MAS2903: Introduction to Bayesian Methods

Dr. Lee Fawcett

Semester 2, 2019—2020

Contact details

My name: Dr. Lee Fawcett

Office: Room 2.07 Herschel Building

Phone: 0191 208 7228

Email: `lee.fawcett@ncl.ac.uk`

www: `www.mas.ncl.ac.uk/~nlf8`

Timetable and Administrative arrangements

- **Lectures** are on Mondays at 3 (**Bedson LT1**) and Thursdays at 1 (**Bedson LT1**). In the first two weeks there will be a third lecture on Thursdays at 2 (**Bedson LT1**; in effect meaning we will have a double lecture on a Thursday!).

- **Problems classes** are in ODD teaching weeks starting in teaching week 3. These take place on Thursdays at 2 (**Bedson LT1**).

- **Drop–in sessions** are at the same time and place in EVEN weeks starting in teaching week 4.

- **Computer practicals**: Tuesdays at 12 (Herschel full PC cluster). These happen *occasionally*, and you will be notified of these well in advance.

Timetable and Administrative arrangements

- **Lectures** are on Mondays at 3 (**Bedson LT1**) and Thursdays at 1 (**Bedson LT1**). In the first two weeks there will be a third lecture on Thursdays at 2 (**Bedson LT1**; in effect meaning we will have a double lecture on a Thursday!).
- **Problems classes** are in ODD teaching weeks starting in teaching week 3. These take place on Thursdays at 2 (**Bedson LT1**).
- **Drop–in sessions** are at the same time and place in EVEN weeks starting in teaching week 4.
- **Computer practicals**: Tuesdays at 12 (Herschel full PC cluster). These happen *occasionally*, and you will be notified of these well in advance.

Timetable and Administrative arrangements

- **Lectures** are on Mondays at 3 (**Bedson LT1**) and Thursdays at 1 (**Bedson LT1**). In the first two weeks there will be a third lecture on Thursdays at 2 (**Bedson LT1**; in effect meaning we will have a double lecture on a Thursday!).
- **Problems classes** are in ODD teaching weeks starting in teaching week 3. These take place on Thursdays at 2 (**Bedson LT1**).
- **Drop–in sessions** are at the same time and place in EVEN weeks starting in teaching week 4.
- **Computer practicals**: Tuesdays at 12 (Herschel full PC cluster). These happen *occasionally*, and you will be notified of these well in advance.

Timetable and Administrative arrangements

- **Lectures** are on Mondays at 3 (**Bedson LT1**) and Thursdays at 1 (**Bedson LT1**). In the first two weeks there will be a third lecture on Thursdays at 2 (**Bedson LT1**; in effect meaning we will have a double lecture on a Thursday!).
- **Problems classes** are in ODD teaching weeks starting in teaching week 3. These take place on Thursdays at 2 (**Bedson LT1**).
- **Drop–in sessions** are at the same time and place in EVEN weeks starting in teaching week 4.
- **Computer practicals**: Tuesdays at 12 (Herschel full PC cluster). These happen *occasionally*, and you will be notified of these well in advance.

So to summarise:

What?	**When?**		**Where?**	**How often?**
Lecture	Monday	3–4	Bedson: LT1	Every week
	Thursday	1–2	Bedson: LT1	Every week
Problems class/ Drop–in session	Thursday	2–3	Bedson: LT1	Every week ha or the other
Computer practical	Tuesday	12–1	Herschel: PC cluster	Occasionally
Office hours	~~Monday~~ Tuesday	~~2–3~~ 12–1	My office	Every week
	Wednesday	1–2		

Assessment

Assessment is by:

- End of semester exam in May/June (**85%**)
- In course assessment, including written solution to problems and computer practical work (**5%**)
- Mid-semester test (**10%**) (~~Monday 9th March 3pm~~ Monday 16th March 3pm)
- No CBAs!

There will be two assignments, each one having a mixture of written and practical questions taken from Chapter 5 of the notes.

Solutions to the assignment questions should be submitted by the dates given in your notes.

We will work through the unassessed questions from Chapter 5 in problems classes and practicals.

Assessment

Assessment is by:

- End of semester exam in May/June (**85%**)
- In course assessment, including written solution to problems and computer practical work (**5%**)
- Mid-semester test (**10%**) (~~Monday 9th March 3pm~~ Monday 16th March 3pm)
- No CBAs!

There will be two assignments, each one having a mixture of written and practical questions taken from Chapter 5 of the notes.

Solutions to the assignment questions should be submitted by the dates given in your notes.

We will work through the unassessed questions from Chapter 5 in problems classes and practicals.

Assessment

Assessment is by:

- End of semester exam in May/June (**85%**)
- In course assessment, including written solution to problems and computer practical work (**5%**)
- Mid-semester test (**10%**) (~~Monday 9th March 3pm~~ Monday 16th March 3pm)
- No CBAs!

There will be two assignments, each one having a mixture of written and practical questions taken from Chapter 5 of the notes.

Solutions to the assignment questions should be submitted by the dates given in your notes.

We will work through the unassessed questions from Chapter 5 in problems classes and practicals.

Assessment

Assessment is by:

- End of semester exam in May/June (**85%**)
- In course assessment, including written solution to problems and computer practical work (**5%**)
- Mid-semester test (**10%**) (~~Monday 9th March 3pm~~ Monday 16th March 3pm)
- No CBAs!

There will be two assignments, each one having a mixture of written and practical questions taken from Chapter 5 of the notes.

Solutions to the assignment questions should be submitted by the dates given in your notes.

We will work through the unassessed questions from Chapter 5 in problems classes and practicals.

Assessment

Assessment is by:

- End of semester exam in May/June (**85%**)
- In course assessment, including written solution to problems and computer practical work (**5%**)
- Mid-semester test (**10%**) (~~Monday 9th March 3pm~~ Monday 16th March 3pm)
- No CBAs!

There will be two assignments, each one having a mixture of written and practical questions taken from Chapter 5 of the notes.

Solutions to the assignment questions should be submitted by the dates given in your notes.

We will work through the unassessed questions from Chapter 5 in problems classes and practicals.

Recommended textbooks

- "*Bayes' Rule: A Tutorial Introduction to Bayesian Analysis*" - James Stone
- "*Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*" - John Krushke
- "*Bayesian Statistics: An Introduction*" - Peter Lee

"Bayes' rule" is a good introduction to the main concepts in Bayesian statistics but doesn't cover everything in this course. The other books are broader references which go well beyond the contents of this course.

