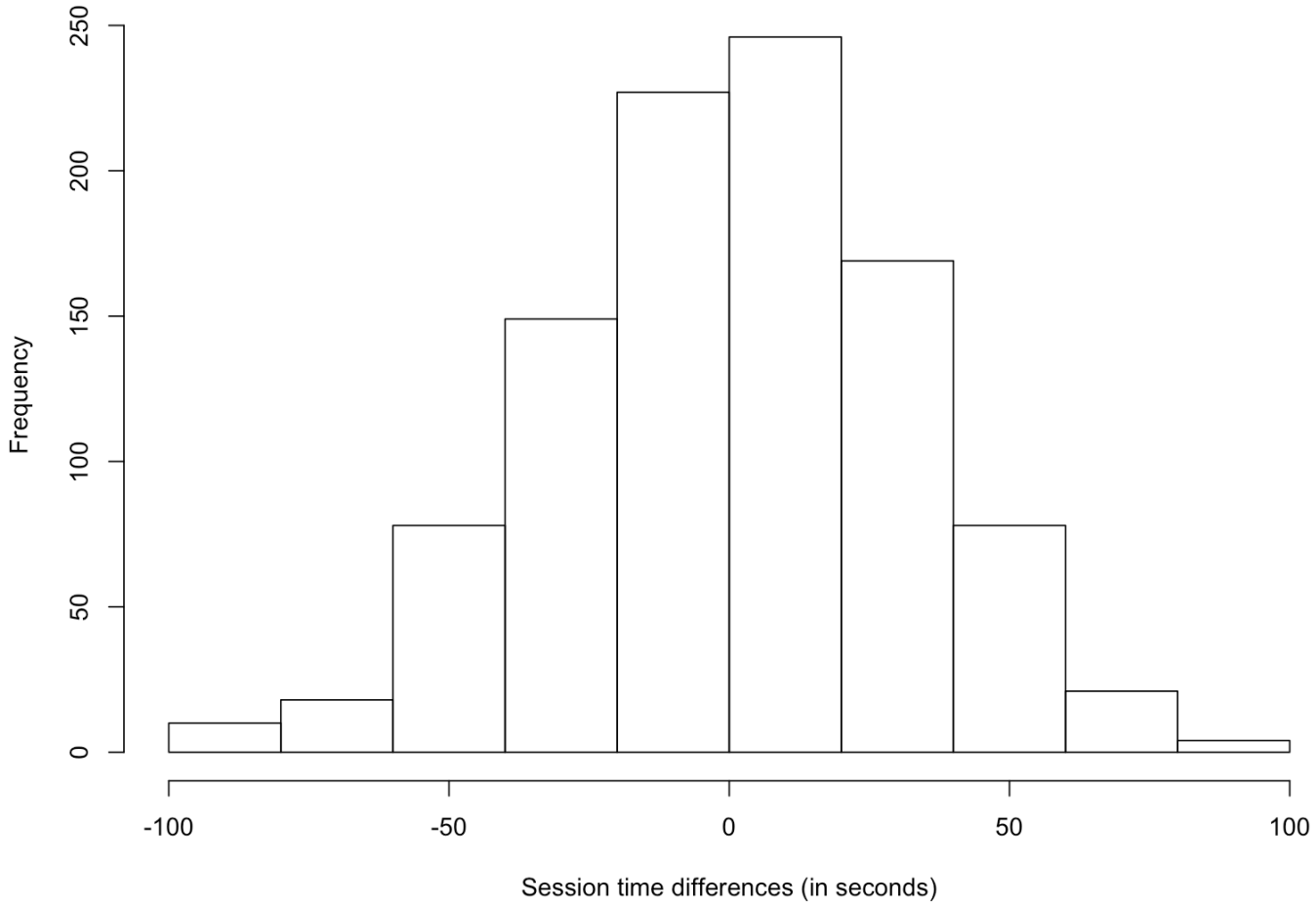
- Permutation test

```
> head(times1)
    Page Time
1 Page A   21
2 Page B  253
3 Page A   35
4 Page B   71
5 Page A   67
6 Page B   85
```

```
x = dataset$Time
n1 = 21
n2 = 15
function1 <- function(x, n1, n2)
{
  n <- n1 + n2
  idx_b <- sample(1:n, n1)
  idx_a <- setdiff(1:n, idx_b)
  mean_diff <- mean(x[idx_b]) - mean(x[idx_a])
  return(mean_diff)
}
```

