

Welcome to Univariate Statistics and Methodology using R

Univariate Statistics...

```
heights
```

```
## # A tibble: 102 × 4
##   gender    HEIGHT submitted      Token
##   <chr>      <dbl> <chr>      <chr>
## 1 Female      173 02/10/2019 08:03:18 fo8qg8m4w53yil2nzuf08qxzpjkgzj0l
## 2 Female      159 02/10/2019 08:03:19 5jjo15xpzs23gdt5esks7jd7ol3uug8p
## 3 Female      164 02/10/2019 08:03:20 q2i04mqiaigqv6w82skh8orq2i04mujp
## 4 Male        187 02/10/2019 08:03:25 2n3x6kpuqjfhv3x2n3x63g7sqs3oshym
## 5 Female      183 02/10/2019 08:03:26 xsi090b2wfjb61xeryhxi0q867xls2i
## 6 NonBinary   173 02/10/2019 08:03:27 feg5s1kerz3dw5090qfeg5s1keg6x08w
## 7 Female      168 02/10/2019 08:03:27 ph3e74tai0l3urg26ph34qfsavvkkq5g
## 8 Female      170 02/10/2019 08:03:30 dbzsfmoir3mkwtlvi7udbz0iw53p0r83
## 9 Male        175 02/10/2019 08:03:36 qt2vujoe9ka2867o28sq2vujm9cz1h9
## 10 Male       165 02/10/2019 08:03:36 cvnzmtz3ej4oa54jfbpcvnmqcgr3jx
## # ... with 92 more rows
## # i Use `print(n = ...)` to see more rows
```

- one row represents one set of observations

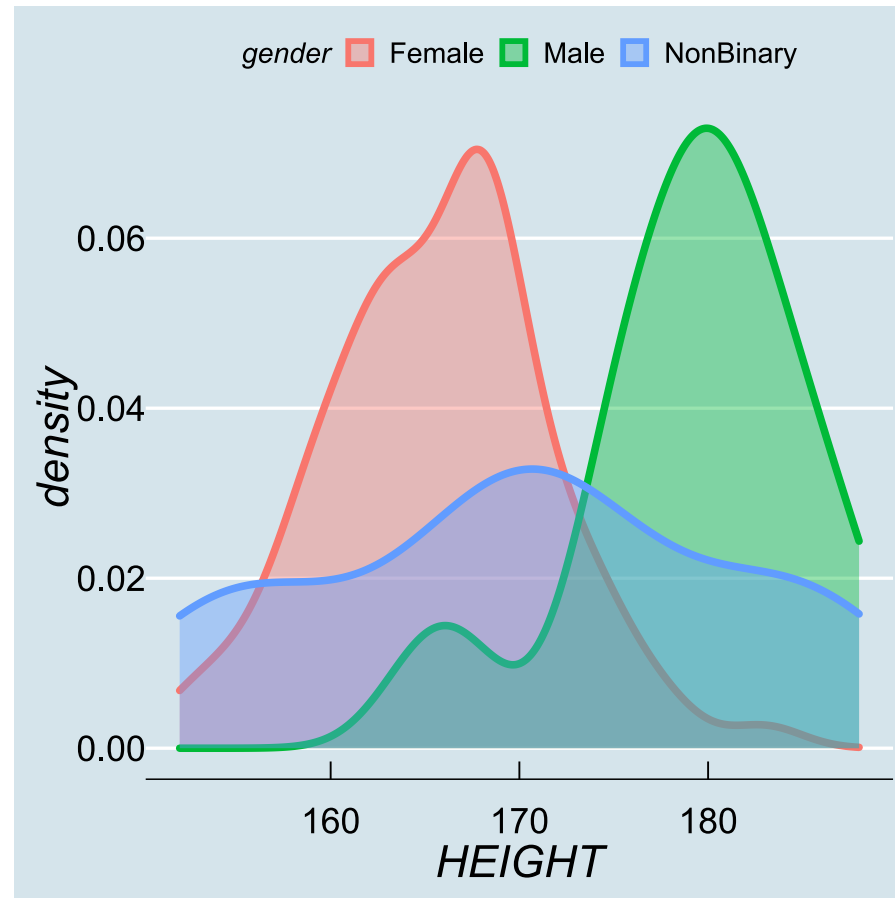
...and Methodology...

- how do we measure height?
- how do we determine gender?
- how do we collect data in a way that is generalisable?



...Using R...

```
heights %>% ggplot(aes(x=HEIGHT,color=gender,fill=gender)) +  
  geom_density(size=2,alpha=0.4)
```



...to Form Conclusions

```
heights_t <- heights %>% filter(gender %in% c('Male', 'Female'))
t.test(heights_t$HEIGHT ~ heights_t$gender)
```

```
##
##      Welch Two Sample t-test
##
## data:  heights_t$HEIGHT by heights_t$gender
## t = -9, df = 29, p-value = 6e-10
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
## 95 percent confidence interval:
##  -16.26 -10.25
## sample estimates:
## mean in group Female    mean in group Male
##           165.7           179.0
```

U
S
M
R

Univariate
Statistics
and
Methodology
using
R

What you will learn

Statistics

- foundations of Null Hypothesis Significance Testing
- probability
- samples and distributions
- the normal and the binomial distributions
- testing for significance
 - F -ratio, χ^2 , t -test, others
- the linear model
- multiple linear regression
- assumptions, models, model criticism
- logit regression (generalized linear model)

What you will learn

Statistics

- foundations of Null Hypothesis Significance Testing
- probability
- samples and distributions
- the normal and the binomial distributions
- testing for significance
 - F -ratio, χ^2 , t -test, others
- the linear model
- multiple linear regression
- assumptions, models, model criticism
- logit regression (generalized linear model)

