



```
course = "Simulated data & Monte-Carlo simulation studies"
```

```
lesson_iteration = 1
```

```
lesson_title = "orientation + foundational concepts"
```

```
auth = "Ian Hussey"
```

```
dept = "Psychology of Digitalisation"
```

Background

of the course

Me

- I'm a user of stats, not a statistician or mathematician.
- I'm a user of code. I'm self taught, not a Computer Science graduate or trained coder.
- I use simulations to teach myself, and others, about quantitative methods to use them in research.

Background

of the course

aut = "Ian Hussey";

dept = "Psychology of Digitalisation || Digitalisation of Psychology"

You

- Why this workshop?
- What are your wants and needs from it?

Background

of the course

Why simulate?

- It gives you access to ground truth
- Take no-one's word, not even R's
- Helps you avoid unintentional p-hacking
 - Learn how to use a method before applying it to your real data.
 - Significant results no longer function as a stop signal for you to consider the analysis correct/complete.



Foundational Concepts

in Monte Carlo Simulation Studies

What is a Monte Carlo simulation?

- There is no consensus on how Monte Carlo should be defined!
 - Monte Carlo methods for quantitative (social) science methods research
 - This course
 - Monte Carlo methods as part of data analysis (e.g., MCMC in Bayesian data analysis)
 - Monte Carlo methods for the solution of general numerical problems (e.g., Monte Carlo integration)
 - Not this course

Foundational Concepts

in Monte Carlo Simulation Studies

Core components of a simulation

1. Generate pseudo-random data set with known properties
2. Analyse data with a statistical method
3. Repeat 1 & 2 many times ('iterations')
4. Summarize results across iterations
5. Make it an experiment
 - Systematically vary parameters in Step 1 (between factor)
 - Compare different ways to do Step 2 (within factor)

Foundational Concepts

in Monte Carlo Simulation Studies

Simulations to increase understanding

What is the distribution of *p* values under the null hypothesis?

Foundational Concepts

in Monte Carlo Simulation Studies

aut = "Ian Hussey";

dept = "Psychology of Digitalisation || Digitalisation of Psychology"

Simulations to increase understanding

What is the distribution of *p* values under the null hypothesis?

```
res <- replicate(10000, t.test(rnorm(n = 50, m = 0, sd = 1), rnorm(n = 50, m = 0, sd = 1))$p.value)  
res > hist()
```

Foundational Concepts

in Monte Carlo Simulation Studies

Simulations to increase understanding

What is the distribution of *p* values under the null hypothesis?

```
do it many times      analyze  
res ← replicate(10000, t.test(rnorm(n = 50, m = 0, sd = 1), rnorm(n = 50, m = 0, sd = 1))$p.value)  
generate           generate  
res > hist()  
summarise across iterations
```

