THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Analysis of Engineering and Scientific Data

Semester 1 – 2019

Sabrina Streipert s.streipert@uq.edu.au

Lecture 1

1. Course Information
2. Probability versus Statistics
3. Deterministic versus Stochastic Model
4. Statistical Inference
5. Introduction to R-Studio

Useful information

- Information about the course, its learning objectives, etc is available on the *Electronic Course Profile*.
- The course has a blackboard page.
- The course also has a webpage.
- Syllabus on blackboard and the course webpage:
  https://courses.smp.uq.edu.au/STAT2201/2019a/

Supportive literature for this course:

Applied Statistics and Probability for Engineers

by Douglas C. Montgomery and George C. Runger *(6^{th} edition)*

Grading

- 2 hour final exam

  Need 40% to pass the course.
  The mark on your final exam makes 60% of the total course grade.

- 6 homework assignments

  Your best 5 homework assignments will be counting towareds your total course grade.

  This grade will make 40% of the total course grade.

- 6 tutorial sessions to help with the homework assignments.

Homework Assignments

Homework as **R Markdown** document.

Week	Dates	Lecture	Tutorial	Assignment Due
1	25/2–01/3	✓	x	x
2	04/3–08/3	✓	✓	x
3	11/3–15/3	✓	x	✓
4	18/3–22/3	✓	✓	x
5	25/3–29/3	✓	x	✓
6	01/4–05/4	✓	✓	x
7	08/4–12/4	✓	✓	✓
8	15/4–18/4	✓	x	x
9	22/4–26/4	x	x	x
10	29/4–03/5	✓	✓	✓
11	07/5–10/4	✓	x	✓
12	13/5–17/5	✓	✓	x
13	20/5–24/5	x	x	✓
14	27/5–31/5	x	x	x

Consultation Hours

Consultation Hour:

- **Friday:** **2-3pm** 82 E (Seddon Building) - 409

- **By appointment:** s.streipert@uq.edu.au

Tutors will not provide consultation hours.

Lecture Notes

- The morning of each class, the preliminary version of the lecture slides is available in a workbook format on blackboard.

- Lecture Notes will be uploaded on *blackboard* in the section of *Learning Resources* as well as on the course webpage by 12pm the day after lecture.

- Lecture notes follow the supportive literature:

  Applied Statistics and Probability for Engineers – Montgomery and Runger

Programming Language

In this course, we will be using the statistics software **R**.

(created by Ross Ihaka and Robert Gentleman at the University of Auckland, NZ in 1992)

Benefits of R:

- R is a programming language and free software environment.
- R is widely used among statisticians and data miners.
- R has a large support community online.
- R is an implementation of the S programming language.

Other languages: S, Matlab, Python, Julia.

Probability vs. Statistics

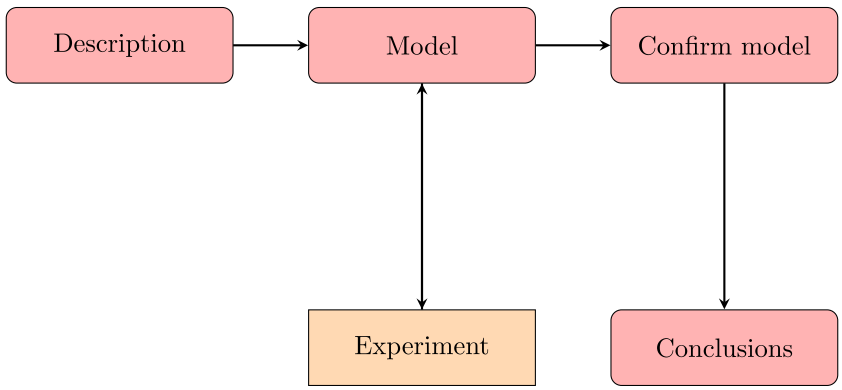
Probability deals with predicting the likelihood of future events.
Q: What is the likelihood of a rainy day tomorrow ?

Statistics involves the analysis of the frequency of past events and with it, collecting data.
Q: How long do patients have to wait in the emergency room? - important questions because some patients leave without treatment...

Statistics can be used for creating **probabilistic** models.

Data Science is an emerging field, combining statistics, big-data, machine learning and computational techniques.

Engineering Method



Deterministic Models

To explain processes, we can use **deterministic** model.

The output of a **deterministic** model is fully determined by the parameter values and the initial conditions.
For example:

$$I = \frac{V}{R} \qquad I(t_0) = I_0,$$

where

- I is the current,
- V is the voltage,
- R is the resistance.

What are external factors that could impact this result?
Will you always get the same results if V and R are the same?

Stochastic Models

Given a deterministic model, there are many different ways to add stochasticity:

For example:

- The measurement observation for I could be wrong:

$$I_{obs} = I_{true} + \epsilon = \frac{V}{R} + \epsilon$$

- The measurement of the voltage could contain an error:

$$I = \frac{V_{obs}}{R} = \frac{V_{true} + \epsilon}{R}$$

- Similarly for the resistance R:

$$I = \frac{V}{R_{obs}} = \frac{V}{R_{true} + \epsilon}$$

- And others, such as combinations of the above.

Statistical Inference

In the previous example, we reasoned from general laws to specific cases.

But what if there is no general law at hand?