Other stuff

- Notes (with gaps) will be handed out in lectures – you should fill in the gaps during lectures.
- A summarised version of the notes will be used in lectures as slides.
  - **Listen and learn**
  - **Write down**
  - **Announcements**
- These notes and slides will be posted on the **course website** and/or BlackBoard after each topic is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

Other stuff

- Notes (with gaps) will be handed out in lectures – you should fill in the gaps during lectures.
- A summarised version of the notes will be used in lectures as slides.
  - **Listen and learn**
  - **Write down**
  - **Announcements**
- These notes and slides will be posted on the **course website** and/or BlackBoard after each topic is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

Other stuff

- Notes (with gaps) will be handed out in lectures – you should fill in the gaps during lectures.
- A summarised version of the notes will be used in lectures as slides.
  - **Listen and learn**
  - **Write down**
  - **Announcements**
- These notes and slides will be posted on the **course website** and/or BlackBoard after each topic is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

MAS2602: Feedback on your feedback

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

MAS2602: Feedback on your feedback

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

MAS2602: Feedback on your feedback

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

MAS2602: Feedback on your feedback

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

MAS2602: Feedback on your feedback

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

MAS2602: Feedback on your feedback

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

MAS2602: Feedback on your feedback

- **Open-ended comments**
  - Lee: Lectures felt rushed at times :-(
  - Lee: Speak more slowly and less Jordy ??
  - More Chris, less Lee :-(
  - Lee: More computing, less Stats
  - Lee: Polish your R-side so it matches Chris's side
  - Lee: Need more R prep/revision at the start
  - Lee: Wired guy
  - Chris > Lee, tbh

Preface to lecture notes:

What is Bayesian Statistics?

Statistics: an alternative approach

- Up until now, you have been taught Probability and Statistics according to a particular way of thinking
- The **Bayesian paradigm** offers another way of seeing things
- Your ideas about Probability and Statistics are deeply entrenched...perhaps so much so that at first, *Bayesian Statistics* might take you a while to grasp!
- Not because it's hard! Just because you have become so conditioned to think in a particular way.
- We want to broaden your horizons a bit!

Statistics: an alternative approach

- Up until now, you have been taught Probability and Statistics according to a particular way of thinking
- The **Bayesian paradigm** offers another way of seeing things
- Your ideas about Probability and Statistics are deeply entrenched...perhaps so much so that at first, *Bayesian Statistics* might take you a while to grasp!
- Not because it's hard! Just because you have become so conditioned to think in a particular way.
- We want to broaden your horizons a bit!

Statistics: an alternative approach

- Up until now, you have been taught Probability and Statistics according to a particular way of thinking
- The **Bayesian paradigm** offers another way of seeing things
- Your ideas about Probability and Statistics are deeply entrenched...perhaps so much so that at first, *Bayesian Statistics* might take you a while to grasp!
- Not because it's hard! Just because you have become so conditioned to think in a particular way.
- We want to broaden your horizons a bit!

Statistics: an alternative approach

There are two main approaches to statistics: **frequentist** (or classical) statistics and **Bayesian** statistics.

All of the statistics teaching you've encountered so far is likely to be about frequentist methods. Bayesian methods are substantially different and can feel quite strange to start with.

So, before starting the main course material, this **non-examinable** preface explains the concepts behind Bayesian statistics and how they differ from frequentist approaches.

Statistics: an alternative approach

One way of defining Statistics is as a way to learn about the world from some data which is subject to random variation.

For example, in **climate science** we want to learn about the climate given some imperfect measurements.

Some statistical questions that we might ask in this field are:

- What is our best estimate of the world temperature this year? This is a point estimation problem.
- What is a plausible range of values? This is an interval estimation (or uncertainty quantification) problem.
- What will the temperature be in 100 years? This is a prediction problem.
- Is the climate warming? This is a hypothesis testing problem.

Statistics: an alternative approach

One way of defining Statistics is as a way to learn about the world from some data which is subject to random variation.

For example, in **climate science** we want to learn about the climate given some imperfect measurements.

Some statistical questions that we might ask in this field are:

- What is our best estimate of the world temperature this year? This is a point estimation problem.
- What is a plausible range of values? This is an interval estimation (or uncertainty quantification) problem.
- What will the temperature be in 100 years?This is a prediction problem.
- Is the climate warming?This is a hypothesis testing problem.

Statistics: an alternative approach

One way of defining Statistics is as a way to learn about the world from some data which is subject to random variation.

For example, in **climate science** we want to learn about the climate given some imperfect measurements.

Some statistical questions that we might ask in this field are:

- What is our best estimate of the world temperature this year? This is a point estimation problem.
- What is a plausible range of values? This is an interval estimation (or uncertainty quantification) problem.
- What will the temperature be in 100 years?This is a prediction problem.
- Is the climate warming?This is a hypothesis testing problem.

Statistics: an alternative approach

One way of defining Statistics is as a way to learn about the world from some data which is subject to random variation.

For example, in **climate science** we want to learn about the climate given some imperfect measurements.

Some statistical questions that we might ask in this field are:

- What is our best estimate of the world temperature this year? This is a point estimation problem.
- What is a plausible range of values? This is an interval estimation (or uncertainty quantification) problem.
- What will the temperature be in 100 years?This is a prediction problem.
- Is the climate warming?This is a hypothesis testing problem.

Statistics: an alternative approach

One way of defining Statistics is as a way to learn about the world from some data which is subject to random variation.

For example, in **climate science** we want to learn about the climate given some imperfect measurements.

Some statistical questions that we might ask in this field are:

- What is our best estimate of the world temperature this year? This is a <u>point estimation</u> problem.
- What is a plausible range of values? This is an <u>interval estimation</u> (or uncertainty quantification) problem.
- What will the temperature be in 100 years? This is a <u>prediction</u> problem.
- Is the climate warming? This is a <u>hypothesis testing</u> problem.

Statistics: an alternative approach

One way of defining Statistics is as a way to learn about the world from some data which is subject to random variation.

For example, in **climate science** we want to learn about the climate given some imperfect measurements.

Some statistical questions that we might ask in this field are:

- What is our best estimate of the world temperature this year? This is a point estimation problem.
- What is a plausible range of values? This is an interval estimation (or uncertainty quantification) problem.
- What will the temperature be in 100 years?This is a prediction problem.
- Is the climate warming?This is a hypothesis testing problem.

Statistics: an alternative approach

One way of defining Statistics is as a way to learn about the world from some data which is subject to random variation.

