```
course = "Improving your statistical inferences through simulation studies in R"


lesson_iteration = 1
lesson_title = "orientation + foundational concepts"


auth = "Ian Hussey"
dept = "Psychology of Digitalisation"
```

Background

of the course

Why am I here?

- 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

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.

Organisation

of the course

What we will cover

- Data simulation from scratch, with a focus on:
    - Visibility of intermediate steps and data
    - Maximising code reusability

- Sometimes using dedicated data simulation packages
    - Easier to use, but often harder to understand

- Very little math
    - Often the point of simulation is to avoid math

- Lots of code
    - tidyverse wherever possible

Organisation

of the course

Requirements & assessment

- Requirements
  - Weekly attendance
    - NO MEETING NEXT WEEK (28-02-2024)
  - At home exercises and assignments (GitHub commits)

- Assessment
  - End of course assignment:
    - Choose, design, implement, and report a simulation study
    - Scope to be determined in class

Organisation

of the course

Duzen

Call me Ian

Assessment

Weekly assignments (not graded)
Final assignment (graded)

Attendance

Weekly meetings in B202
You can miss max 2 sessions

Communication

Slack wherever possible
Email if necessary: ian.hussey@unibe.ch

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

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. Collect and aggregate 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)

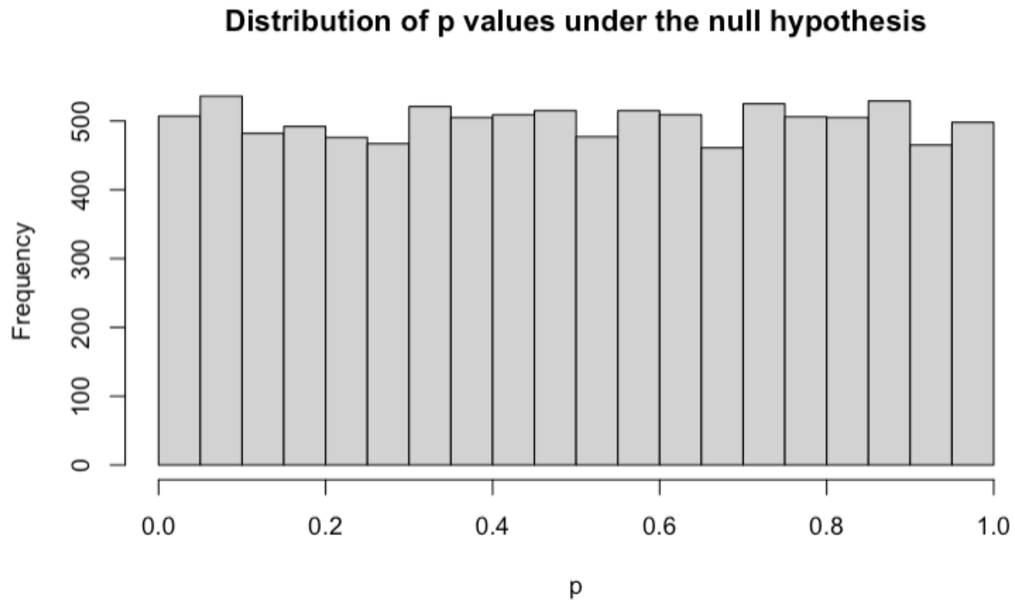
Foundational Concepts

in Monte Carlo Simulation Studies

Core components of a simulation

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

```r
replicate( # 3. repeat 1 & 2 many times ('iterations')
  n = 10000,
  expr = t.test( # 2. analyse data with a statistical method
    x = rnorm(n = 50, mean = 0, sd = 1), # 1. generate pseudo-random data set with known properties
    y = rnorm(n = 50, mean = 0, sd = 1)
  )$p.value
) ▷
  hist(main = "Distribution of p values under the null hypothesis",
       xlab = "p") # 4. collect and aggregate results across iterations
```



aut = "Ian Hussey";

logy"

Preparation

for next lesson

No meeting next week!
Meet in two weeks: 06-03-2024

- Self-study
  - 1_foundational_concepts__lesson.Rmd
  - 2_general_structure_of_a_simulation__lesson.Rmd

- Read
  - Lakens (2015) Always use Welch's t-test instead of Student's t-test. Blog post: https://daniellakens.blogspot.com/2015/01/always-use-welchs-t-test-instead-of.html
    - Relevant to the assignment, but uses the for-loop workflow we want to avoid.

- Assignment
  - 2_general_structure_of_a_simulation__assignment.Rmd

- Solution will be made available [when?]
  - 2_general_structure_of_a_simulation__solution.Rmd