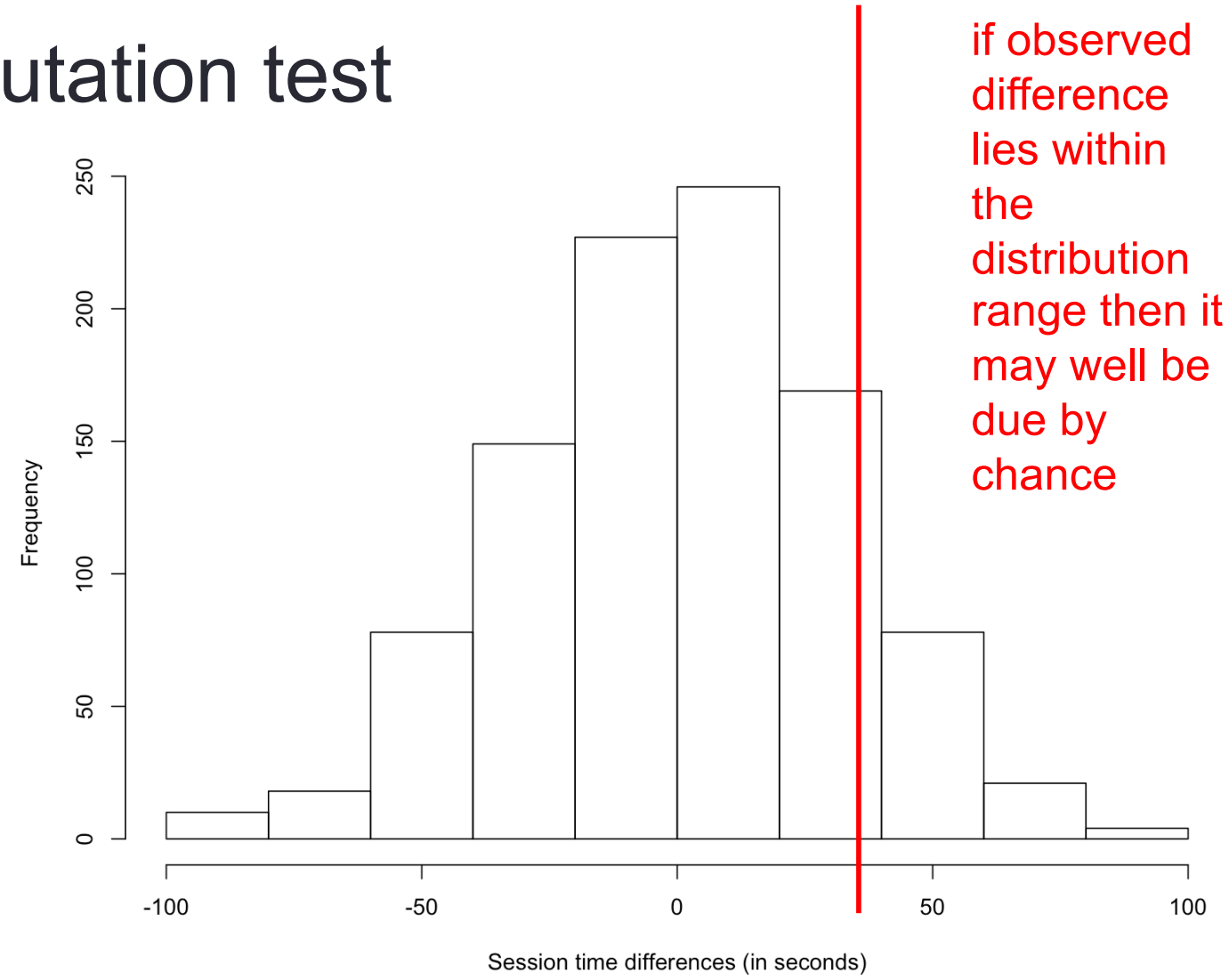
Example – Time spent in website designs

- Permutation test



Example – Time spent in website designs