For example, in **climate science** we want to learn about the climate given some imperfect measurements.

Some statistical questions that we might ask in this field are:

- What is our best estimate of the world temperature this year? This is a <u>point estimation</u> problem.
- What is a plausible range of values? This is an <u>interval estimation</u> (or uncertainty quantification) problem.
- What will the temperature be in 100 years? This is a <u>prediction</u> problem.
- Is the climate warming? This is a <u>hypothesis testing</u> problem.

Statistics: an alternative approach

One way of defining Statistics is as a way to learn about the world from some data which is subject to random variation.

For example, in **climate science** we want to learn about the climate given some imperfect measurements.

Some statistical questions that we might ask in this field are:

- What is our best estimate of the world temperature this year? This is a point estimation problem.
- What is a plausible range of values? This is an interval estimation (or uncertainty quantification) problem.
- What will the temperature be in 100 years?This is a prediction problem.
- Is the climate warming?This is a hypothesis testing problem.

Statistics: an alternative approach

Frequentist and Bayesian methods address all these types of problem: point estimation, interval estimation, prediction and hypothesis testing.

But they use different approaches to do so.

Some typical frequentist approaches include:

- Least squares (point estimation)
- Maximum likelihood (point estimation)
- Confidence intervals (interval estimation)
- Test statistics and *p*-values (hypothesis testing)

Bayesian Statistics doesn't use any of these familiar methods!

Bayes goes to war!

- The work of Thomas Bayes (see later) was published in 1764, 3 years after his death
- Over the course of the next 100–150 years, it received little attention
- In fact, some key figures in Statistics – e.g. **R.A. Fisher** – outrightly rejected the idea of Bayesian statistics
- By the start of WW2, Bayes' rule was **virtually taboo** in the world of Statistics!
- During WW2, some of the world's leading mathematicians resurrected Bayes' rule in deepest secrecy to crack the coded messages of the Germans

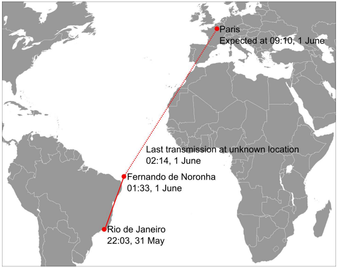
Bayes goes to war!

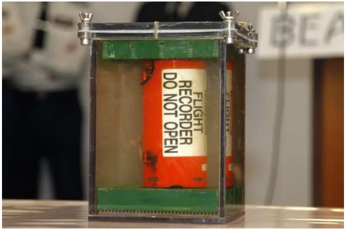
- **Alan Turing** – mathematician working at Bletchley Park
- Designed the ‘bombe’ – an electro–mechanical machine for testing every possible permutation of a message produced on the Enigma machine — could take up to **4 days** to decode a message
- New system: **Banburismus** – where Bayesian methods were used to quantify the belief in guesses of a stretch of letters in an Enigma message
- Certain permutations – unlikely to be the original message – were ‘thrown out’ before they were even tested
- Greatly reduced the time it took to crack Enigma codes

Bayes goes to war!

- **Alan Turing** – mathematician working at Bletchley Park
- Designed the 'bombe' – an electro–mechanical machine for testing every possible permutation of a message produced on the Enigma machine — could take up to **4 days** to decode a message
- New system: **Banburismus** – where Bayesian methods were used to quantify the belief in guesses of a stretch of letters in an Enigma message
- Certain permutations – unlikely to be the original message – were 'thrown out' before they were even tested
- Greatly reduced the time it took to crack Enigma codes

Air France 447





Air France 447

- Black boxes could be anywhere within an area of the South Atlantic the size of Switzerland (**6,500 square miles**)
- Mid–Atlantic ridge – between two tectonic plates – just as Mountainous as Switzerland!
- So remote – scientists have not yet charted the sea–bed
- Search method: sonar – detectors emitting sound waves which would bounce back once they hit something
- After two years of meticulously searching an area north of the plane's trajectory (after analysing debris drift) $\longrightarrow$ **nothing**

Air France 447

- April 2011: *Metron, Inc.*, of Retson, Virginia, hired to launch a **Bayesian review** of the entire search effort
- Included in the analysis were:
  - Data from 9 previous airline crashes involving loss of pilot control – reduced search area to **1,600 square miles**
  - **Expert opinions** on the credibility of the flight data
  - **Expert opinions** about whether or not the black box 'pingers' might have become damaged on impact
  - Positions/recovery times of bodies found drifting – **expert opinions** assigned to the reliability of this data because of the turbulent equatorial waters
  - **Expert information** from oceanographers: Sea state, visibility, underwater geography,...
- All information combined through **Bayes Theorem**: After one week, black boxes found!

Air France 447

- April 2011: *Metron, Inc.*, of Retson, Virginia, hired to launch a **Bayesian review** of the entire search effort
- Included in the analysis were:
  - Data from 9 previous airline crashes involving loss of pilot control – reduced search area to **1,600 square miles**
  - **Expert opinions** on the credibility of the flight data
  - **Expert opinions** about whether or not the black box 'pingers' might have become damaged on impact
  - Positions/recovery times of bodies found drifting – **expert opinions** assigned to the reliability of this data because of the turbulent equatorial waters
  - **Expert information** from oceanographers: Sea state, visibility, underwater geography,...
- All information combined through **Bayes Theorem**: After one week, black boxes found!

Air France 447

- April 2011: *Metron, Inc.*, of Retson, Virginia, hired to launch a **Bayesian review** of the entire search effort
- Included in the analysis were:
  - Data from 9 previous airline crashes involving loss of pilot control – reduced search area to **1,600 square miles**
  - **Expert opinions** on the credibility of the flight data
  - **Expert opinions** about whether or not the black box 'pingers' might have become damaged on impact
  - Positions/recovery times of bodies found drifting – **expert opinions** assigned to the reliability of this data because of the turbulent equatorial waters
  - **Expert information** from oceanographers: Sea state, visibility, underwater geography,...
- All information combined through **Bayes Theorem**: After one week, black boxes found!

Air France 447

- April 2011: *Metron, Inc.*, of Retson, Virginia, hired to launch a **Bayesian review** of the entire search effort
- Included in the analysis were:
  - Data from 9 previous airline crashes involving loss of pilot control – reduced search area to **1,600 square miles**
  - **Expert opinions** on the credibility of the flight data
  - **Expert opinions** about whether or not the black box ‘pingers’ might have become damaged on impact
  - Positions/recovery times of bodies found drifting – **expert opinions** assigned to the reliability of this data because of the turbulent equatorial waters
  - **Expert information** from oceanographers: Sea state, visibility, underwater geography,...
- All information combined through **Bayes Theorem**: After one week, black boxes found!