We can also, given a set of measurements, reason from a sample to a population (i.e. to a more general case).

$\longrightarrow$ **statistical inference**

Data Collection for statistical inference

- retrospective study
- observational study
- designed experiment

Important in this case is the understanding of:

- samples
- impact factors
- limitations
- assumptions

Models by statistical inference

Given:

Measurements (output): y ($\vec{y} = (y_1, \ldots, y_N)$)
Impact factors (explanatory variables): $x_1, x_2 \ldots, x_k$ ($x_i = \vec{x}_i$).

Aim: Find f such that

$$y \approx f(x_1, \ldots, x_k)$$

(*Regression Models*)

Probability is then used to analyse the statistical inference.

Chapter 2

1. Sample Space, Outcomes and Events
2. Probability
3. Conditional Probability and Independence
4. Birthday Problem (Monte Carlo)

Random Experiment

= experiment that yields different results each time

Examples:

- Tossing a coin
- Selecting a ball from an urn containing coloured balls
- Amount of rain in Queensland
- Measuring the current on a physics experiment
- Asking for someone's birthday

Sample Space

Random experiment $\longrightarrow$ different results $\longrightarrow$ define sample space

Sample Space = Ω = set of all possible (non-decomposible) outcomes.

non-decomposible outcomes means that it can not be a combination of other outcomes

Example: Having the birthday at the 25th of March 1998 versus having the birthday in March 1998.

Note: Only one outcome can occur.

Sample Space - Examples

- Tossing a coin
- The amount of water in the Fitzroy river?
- The amount of students that share the same birthday in this class?
- The number of professors in Mechanical Engineering at UQ
- The salary you might earn as an engineer?

Sample Space - Types

Ω

Discrete | Continuous

$\longrightarrow$ revisit examples on previous slide and determine the type of sample space.

Subset of Sample Space (Event)

Any subset of the sample space = **Event**

Events are usually denoted by upper-case letters $A, B, C, \ldots$.

Event A occurs if the outcome of the experiment is one of the elements in A.

Example:

- Having a repeated number in a 6-digit phone code.
- Having all students here have their birthday in March.
- Having a new-born weight more than 4kg.
- Having the same outcome when tossing a coin three times.

Events

Say you wonder about the months in which all Stat 2201 students have their birthday in.

What is the sample space?

$\Omega = \{Jan, Feb, \ldots, Dec\}$.

What could be a possible event here?

$A = \{June, July, August\}$

What do you think is the "smallest" event?

Having a student not having birthday in any of the month January to December. **Null Event** $A = \emptyset$

What do you think is the "biggest" event?

The event that all students have their birthday in any of the months, that is $A = \{1, 2, 3 \ldots, 12\} = \Omega$ (certain event).

Set Operations

Events $A, B, C, \ldots$ are **(sub-)sets** of the sample space Ω

a) $A \cup B$ "Union of A and B"

b) $A \cap B$ "Intersection of A and B"

c) $\overline{A} = A^c = \Omega \backslash A$ "Complement of A" = Everything except A.

d) $A \subset B$ "A is a subset of B"

Example:

$\Omega = \{Jan, Feb, \ldots, Dec\},\ A = \{June, July, Aug\},\ B = \{Aug, Oct, Dec\}$.
Then

a) $A \cup B = \{June, July, Aug, Oct, Dec\}$

b) $A \cap B = \{Aug\}$

c) $A^c = \overline{A} = \{Jan, Feb, March, May, Sept, Oct, Nov, Dec\}$

d) For example $C = \{July, Aug\} \subset A = \{June, July, Aug\}$

Remarks to Sets

If $A \cap B = \emptyset$, we say that event A and event B are **"mutually exclusive"** or **"disjoint"**.

If $A \subset B$ and $C \subset B$ does this imply that $A \cap C \neq \emptyset$?

No, not necessarily, for example $B = \{1, 2, 3, 4, 5\}$ and $A = \{1, 2\}$ and $C = \{4\}$.

$A \cap \overline{A} = \emptyset$.

Example:

Consider a digital scale that provides weights to the nearest gram.
What is the sample space?

$$\Omega = \mathbb{N}_0 = \{0, 1, 2, 3, 4, \ldots\}.$$

Let A be the event that a weight exceeds 10 grams $\rightarrow A = \{11, 12, 13, \ldots\}$.

Let B be the event that a weight is greater than or equal to 7 grams
$\rightarrow B = \{7, 8, 9, \ldots\}$.

Let C be the event that a weight is less than or equal to 16 grams
$\rightarrow C = \{0, 1, 2, 3, \ldots, 16\}$.

Find $A \cup B$ $\quad A \cup B = \{7, 8, 9, \ldots\}$.

Find $A \cup (B \cap C)$

$$A \cup (B \cap C) = \{11, 12, \ldots\} \cup \{7, 8, 9, \ldots, 15, 16\} = \{7, 8, 9, \ldots\}$$

Find $\overline{(A \cup B)} \cap C$

$$\overline{(A \cup B)} \cap C = \overline{\{7, 8, 9, \ldots\}} \cap \{0, 1, 2, \ldots, 15, 16\} =$$
$$\{0, 1, 2, 3, 4, 5, 6\} \cap \{0, 1, 2, \ldots, 15, 16\} = \{0, 1, 2, 3, 4, 5, 6\}$$