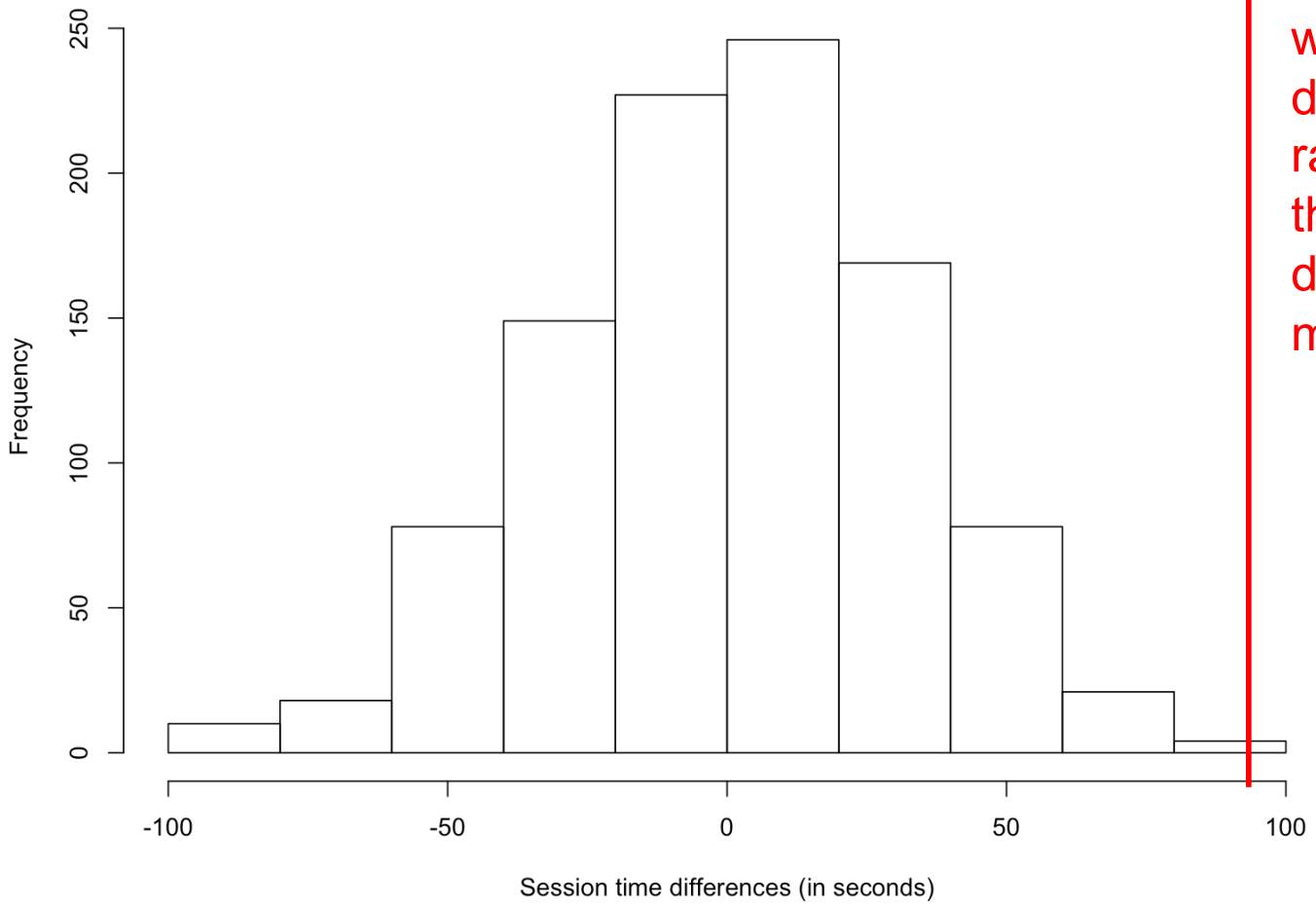
- Permutation test



if observed difference lies within the distribution range then it may well be due by chance