Air France 447

- April 2011: *Metron, Inc.*, of Retson, Virginia, hired to launch a **Bayesian review** of the entire search effort
- Included in the analysis were:
  - Data from 9 previous airline crashes involving loss of pilot control – reduced search area to **1,600 square miles**
  - **Expert opinions** on the credibility of the flight data
  - **Expert opinions** about whether or not the black box 'pingers' might have become damaged on impact
  - Positions/recovery times of bodies found drifting – **expert opinions** assigned to the reliability of this data because of the turbulent equatorial waters
  - **Expert information** from oceanographers: Sea state, visibility, underwater geography,...
- All information combined through **Bayes Theorem**: After one week, black boxes found!

Air France 447

- April 2011: *Metron, Inc.*, of Retson, Virginia, hired to launch a **Bayesian review** of the entire search effort
- Included in the analysis were:
  - Data from 9 previous airline crashes involving loss of pilot control – reduced search area to **1,600 square miles**
  - **Expert opinions** on the credibility of the flight data
  - **Expert opinions** about whether or not the black box ‘pingers’ might have become damaged on impact
  - Positions/recovery times of bodies found drifting – **expert opinions** assigned to the reliability of this data because of the turbulent equatorial waters
  - **Expert information** from oceanographers: Sea state, visibility, underwater geography,...
- All information combined through **Bayes Theorem**: After one week, black boxes found!

Air France 447

- April 2011: *Metron, Inc.*, of Retson, Virginia, hired to launch a **Bayesian review** of the entire search effort
- Included in the analysis were:
  - Data from 9 previous airline crashes involving loss of pilot control – reduced search area to **1,600 square miles**
  - **Expert opinions** on the credibility of the flight data
  - **Expert opinions** about whether or not the black box 'pingers' might have become damaged on impact
  - Positions/recovery times of bodies found drifting – **expert opinions** assigned to the reliability of this data because of the turbulent equatorial waters
  - **Expert information** from oceanographers: Sea state, visibility, underwater geography,...
- All information combined through **Bayes Theorem**: After one week, black boxes found!

Body Mass Index

Suppose we are interested in whether or not there is any real difference between the Body Mass Index (BMI) of Maths & Stats students and Food & Human Nutrition students.

BMI is often used as a measure of 'fatness' or 'thinness': it is the ratio of a person's weight (measured in kg) to their squared height (measured in metres).

In a two–sample t–test, we often test the null hypothesis that there is no difference between the population means of the two groups:

$$H_0 : \mu_M = \mu_F,$$

How can we test this hypothesis?

Body Mass Index

Suppose we are interested in whether or not there is any real difference between the Body Mass Index (BMI) of Maths & Stats students and Food & Human Nutrition students.

BMI is often used as a measure of ‘fatness’ or ‘thinness’: it is the ratio of a person’s weight (measured in kg) to their squared height (measured in metres).

In a two–sample t–test, we often test the null hypothesis that there is no difference between the population means of the two groups:

$$H_0 : \mu_M = \mu_F,$$

How can we test this hypothesis?

Body Mass Index

Suppose we are interested in whether or not there is any real difference between the Body Mass Index (BMI) of Maths & Stats students and Food & Human Nutrition students.

BMI is often used as a measure of ‘fatness’ or ‘thinness’: it is the ratio of a person’s weight (measured in kg) to their squared height (measured in metres).

In a two–sample *t*–test, we often test the null hypothesis that there is no difference between the population means of the two groups:

$$H_0 : \mu_M = \mu_F,$$

How can we test this hypothesis?

Body Mass Index

Suppose we are interested in whether or not there is any real difference between the Body Mass Index (BMI) of Maths & Stats students and Food & Human Nutrition students.

BMI is often used as a measure of 'fatness' or 'thinness': it is the ratio of a person's weight (measured in kg) to their squared height (measured in metres).

In a two–sample *t*–test, we often test the null hypothesis that there is no difference between the population means of the two groups:

$$H_0 : \mu_M = \mu_F,$$

How can we test this hypothesis?

Body Mass Index

Suppose we are interested in whether or not there is any real difference between the Body Mass Index (BMI) of Maths & Stats students and Food & Human Nutrition students.

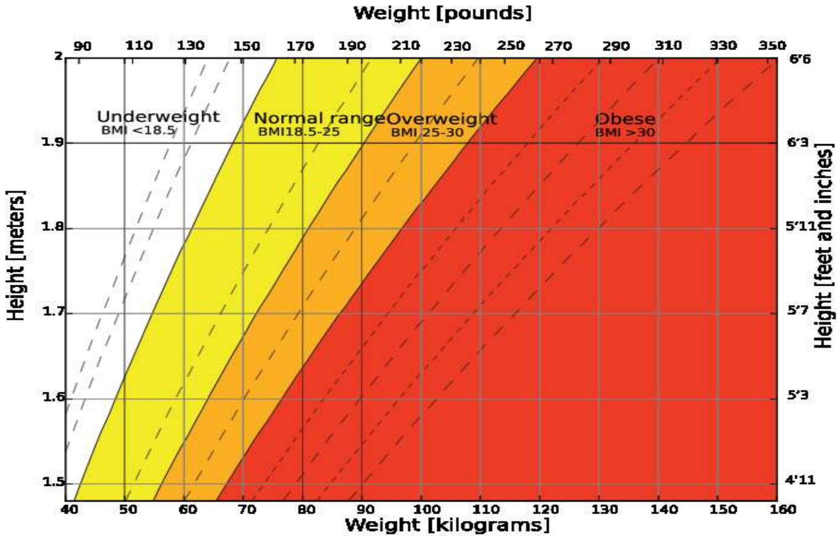
BMI is often used as a measure of ‘fatness’ or ‘thinness’: it is the ratio of a person’s weight (measured in kg) to their squared height (measured in metres).

In a two–sample *t*–test, we often test the null hypothesis that there is no difference between the population means of the two groups:

$$H_0 : \mu_M = \mu_F,$$

How can we test this hypothesis?

