STAT115: Introduction to Biostatistics

University of Otago

Ōtākou Whakaihu Waka

Lecture 1: What is statistics? And biostatistics?

- Learning from data
- Wikipedia:
  - "Statistics is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data" [1]
  - It's all about data.
  - Learning from data involves all of those concepts
- What's *biostatistics* then?
  - Biostatistics is the application of statistics in health-related fields, including medicine, biology, and public health.[2]



[1] I got a similar answer when I asked ChatGPT.

[2] Wikipedia doesn't do quite so well on this one.

Data

- Data is all around us
  - ▶ It informs us about the natural world, business, society, disease, ...
- In the past, data sets tended to be small
  - ▶ Data was expensive to collect (it often still is!)
  - ▶ Much was done with pen and paper
- It is now common to have large data sets
  - ▶ Computing is an essential part of modern statistics

Then

Now

Proportion of SARS−CoV−2 Variants Over Time in New Zealand
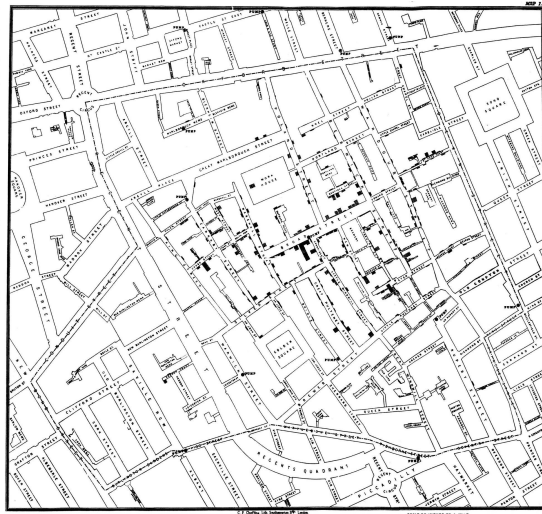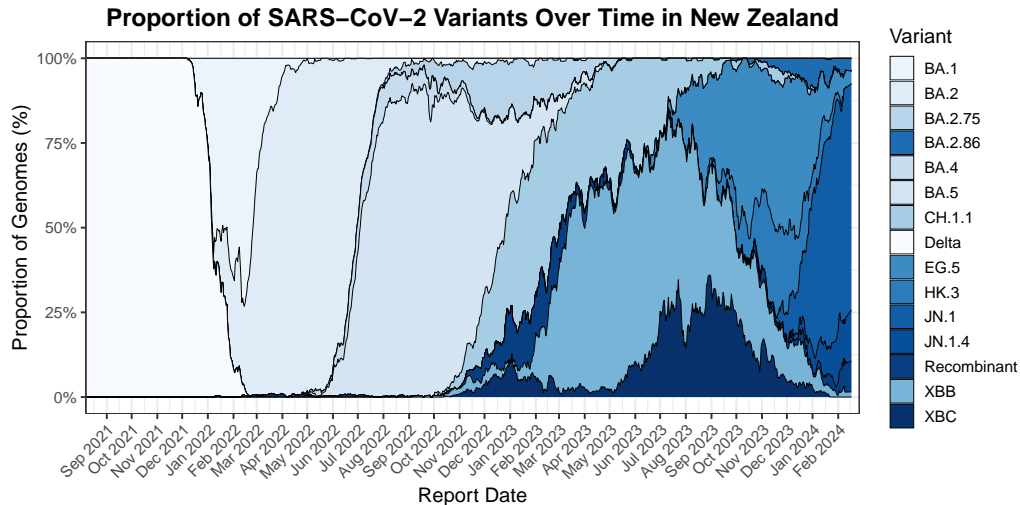
Then and Now

- *Then*:
  - ▸ John Snow's celebrated 1854 map of cholera cases in London.
  - ▸ A few hundred cases.
  - ▸ Time consuming.

- *Now*:
  - ▸ Breakdown of genetic strains of COVID in NZ, through time.
  - ▸ Based on 13860 records.
  - ▸ Very quick.[3]

---

[3] https://mathstatfiles.otago.ac.nz/STAT115/covidplot.r

Data

- We will see a lot of data in this course
  - ▶ Lectures
  - ▶ Tutorial exercises
  - ▶ Assignments
- Data comes from a variety of sources
  - ▶ Variety of subject areas
  - ▶ You will hopefully see examples that interest you
  - ▶ Statistical methods are easily transferable

Variation and Uncertainty

- Data produced by processes full of variability and uncertainty
  - ▶ Contemporaneous opinion polls can give different results
  - ▶ Patients vary in response to treatment
  - ▶ Tomorrow's weather impossible to predict with perfect accuracy
- Statistics helps us learn in the presence of uncertainty
  - ▶ Quantify uncertainty
    - – Use probability
 - ▶ Understand sources of variability
    - – Separate signal from noise

Example: Harold Shipman

- Harold Shipman was a notorious serial killer
- He had 215 confirmed kills
  - A further 45 suspected kills
- He was a British GP
- His victims were predominantly:
  - Older
  - Female
- What does this have to do with statistics?

Shipman's statistical legacy

Harold Shipman, who in January committed suicide in prison, has become notorious the world over as one of the most prolific serial killers of all time. His case has also seriously dented public confidence in doctors. **David Spiegelhalter** and **Nicky Best** explain how industrial quality control techniques could be adapted to signal when death rates among a doctor's patients are surprisingly high, and the tricky issues that would arise in implementing such a monitoring system.