Example – Time spent in website designs

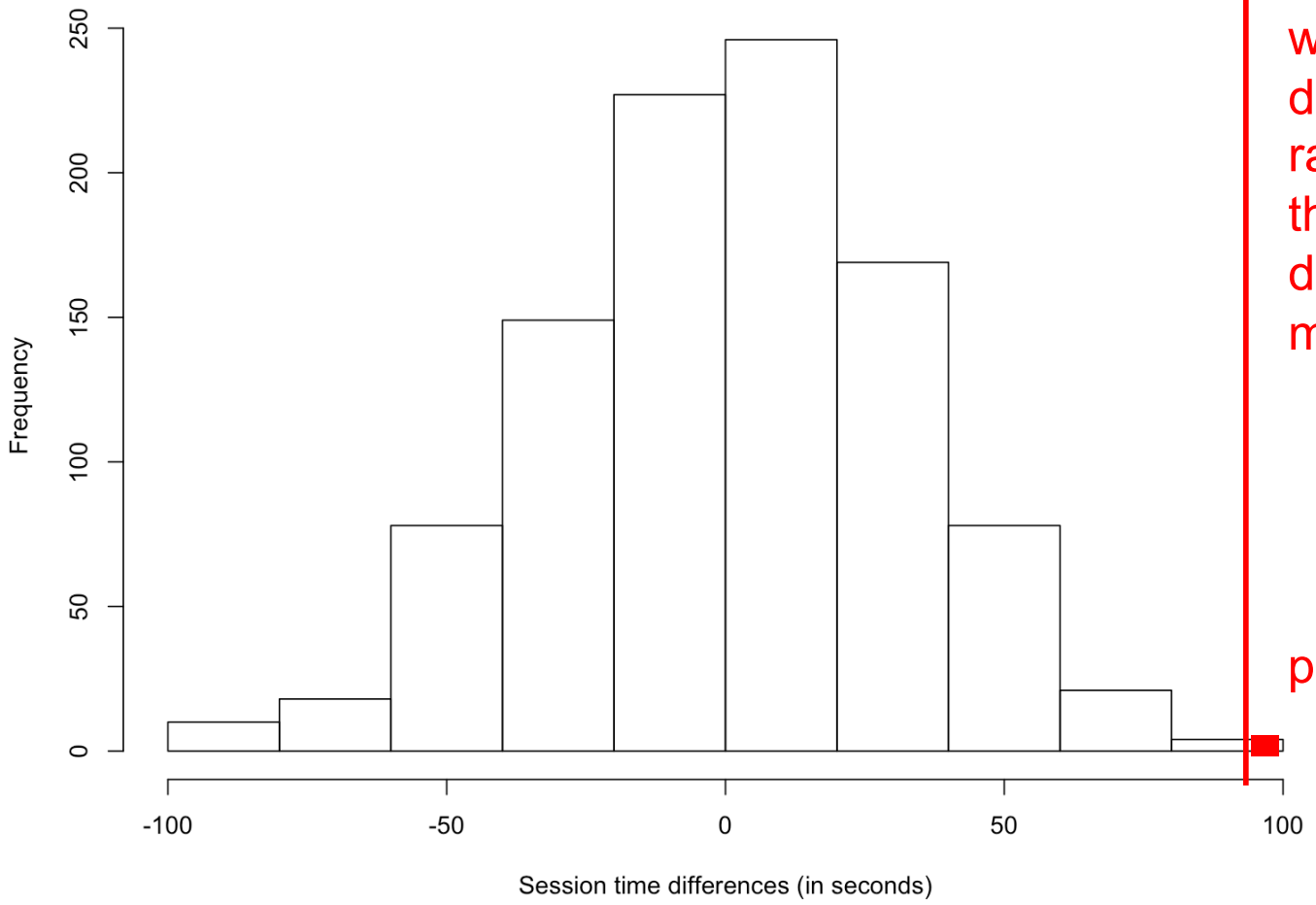
- Permutation test



if observed difference lies not within the distribution range then the TRUE difference may not be 0

Example – Time spent in website designs

- Permutation test



if observed difference lies not within the distribution range then the TRUE difference may not be 0

p-value

Example – Select best design

- The General Products Company sells a variety of household products.
- One of its products, a bath soap, is not selling well.
- To improve sales, General Products decided to introduce more attractive packaging.
- The company developed two new designs, A and B

  (A is more expensive than B)
- To determine which design sells better, two supermarkets were selected
- In one supermarket, the soap was offered using design A while in the other supermarket, design B was used

Example – Select best design

- The product scanner at each supermarket tracked every buyer of soap over a 1-week period.
- The supermarkets recorded the last four digits of the scanner code for each of the five brands of soap the supermarket sold.
- The code for General Products soap is 9077

  (other codes denote other brands).
- Management has decided to use design A only if there is enough evidence to conclude that, that design is better.
- Which design should the company use?