Body Mass Index



Suppose we obtain the following BMI values for a random sample of students from both groups:

Maths & Stats	23.1	28.3	35.3	24.0	32.4	30.5	30.1	21.9	22.1
	29.9	27.2	33.2						
Food & Human Nutrition	25.5	27.3	21.0	23.3	19.1	22.2	29.5	27.5	23.2

From this we can obtain the summaries:

$$n_M = 12 \qquad \bar{x}_M = 28.17 \qquad s_M = 4.54$$

$$n_F = 9 \qquad \bar{x}_F = 24.29 \qquad s_F = 3.40$$

Suppose we obtain the following BMI values for a random sample of students from both groups:

Maths & Stats	23.1	28.3	35.3	24.0	32.4	30.5	30.1	21.9	22.1
	29.9	27.2	33.2						
Food & Human Nutrition	25.5	27.3	21.0	23.3	19.1	22.2	29.5	27.5	23.2

From this we can obtain the summaries:

$$n_M = 12 \qquad \bar{x}_M = 28.17 \qquad s_M = 4.54$$

$$n_F = 9 \qquad \bar{x}_F = 24.29 \qquad s_F = 3.40$$

Body Mass Index

Suppose we obtain the following BMI values for a random sample of students from both groups:

Maths & Stats	23.1	28.3	35.3	24.0	32.4	30.5	30.1	21.9	22.1
	29.9	27.2	33.2						
Food & Human Nutrition	25.5	27.3	21.0	23.3	19.1	22.2	29.5	27.5	23.2

From this we can obtain the summaries:

$$n_M = 12 \qquad \bar{x}_M = 28.17 \qquad s_M = 4.54$$

$$n_F = 9 \qquad \bar{x}_F = 24.29 \qquad s_F = 3.40$$

Body Mass Index

The pooled standard deviation is

$$
\begin{aligned}
s &= \sqrt{\frac{(n_M - 1)s_M^2 + (n_F - 1)s_F^2}{n_M + n_F - 2}} \\
&= \sqrt{\frac{11 \times 4.54^2 + 8 \times 3.4^2}{12 + 9 - 2}} \\
&= 4.099
\end{aligned}
$$

Then the test statistic is

$$
\begin{aligned}
t &= \frac{|\bar{x}_M - \bar{x}_F|}{s \times \sqrt{\frac{1}{n_M} + \frac{1}{n_F}}} \\
&= 2.147.
\end{aligned}
$$

Body Mass Index

The pooled standard deviation is

$$s = \sqrt{\frac{(n_M - 1)s_M^2 + (n_F - 1)s_F^2}{n_M + n_F - 2}}$$

$$= \sqrt{\frac{11 \times 4.54^2 + 8 \times 3.4^2}{12 + 9 - 2}}$$

$$= 4.099$$

Then the test statistic is

$$t = \frac{|\bar{x}_M - \bar{x}_F|}{s \times \sqrt{\frac{1}{n_M} + \frac{1}{n_F}}}$$

$$= 2.147.$$

Body Mass Index

The pooled standard deviation is

$$s = \sqrt{\frac{(n_M - 1)s_M^2 + (n_F - 1)s_F^2}{n_M + n_F - 2}}$$

$$= \sqrt{\frac{11 \times 4.54^2 + 8 \times 3.4^2}{12 + 9 - 2}}$$

$$= 4.099$$

Then the test statistic is

$$t = \frac{|\bar{x}_M - \bar{x}_F|}{s \times \sqrt{\frac{1}{n_M} + \frac{1}{n_F}}}$$

$$= 2.147.$$

Body Mass Index

The pooled standard deviation is

$$
\begin{aligned}
s &= \sqrt{\frac{(n_M - 1)s_M^2 + (n_F - 1)s_F^2}{n_M + n_F - 2}} \\
&= \sqrt{\frac{11 \times 4.54^2 + 8 \times 3.4^2}{12 + 9 - 2}} \\
&= 4.099
\end{aligned}
$$

Then the test statistic is

$$
\begin{aligned}
t &= \frac{|\bar{x}_M - \bar{x}_F|}{s \times \sqrt{\frac{1}{n_M} + \frac{1}{n_F}}} \\
&= 2.147.
\end{aligned}
$$

Body Mass Index

The pooled standard deviation is

$$
\begin{aligned}
s &= \sqrt{\frac{(n_M - 1)s_M^2 + (n_F - 1)s_F^2}{n_M + n_F - 2}} \\
&= \sqrt{\frac{11 \times 4.54^2 + 8 \times 3.4^2}{12 + 9 - 2}} \\
&= 4.099
\end{aligned}
$$

Then the test statistic is

$$
\begin{aligned}
t &= \frac{|\bar{x}_M - \bar{x}_F|}{s \times \sqrt{\frac{1}{n_M} + \frac{1}{n_F}}} \\
&= 2.147.
\end{aligned}
$$

Body Mass Index

The pooled standard deviation is

$$
\begin{aligned}
s &= \sqrt{\frac{(n_M - 1)s_M^2 + (n_F - 1)s_F^2}{n_M + n_F - 2}} \\
&= \sqrt{\frac{11 \times 4.54^2 + 8 \times 3.4^2}{12 + 9 - 2}} \\
&= 4.099
\end{aligned}
$$

Then the test statistic is

$$
\begin{aligned}
t &= \frac{|\bar{x}_M - \bar{x}_F|}{s \times \sqrt{\frac{1}{n_M} + \frac{1}{n_F}}} \\
&= 2.147.
\end{aligned}
$$

The pooled standard deviation is

$$
\begin{aligned}
s &= \sqrt{\frac{(n_M - 1)s_M^2 + (n_F - 1)s_F^2}{n_M + n_F - 2}} \\
&= \sqrt{\frac{11 \times 4.54^2 + 8 \times 3.4^2}{12 + 9 - 2}} \\
&= 4.099
\end{aligned}
$$

Then the test statistic is

$$
\begin{aligned}
t &= \frac{|\bar{x}_M - \bar{x}_F|}{s \times \sqrt{\frac{1}{n_M} + \frac{1}{n_F}}} \\
&= 2.147.
\end{aligned}
$$

Body Mass Index

From t_{19}–tables, we get:

p–value	10%	5%	1%
critical value	1.729	2.093	2.861

Our test statistic $t = 2.147$ indicates that our p–value is less than 5%:

- We have a significant result at the 5% level of significance
- We **reject** H_0
- We conclude that there *is* a significant difference in BMI between M&S and FHN students

Body Mass Index

From t_{19}–tables, we get:

p–value	10%	5%	1%
critical value	1.729	2.093	2.861

Our test statistic $t = 2.147$ indicates that our p–value is less than 5%:

- We have a significant result at the 5% level of significance
- We **reject** H_0
- We conclude that there *is* a significant difference in BMI between M&S and FHN students