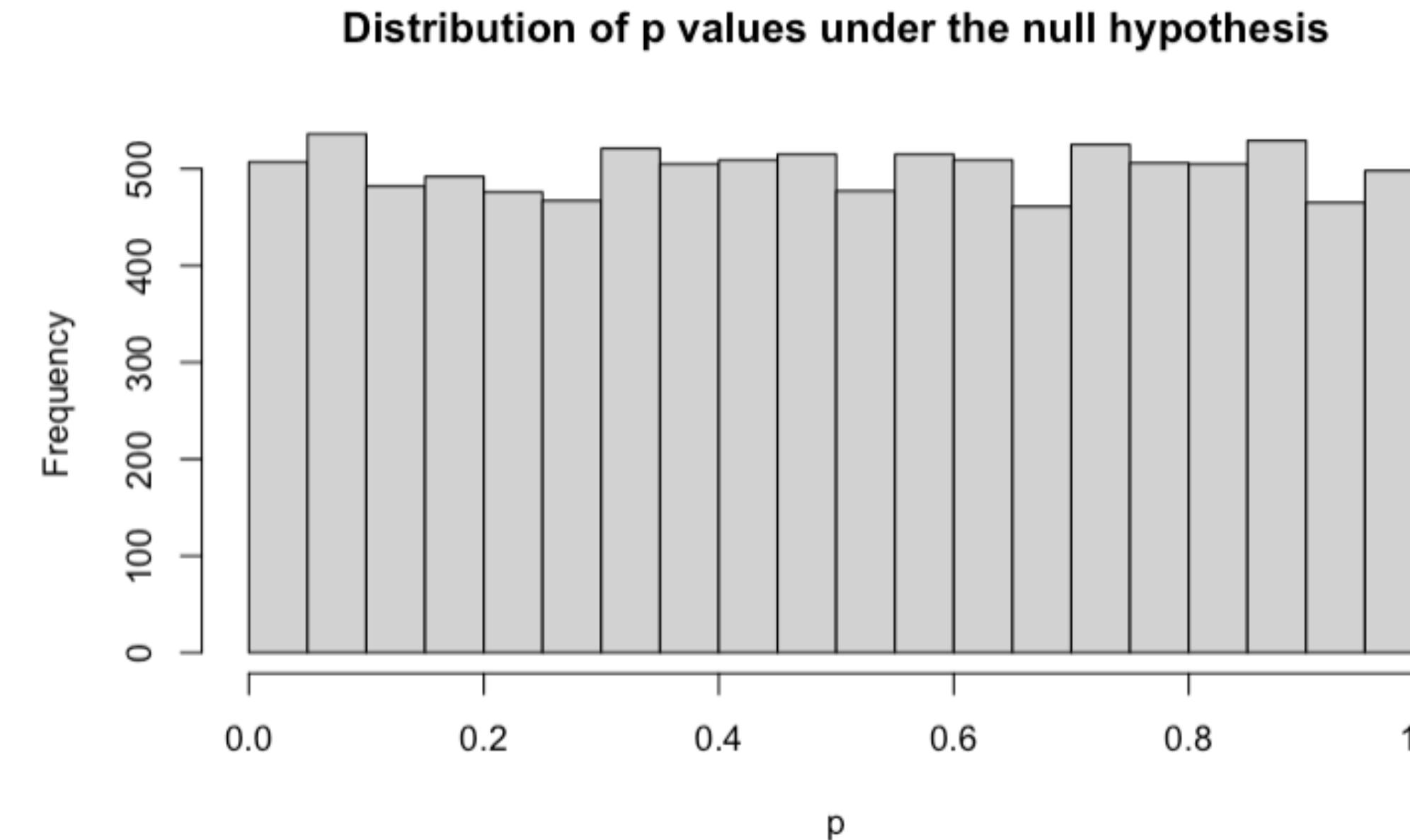
Foundational Concepts

in Monte Carlo Simulation Studies

Simulations to increase understanding

What is the distribution of *p* values under the null hypothesis?

```
do it many times          analyze  
res <- replicate(10000, t.test(rnorm(n = 50, m = 0, sd = 1), rnorm(n = 50, m = 0, sd = 1))$p.value)  
generate                 generate  
res > hist()  
summarise across iterations
```



tidy workflow

maximising transparency and reusability

```
res ← replicate(10000, t.test(rnorm(n = 50, m = 0, sd = 1), rnorm(n = 50, m = 0, sd = 1))$p.value)  
res > hist()
```

This code does not read in a linear fashion, eg top to bottom or left to right.

It makes its internal workings invisible - you cannot view the data or any intermediate steps