Example – Select best design

	Super1	Super2
brand	180	155
others	724	883
total	904	1038

Example – Select best design

	Super1	Super2
brand	180	155
others	724	883
total	904	1038

Super1	Super2
0.199115	0.1493256

It seems that design A is more attractive to the users

Example – Select best design

- The difference of proportions of sales of design A and design B is

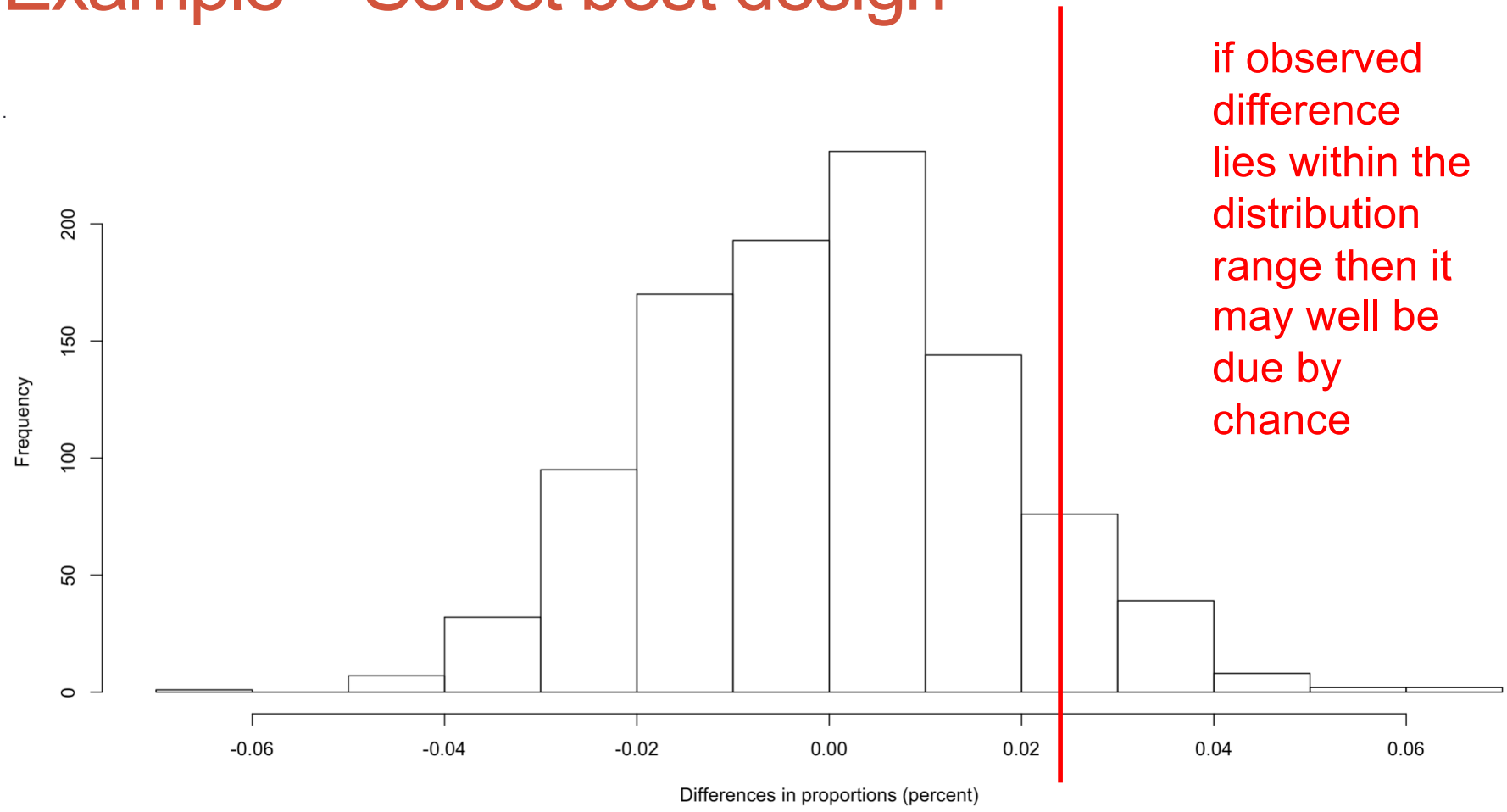
```
> obs_diff = p1hat - p2hat
> obs_diff
[1] 0.04978942
```