Body Mass Index

From t_{19}–tables, we get:

p–value	10%	5%	1%
critical value	1.729	2.093	2.861

Our test statistic $t = 2.147$ indicates that our p–value is less than 5%:

- We have a significant result at the 5% level of significance
- We **reject** H_0
- We conclude that there *is* a significant difference in BMI between M&S and FHN students

Body Mass Index

From t_{19}–tables, we get:

p–value	10%	5%	1%
critical value	1.729	2.093	2.861

Our test statistic $t = 2.147$ indicates that our p–value is less than 5%:

- We have a significant result at the 5% level of significance
- We **reject** H_0
- We conclude that there *is* a significant difference in BMI between M&S and FHN students

From t_{19}–tables, we get:

p–value	10%	5%	1%
critical value	1.729	2.093	2.861

Our test statistic $t = 2.147$ indicates that our p–value is less than 5%:

- We have a significant result at the 5% level of significance
- We **reject** H_0
- We conclude that there *is* a significant difference in BMI between M&S and FHN students

From t_{19}–tables, we get:

p–value	10%	5%	1%
critical value	1.729	2.093	2.861

Our test statistic $t = 2.147$ indicates that our p–value is less than 5%:

- We have a significant result at the 5% level of significance
- We **reject** H_0
- We conclude that there *is* a significant difference in BMI between M&S and FHN students

Body Mass Index

Notice from R that the *exact* p–value is 0.004488 = 4.488%.

- As we have already concluded, we would **reject** H_0 at the 5% level;
- However, if we work at the 1% level of significance, we would **retain** H_0!

What are we to do? Convention tells us to work at the 5% level, but why? Some practitioners work at the 1% level!

Body Mass Index

Notice from R that the *exact* p–value is 0.004488 = 4.488%.

- As we have already concluded, we would **reject** H_0 at the 5% level;
- However, if we work at the 1% level of significance, we would **retain** H_0!

What are we to do? Convention tells us to work at the 5% level, but why? Some practitioners work at the 1% level!

Body Mass Index

Notice from R that the *exact* p–value is 0.004488 = 4.488%.

- As we have already concluded, we would **reject** H_0 at the 5% level;
- However, if we work at the 1% level of significance, we would **retain** H_0!

What are we to do? Convention tells us to work at the 5% level, but why? Some practitioners work at the 1% level!

In MAS1604 you were introduced to **Bayes' Theorem**.

As we shall see, Bayes' Theorem gives us a way of combining subjective assessments with observed data.

For example, what if – prior to observing the data – we believed that Food & Human Nutrition students were likely to have a **considerably lower BMI** than Maths & Stats students?

What if, from a previous study, we knew that Maths & Stats students were quite likely to have a **BMI somewhere between 25 and 33**?

Body Mass Index

Would it not be sensible to build this information into our analysis before forming our conclusions using the data alone?

This could be a good idea – we have relatively small samples, which could be biased!

A **Bayesian analysis** can do this!

Some people argue against this on the grounds that it is subjective... But as we have just seen, the usual approach to hypothesis testing is also subjective!

Not only are hypothesis tests tricky to interpret... but so are **confidence intervals**.

Would it not be sensible to build this information into our analysis before forming our conclusions using the data alone?

This could be a good idea – we have relatively small samples, which could be biased!

A **Bayesian analysis** can do this!

Some people argue against this on the grounds that it is subjective... But as we have just seen, the usual approach to hypothesis testing is also subjective!

Not only are hypothesis tests tricky to interpret... but so are **confidence intervals**.

Would it not be sensible to build this information into our analysis before forming our conclusions using the data alone?

This could be a good idea – we have relatively small samples, which could be biased!

A **Bayesian analysis** can do this!

Some people argue against this on the grounds that it is subjective... But as we have just seen, the usual approach to hypothesis testing is also subjective!

Not only are hypothesis tests tricky to interpret... but so are **confidence intervals**.

Would it not be sensible to build this information into our analysis before forming our conclusions using the data alone?

This could be a good idea – we have relatively small samples, which could be biased!

A **Bayesian analysis** can do this!

Some people argue against this on the grounds that it is subjective... But as we have just seen, the usual approach to hypothesis testing is also subjective!

Not only are hypothesis tests tricky to interpret... but so are **confidence intervals**.

Body Mass Index

Obtain a 95% confidence interval for the population mean BMI for Maths & Stats students.

We use

$$\bar{x}_M \pm t\frac{s_M}{\sqrt{n_M}}, \qquad \text{giving}$$

$$28.17 \quad \pm \quad 2.201 \times \frac{4.54}{\sqrt{12}} \qquad \text{i.e.}$$

$$28.17 \quad \pm \quad 2.743,$$

giving (25.28, 31.05).

Body Mass Index

Obtain a 95% confidence interval for the population mean BMI for Maths & Stats students.

We use

$$\bar{x}_M \pm t\frac{s_M}{\sqrt{n_M}}, \qquad \text{giving}$$

$$28.17 \quad \pm \quad 2.201 \times \frac{4.54}{\sqrt{12}} \qquad \text{i.e.}$$

$$28.17 \quad \pm \quad 2.743,$$

giving (25.28, 31.05).

Body Mass Index

Obtain a 95% confidence interval for the population mean BMI for Maths & Stats students.

We use

$$\bar{x}_M \pm t\frac{s_M}{\sqrt{n_M}}, \qquad \text{giving}$$

$$28.17 \quad \pm \quad 2.201 \times \frac{4.54}{\sqrt{12}} \qquad \text{i.e.}$$

$$28.17 \quad \pm \quad 2.743,$$

giving (25.28, 31.05).

Body Mass Index

Obtain a 95% confidence interval for the population mean BMI for Maths & Stats students.

We use

$$\bar{x}_M \pm t\frac{s_M}{\sqrt{n_M}}, \qquad \text{giving}$$

$$28.17 \quad \pm \quad 2.201 \times \frac{4.54}{\sqrt{12}} \qquad \text{i.e.}$$

$$28.17 \quad \pm \quad 2.743,$$

giving (25.28, 31.05).

Body Mass Index

What is $\Pr(25.28 < \mu_M < 31.05)$?

NOT 0.95!! In the frequentist approach, population parameters (in this case μ_M) are NOT random variables!

This means μ_M is a fixed (but unknown) quantity. So it's either *in* the interval, or it's *not*, i.e.

$$\Pr(25.28 < \mu_M < 31.05) = 1 \qquad \text{or}$$

$$\Pr(25.28 < \mu_M < 31.05) = 0.$$

Nothing else!