R

- basic programming
- using libraries
- using an IDE (**RStudio**)
- data types
 - data manipulation
 - visualisation (graphs)
- functions
- running statistical models
- **RMarkdown**
 - literate programming
 - document creation

Shape of the Course



Lectures

Usually includes sections of live coding



Readings/Walkthroughs

You're encouraged to work along with these



Labs (Exercises)

A place to work through exercises with colleagues
(with help on-hand from a team of tutors)



Discussion forums and support



Assessment

Lectures & Readings

- broadly, about *concepts*
 - statistics
 - coding

Lectures & Readings

- broadly, about *concepts*
 - statistics
 - coding

Exercises

- broadly, *how to*
 - coding
 - data manipulation
 - statistics
- lots of hints and links to the readings

solutions will be available at the end of each week

Labs

- A time and place to work on the exercises
 - work should be done in **RStudio**
 - working in small groups is encouraged
 - a team of tutors will be on-hand to answer any questions

Discussions

The screenshot displays a Piazza discussion forum. On the left is a sidebar with a list of questions, each with a title, a brief description, a date (4/28/21), and a status icon. The question 'Comparing models' is highlighted in yellow. The main area on the right shows the full details of this question. At the top, it says 'question @93' with icons for search, star, and lock, and a '37 views' badge. The title is 'Comparing models'. The question text asks: 'When comparing models, should they all use the same optimiser? So if my base model converged fine with the default optimiser but a more complex model didn't so I switched optimiser, is it still ok to compare the two using anova()?' Below the question, there is a 'Thank you!! :)' and a 'other' tag. At the bottom of the question area, there are 'edit' and 'good question' buttons, and a '0' count. Below the question area, there is a section for 'the instructors' answer, where the text reads: 'That's a very deep question. You said that the default optimiser didn't return errors. So, I'll have to make an assumption here. Assuming that the default optimiser reached the right optimum, then yes - you can compare them if one used the default optimiser and the other model user another optimiser.'

- piazza discussion forums for the course on Learn
 - ask questions, share experiences, talk to the course team
 - post anonymously if preferred
 - an important way to keep in touch

Support



we are here to help you

- lectures: feel free to ask questions at any point
- lab sessions: make use of the tutoring team (they *want* to help!)
- piazza discussion forums: any time
- office hours: see Learn page for details



Interim Assessments (20%)

- four online multiple-choice quizzes
- for each quiz, **one attempt** which must be completed within 60 min

Quizzes released on Mondays of Weeks 2, 4, 7 and 9.

must be completed within 7 days of release



Final Assessment (80%)

- check a dataset for consistency
- explore hypotheses about how variables are related
 - conduct and interpret appropriate statistical tests
- produce suitable graphics
- present your workings and conclusions in a markdown document



Final Assessment (80%)

- **check** a dataset for consistency
- **explore** hypotheses about how variables are related
 - conduct and **interpret** appropriate statistical tests
- produce **suitable** graphics
- present your workings and **conclusions** in a markdown document



Final Assessment (80%)

- **check** a dataset for consistency
- **explore** hypotheses about how variables are related
 - conduct and **interpret** appropriate statistical tests
- produce **suitable** graphics
- present your workings and **conclusions** in a markdown document

coursework released: 17 Nov 2022

coursework submission deadline: 15 Dec 2022

Tips for survival

1. active engagement!
2. use the piazza forums and other forms of support whenever you need them
3. keep on top of the coursework
4. remember that some things will feel difficult at first
 - what's hard for you may be easy for others
 - what's easy for you may be hard for others
5. most importantly, *don't give up*

Tips for improving the course

- we have worked hard to make this course work, but some things will undoubtedly still be improvable
- please feed back to us!
 - during labs/after lectures, via email/forums/anonymously on Learn via "Have Your Say"
- please bear with us!
 - we are doing our very best
- any good course is a *conversation* between teachers and learners
 - we try our best to listen, but this also will only work with your help

first tasks

1. help us get to know you
2. get the software
3. fill in the survey

1. help us get to know you

1. Log into Learn, navigate to the USMR course page
2. look for the Piazza Forum button
 - it should automatically enroll you on Piazza
3. post a little introduction of yourself

Suggested content:

- pictures of pets/animals very welcome.
- answer a randomly generated question from <https://learnhip.com/randomq/>

2. get the software

Option A

PPLS RStudio Server: <https://rstudio.ed.ac.uk>

- no installation required
- accessed via a web browser (requires internet)

Option B

Install it yourself. Instructions at <https://edin.ac/3B0oi5A>

- process is a bit more involved

Our Recommendation

Go with option A for now.

After the course (over the christmas break), move to option B in time for Semester 2.

2. get the software

Option A

PPLS RStudio Server: <https://rstudio.ed.ac.uk>

- no installation required
- accessed via a web browser (requires internet)

- *The PPLS Rstudio Server will undergo regular maintenance on the first Sunday of every month, so there may be a short interruption to the service on these days*
- *Our agreement with RStudio allows us the use of RStudio Server for teaching purposes only. **Please do not use the RStudio server for your dissertation***

Option B

Install it yourself. Instructions at <https://edin.ac/3B0oi5A>

- process is a bit more involved

3. fill in the survey

- Found at <https://edin.ac/3CHTRm8>
- Takes 10-15 mins to complete
- Provides real data for the examples and exercises

Acknowledgements

- icons by Becris, tezar tantula, ProSymbols, Arjan Farzkenari, Econceptive from the [Noun Project](#)