- This observed difference is random
- If the true difference is greater than zero, then we conclude that design A is more attractive

Example – Permutation test for proportions

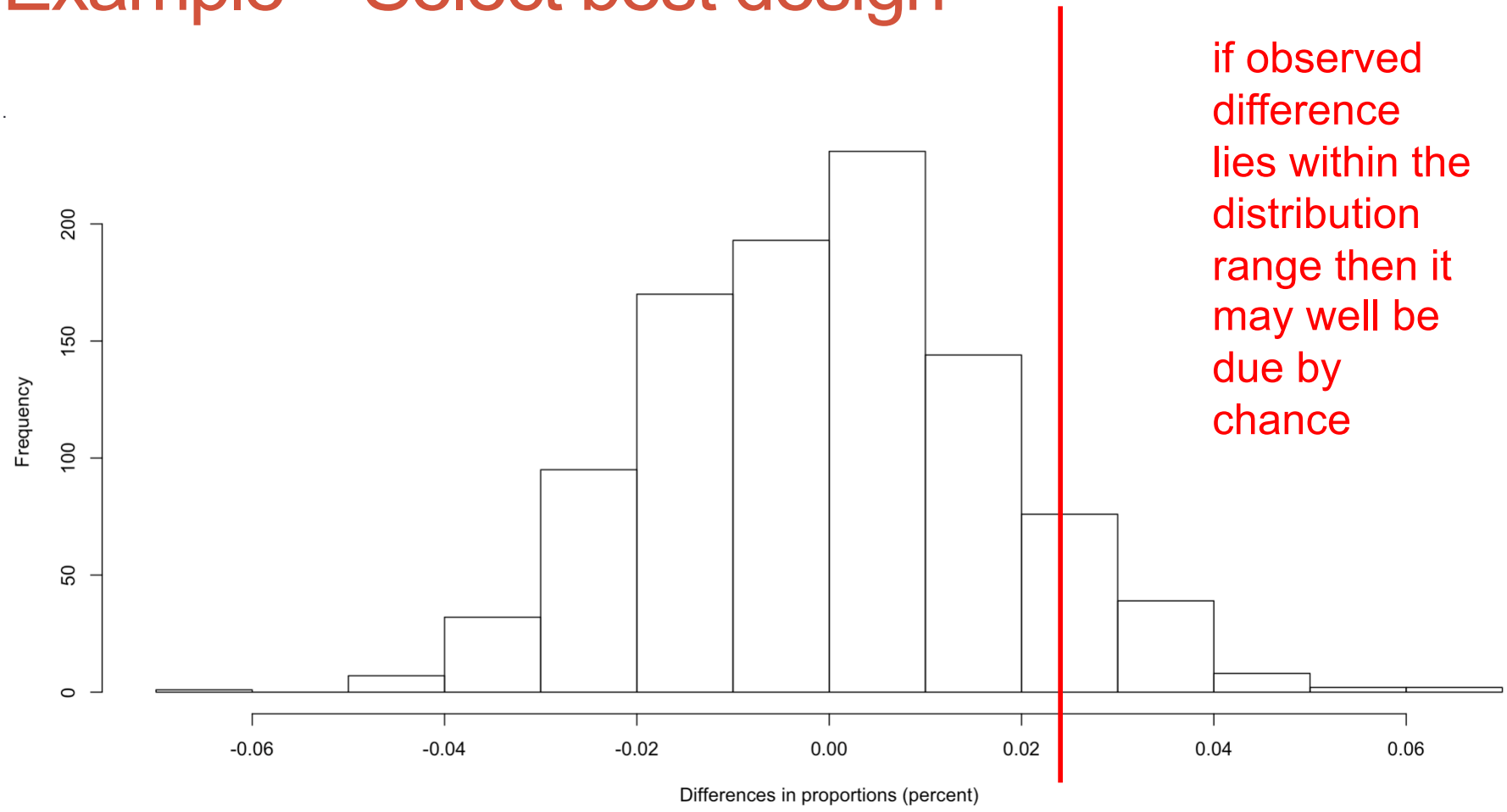
- Combine all 1942 sales in a single set
- Simulate Supermarket 1 sales
  - Randomly choose 904
  - Find the proportion of sales from design A
- For Supermarket 2,
  - Find the proportion of sales from design B in the remaining 1038
- Find the difference in these two proportions
- Repeat