Equivalent intervals in the **Bayesian framework** have more natural interpretations!

Body Mass Index

What is $\Pr(25.28 < \mu_M < 31.05)$?

NOT 0.95!! In the frequentist approach, population parameters (in this case μ_M) are NOT random variables!

This means μ_M is a fixed (but unknown) quantity. So it's either *in* the interval, or it's *not*, i.e.

$$\Pr(25.28 < \mu_M < 31.05) = 1 \qquad \text{or}$$

$$\Pr(25.28 < \mu_M < 31.05) = 0.$$

Nothing else!

Equivalent intervals in the **Bayesian framework** have more natural interpretations!

Body Mass Index

What is $\Pr(25.28 < \mu_M < 31.05)$?

NOT 0.95!! In the frequentist approach, population parameters (in this case μ_M) are NOT random variables!

This means μ_M is a fixed (but unknown) quantity. So it's either *in* the interval, or it's *not*, i.e.

$$\Pr(25.28 < \mu_M < 31.05) = 1 \qquad \text{or}$$

$$\Pr(25.28 < \mu_M < 31.05) = 0.$$

Nothing else!

Equivalent intervals in the **Bayesian framework** have more natural interpretations!

Body Mass Index

What is $\Pr(25.28 < \mu_M < 31.05)$?

NOT 0.95!! In the frequentist approach, population parameters (in this case μ_M) are NOT random variables!

This means μ_M is a fixed (but unknown) quantity. So it's either *in* the interval, or it's *not*, i.e.

$$\Pr(25.28 < \mu_M < 31.05) = 1 \qquad \text{or}$$

$$\Pr(25.28 < \mu_M < 31.05) = 0.$$

Nothing else!

Equivalent intervals in the **Bayesian framework** have more natural interpretations!

Body Mass Index

What is $\Pr(25.28 < \mu_M < 31.05)$?

NOT 0.95!! In the frequentist approach, population parameters (in this case μ_M) are NOT random variables!

This means μ_M is a fixed (but unknown) quantity. So it's either *in* the interval, or it's *not*, i.e.

$$\Pr(25.28 < \mu_M < 31.05) = 1 \qquad \text{or}$$

$$\Pr(25.28 < \mu_M < 31.05) = 0.$$

Nothing else!

Equivalent intervals in the **Bayesian framework** have more natural interpretations!

Frequentist/Subjectivist debate

One of the main arguments *against* working within the Bayesian paradigm is that it is **subjective**.

Although we have not yet thought about *how* the Bayesian framework combines subjective assessments with the data, we have said that this is what it does.

Surely we should strive to be as *objective* as we can?

Actually, to be able to incorporate a person's beliefs into an analysis of sample data is surely the **right thing to do** — especially if that person is an expert!

This is in line with how scientific experiments are carried out:

- The experimenter usually **knows something**;
- She then carries out the experiment in which **data are collected**;
- the experimenter then **updates her beliefs** from these results.

Frequentist/Subjectivist debate

Actually, to be able to incorporate a person's beliefs into an analysis of sample data is surely the **right thing to do** — especially if that person is an expert!

This is in line with how scientific experiments are carried out:

- The experimenter usually **knows something**;
- She then carries out the experiment in which **data are collected**;
- the experimenter then **updates her beliefs** from these results.

Frequentist/Subjectivist debate

Actually, to be able to incorporate a person's beliefs into an analysis of sample data is surely the **right thing to do** — especially if that person is an expert!

This is in line with how scientific experiments are carried out:

- The experimenter usually **knows something**;
- She then carries out the experiment in which **data are collected**;
- the experimenter then **updates her beliefs** from these results.

Frequentist/Subjectivist debate

Actually, to be able to incorporate a person's beliefs into an analysis of sample data is surely the **right thing to do** — especially if that person is an expert!

This is in line with how scientific experiments are carried out:

- The experimenter usually **knows something**;
- She then carries out the experiment in which **data are collected**;
- the experimenter then **updates her beliefs** from these results.

In other words, the data are used to refine what the experimenter knows.

And that is what MAS2903 is all about – we will consider how

- **prior beliefs** can be combined with
- **experimental data** to form
- **posterior beliefs** – which include both the prior knowledge and what we have learnt from the data

In other words, the data are used to refine what the experimenter knows.

And that is what MAS2903 is all about – we will consider how

- **prior beliefs** can be combined with
- **experimental data** to form
- **posterior beliefs** – which include both the prior knowledge and what we have learnt from the data

In other words, the data are used to refine what the experimenter knows.

And that is what MAS2903 is all about – we will consider how

- **prior beliefs** can be combined with
- **experimental data** to form
- **posterior beliefs** – which include both the prior knowledge and what we have learnt from the data

In other words, the data are used to refine what the experimenter knows.

And that is what MAS2903 is all about – we will consider how

- **prior beliefs** can be combined with
- **experimental data** to form
- **posterior beliefs** – which include both the prior knowledge and what we have learnt from the data

In other words, the data are used to refine what the experimenter knows.

And that is what MAS2903 is all about – we will consider how

- **prior beliefs** can be combined with
- **experimental data** to form
- **posterior beliefs** – which include both the prior knowledge and what we have learnt from the data

Thomas Bayes (1702–1761)

- Presbyterian minister in Tunbridge Wells, Kent
- Bayes' solution to a problem of "inverse probability" was presented in the *Essay towards Solving a Problem in the Doctrine of Chances* (1764)
- Work published posthumously by his friend Richard Price in the *Philosophical Transactions of the Royal Society of London*.
- This work gives us **Bayes' Theorem**, a key result on which most of the work in this course rests.

Bayesian Statistics: leading players

Thomas Bayes (1702–1761)

- Presbyterian minister in Tunbridge Wells, Kent
- Bayes' solution to a problem of "inverse probability" was presented in the *Essay towards Solving a Problem in the Doctrine of Chances* (1764)
- Work published posthumously by his friend Richard Price in the *Philosophical Transactions of the Royal Society of London*.
- This work gives us **Bayes' Theorem**, a key result on which most of the work in this course rests.

Bayesian Statistics: leading players

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Bruno di Finetti (1906–1985)

- Italian mathematical probabilist, developed his ideas on subjective probability in the 1920s.
- Famed for saying

"Probability does not exist"

- By this, he meant it has no *objective* existence – i.e. it is not a feature of the world around us.
- It is a measure of degree–of–belief – your belief, my belief, someone else's belief – all of which could be different.