Dr Harold Shipman arrives at Ashton-under-Lyme police station (photograph copyright Chris Glasow, REX syndication)
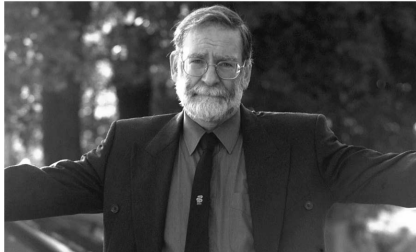
Example: Harold Shipman

Variability

- Could statistics have detected Harold Shipman's offending earlier?
- A patient dying is not unusual
  - ▶ Older patients tend to be more likely to die
- The number of patient deaths varies
- The timing of patient deaths varies
- The death rate varies by age, sex, ...
- Expect variation in the number of death certificates signed by different doctors
 - ▶ Some would sign more, some less

Example: Harold Shipman

Variability

- A doctor signs one death certificate in one afternoon
  - ▸ Not unusual
- A doctor signs fifty death certificates in one afternoon
  - ▸ Concerning
- Idea: there is some range of values that are 'expected' or 'normal'
- Calculate excess deaths compared to an average doctor
  - ▸ Based on probability
  - ▸ Account for factors like patient age

Example: Harold Shipman

- Excess deaths in 1998: 175 women and 49 men[4]
  - ▶ Close to number of confirmed kills

---

[4] https://mathstatfiles.otago.ac.nz/STAT115/shipman.r
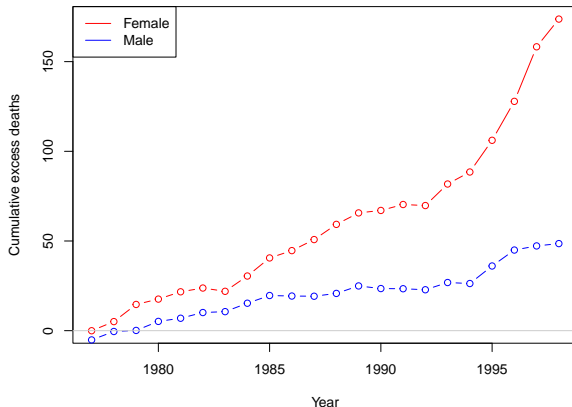
Example: Harold Shipman

Role of statistics

- It may be feasible to monitor doctors using statistics
- Potential issues:
  - ▶ Data availability
  - ▶ Privacy concerns
  - ▶ False positives: unusually high numbers of deaths
    - − By chance
    - − Case mix (e.g. predominantly work in rest homes)
    - − Data quality

Course Organization: Contact email and teaching staff

Contact: <u>stat115@otago.ac.nz</u>

Peter Dillingham   Martin Hazelton   Conor Kresin   Katrina Sharples   Phil Wilcox

Course Organization: Lectures

- Stream 1
  - Mon 8am, Tue 8am, Wed 8am, (Thu 3pm)
- Stream 2
  - Mon 10am, Tue 2pm, Wed 4pm, (Thu 9am)
- **Thursday lectures will not run unless specifically announced**
- Lecture recordings will be posted – but better to attend in person.

Course Organization: Tutorials

- In room MA124, first floor of the Science III building.

- Tutorials are unstreamed and not compulsory but **highly recommended**!

- Attend as many as you wish (space permitting).

- **Tutorial times** are available on the Course Schedule on the STAT115 Blackboard site.
  - They will not appear on your timetable (since not compulsory) so make a note.

- You can get help at the tutorials on any of the paper material including lectures, assignments and use of the R software.

- The tutorial room computers have R installed on them.

Course Organization: Maths Support

- The *basics booklet* on Blackboard covers material you are expected to be familiar with.

- Help available in tutorials.

Course Organization: Assessment

Final mark (F) is a weighted combination of your exam (E) and assignment (A) marks, all expressed as percentages:

$$F = 0.7\,E + 0.3\,A$$

Assignments:

- 10 assignments for the semester: see course schedule for dates.
- Each assignment carries equal weight (so contributes 3% of final mark).
- Practice (non-assessed) versions of assignments available – use them!
- **No extensions are possible**

Course Organization: Assessment

Final Exam:

Duration 3 hours

Format All multi-choice (about 90 questions)

Calculators Any model of calculator provided this is battery powered, silent, truly portable and free of communication capabilities.

Date To be confirmed: set by the Examinations Office (not the Department)

Course Organization: Academic Integrity and Misconduct

Academic integrity means being honest in your studying and assessments. It is the basis for ethical decision-making and behaviour in an academic context. Academic integrity is informed by the values of honesty, trust, responsibility, fairness, respect and courage..

Academic Misconduct, such as plagiarism or cheating, is a breach of Academic Integrity and is taken very seriously by the University. Types of misconduct include plagiarism, copying, unauthorised collaboration, taking unauthorised material into a test or exam, impersonation, and assisting someone else's misconduct. .

See Course Outline for further details.

Course Organization: Blackboard

- Blackboard hosts many resources for STAT115
  - Course Outline (**read this!**)
  - Course Schedule (**diarize this!**)
  - Lecture slides
  - Announcements
  - A basics booklet that contains a review of assumed maths
  - Link to lecture recordings
  - Formulae sheet
  - Revision help
- Link: `blackboard.otago.ac.nz`

Summary

- Statistics is learning from data

- Statistics is about describing and quantifying variability

- Read the Course Outline!