Example – Select best design



if observed difference lies within the distribution range then it may well be due by chance

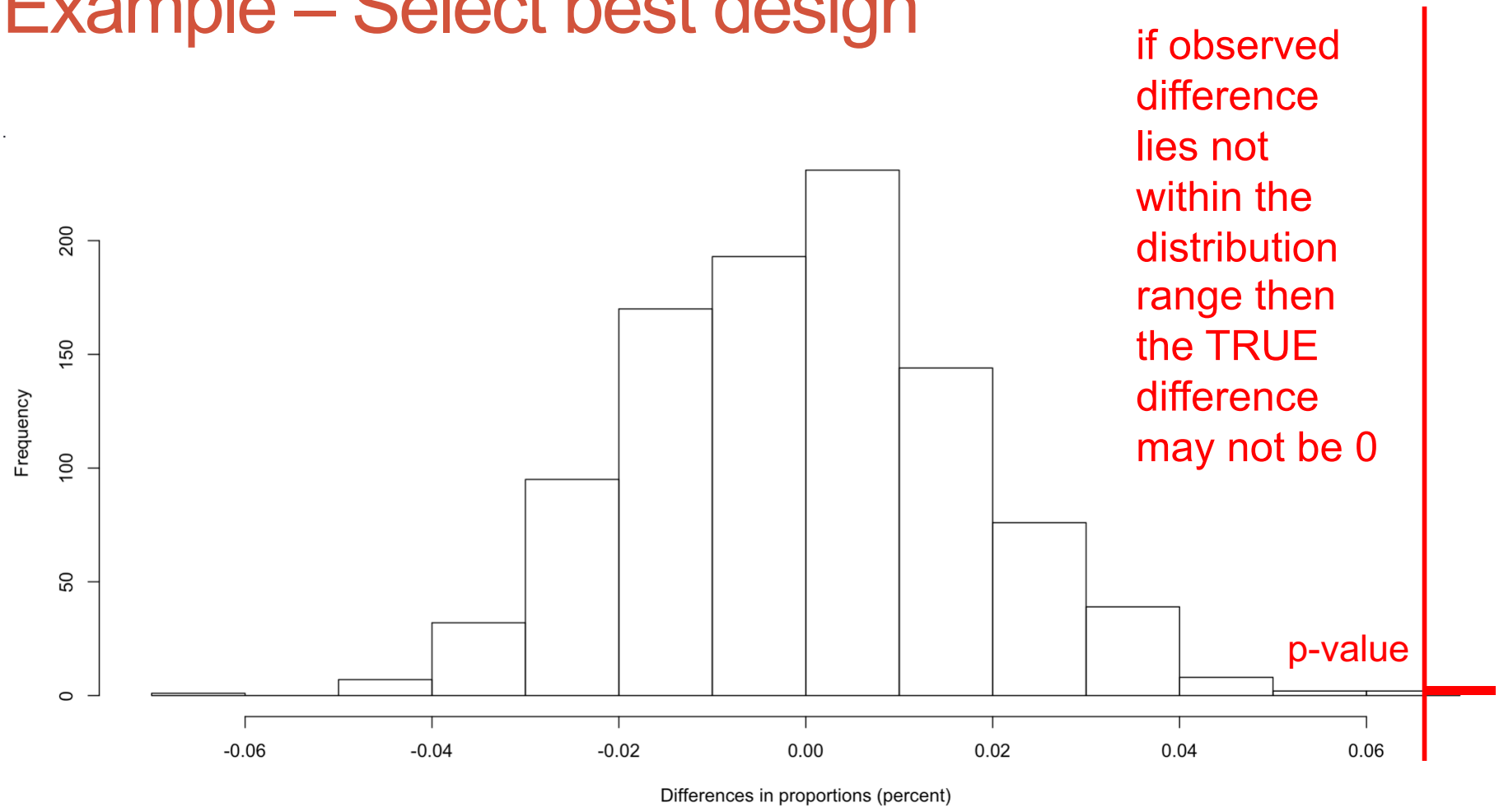
Example – Select best design



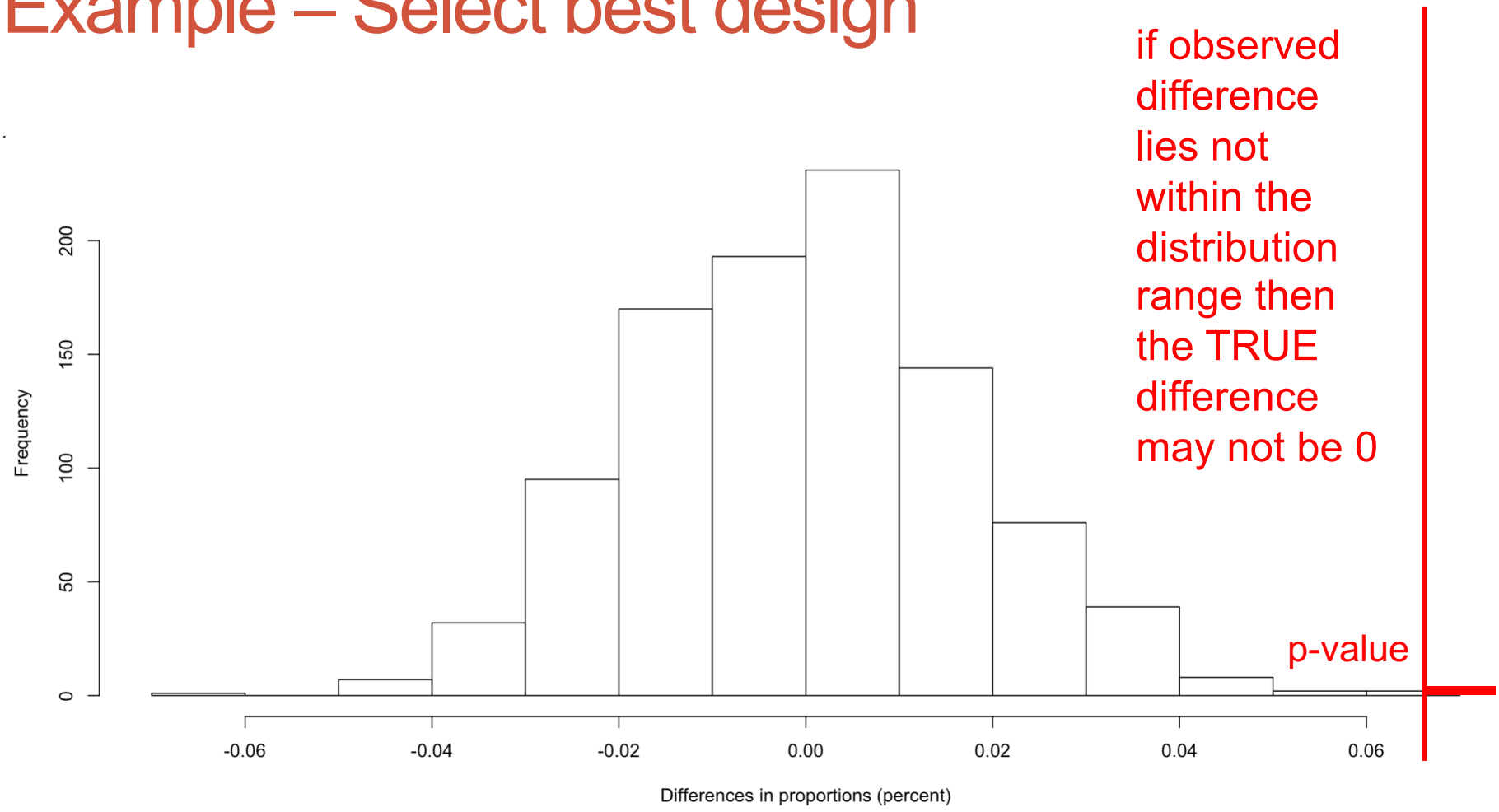
if observed difference lies within the distribution range then it may well be due by chance

We may conclude that no design is better than the other

Example – Select best design



Example – Select best design



We may conclude that one design is better than the other