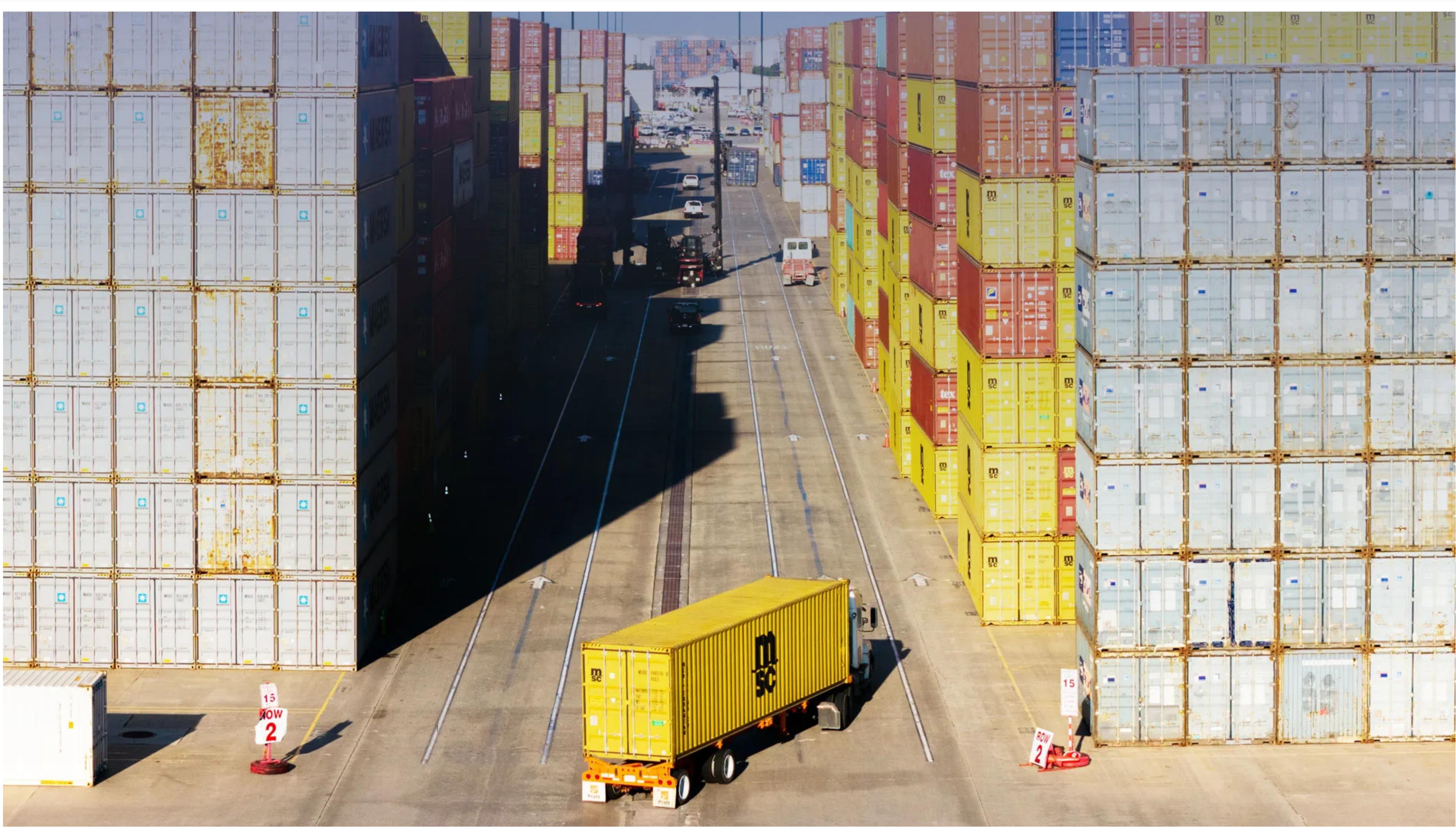
aut = "Ian Hussey";

dept = "Psychology of Digitalisation || Digitalisation of Psychology"



aut = "Ian Hussey";

dept = "Psychology of Digitalisation || Digitalisation of Psychology"



aut = "Ian Hussey";

dept = "Psychology of Digitalisation || Digitalisation of Psychology"

tidy data, tidy workflow

tidy workflow

maximising transparency and reusability

tidy inputs, tidy outputs

maximise code & workflow reusability

maximise transparency of intermediate steps

A tibble: 1 × 10									
estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	method	alternative
0.01230826	0.08646291	0.07415465	0.08756086	0.9303142	198	-0.2648943	0.2895108	Two Sample t-test	two.sided

Foundational Concepts

in Monte Carlo Simulation Studies

Core components of a tidy simulation

1. Generate tidy data
2. Analyse tidy data and return tidy results
3. Repeat 1 & 2 many times ('iterations') in a tidy way
4. Summarize results across iterations
5. Make it an experiment
 - Systematically vary parameters in Step 1 (between factor)
 - Compare different ways to do Step 2 (within factor)

Foundational Concepts

in Monte Carlo Simulation Studies