Bayesian Statistics: leading players

Bruno di Finetti (1906–1985)

- Italian mathematical probabilist, developed his ideas on subjective probability in the 1920s.
- Famed for saying

  "Probability does not exist"

- By this, he meant it has no *objective* existence – i.e. it is not a feature of the world around us.
- It is a measure of degree–of–belief – your belief, my belief, someone else's belief – all of which could be different.

Dennis Lindley (1923—2013)

- Worked hard to find a mathematical basis for the subject of Statistics
- With Leonard Savage, he found a deeper justification for Statistics in **Bayesian theory**
- Turned into a critic of the classical statistical inference he had hoped to justify
- Quoted as saying:

"Uncertainty is a personal matter. It is not the uncertainty, but your uncertainty"

Dennis Lindley (1923—2013)

- Worked hard to find a mathematical basis for the subject of Statistics
- With Leonard Savage, he found a deeper justification for Statistics in **Bayesian theory**
- Turned into a critic of the classical statistical inference he had hoped to justify
- Quoted as saying:

"Uncertainty is a personal matter. It is not the uncertainty, but your uncertainty"

Recent developments

One of the difficulties in early applications of Bayesian methodology was the maths:

Combining prior beliefs with a probability model for the data often resulted in maths that was just **too difficult/impossible to solve by hand**.

Recent developments

Then – in the early 1990s – a technique called "**Markov chain Monte Carlo**" (MCMC for short) was developed.

MCMC is a computer–intensive simulation–based procedure that gets around the problem of hard maths.

If you choose to take **MAS3902: Bayesian Inference** next year, you will be introduced to this technique.

MCMC has **revolutionised the use of Bayesian Statistics**, to the extent that Bayesian data analyses are now routinely used by non–Statisticians in fields as diverse as biology, civil engineering and sociology.

Number of academic staff publishing advanced research using Bayesian methods:

- 1985: **1** (Professor Boys)
- 2018: **Lots!**
  - *Systems Biology:* (Boys, Farrow, Gillespie, Golightly, Heaps, Henderson, Prangle, Wilkinson, Wilson)
  - *Environmental Extremes:* (Fawcett, Walshaw)

Number of academic staff publishing advanced research using Bayesian methods:

- 1985: **1** (Professor Boys)
- 2018: **Lots!**
  - *Systems Biology:* (Boys, Farrow, Gillespie, Golightly, Heaps, Henderson, Prangle, Wilkinson, Wilson)
  - *Environmental Extremes:* (Fawcett, Walshaw)

Number of academic staff publishing advanced research using Bayesian methods:

- 1985: **1** (Professor Boys)
- 2018: **Lots!**
  - *Systems Biology:* (Boys, Farrow, Gillespie, Golightly, Heaps, Henderson, Prangle, Wilkinson, Wilson)
  - *Environmental Extremes:* (Fawcett, Walshaw)

Newcastle's contribution

Number of academic staff publishing advanced research using Bayesian methods:

- 1985: **1** (Professor Boys)
- 2018: **Lots!**
  - *Systems Biology:* (Boys, Farrow, Gillespie, Golightly, Heaps, Henderson, Prangle, Wilkinson, Wilson)
  - *Environmental Extremes:* (Fawcett, Walshaw)

Newcastle's contribution

Number of academic staff publishing advanced research using Bayesian methods:

- 1985: **1** (Professor Boys)
- 2018: **Lots!**
  - *Systems Biology:* (Boys, Farrow, Gillespie, Golightly, Heaps, Henderson, Prangle, Wilkinson, Wilson)
  - *Environmental Extremes:* (Fawcett, Walshaw)

Newcastle's contribution: Environmental Extremes

Use Statistical methods to help plan for environmental extremes:

- **Hurricane–induced waves**: how high should we build a sea–wall to protect New Orleans against another *Hurricane Katrina*?
- **Extreme wind speeds**: how strong should we design buildings so they will not fail in storm–strength wind speeds?
- **Extreme cold spells**: How much fuel should we stockpile to cater for an extremely cold winter?

Newcastle's contribution: Environmental Extremes

Use Statistical methods to help plan for environmental extremes:

- **Hurricane–induced waves**: how high should we build a sea–wall to protect New Orleans against another *Hurricane Katrina*?
- **Extreme wind speeds**: how strong should we design buildings so they will not fail in storm–strength wind speeds?
- **Extreme cold spells**: How much fuel should we stockpile to cater for an extremely cold winter?

Newcastle's contribution: Environmental Extremes

Use Statistical methods to help plan for environmental extremes:

- **Hurricane–induced waves**: how high should we build a sea–wall to protect New Orleans against another *Hurricane Katrina*?
- **Extreme wind speeds**: how strong should we design buildings so they will not fail in storm–strength wind speeds?
- **Extreme cold spells**: How much fuel should we stockpile to cater for an extremely cold winter?

Newcastle's contribution: Environmental Extremes

Use Statistical methods to help plan for environmental extremes:

- **Hurricane–induced waves**: how high should we build a sea–wall to protect New Orleans against another *Hurricane Katrina*?
- **Extreme wind speeds**: how strong should we design buildings so they will not fail in storm–strength wind speeds?
- **Extreme cold spells**: How much fuel should we stockpile to cater for an extremely cold winter?

Newcastle's contribution: Environmental Extremes

We work within the Bayesian framework to combine

- expert information from hydrologists/oceanographers, with
- extreme rainfall data/hurricane–induced sea surge data,

to help estimate how likely an **extreme flooding event is to occur**, or to aid the **design of sea wall defences** in storm–prone regions.

If you're interested, you can see my web–page for more details.

Newcastle's contribution: Environmental Extremes

We work within the Bayesian framework to combine

- expert information from hydrologists/oceanographers, with
- extreme rainfall data/hurricane–induced sea surge data,

to help estimate how likely an **extreme flooding event is to occur**, or to aid the **design of sea wall defences** in storm–prone regions.

If you're interested, you can see my web–page for more details.

Quick quiz

Suppose A, θ and μ_M are constants and Z is a random variable, such that $A = 5$, $\theta = 10$, $\mu_M = 33.5$ and $Z \sim N(0, 1)$.

Write down

1. $\Pr(Z > 0)$
2. $\Pr(-1.96 < Z < 1.96)$
3. $\Pr(2 < A < 4)$
4. $\Pr(7 < \theta < 12)$
5. $\Pr(25.28 < \mu_M < 31.05)$