



Evaluation Methods and Statistics: Investigating human behavior

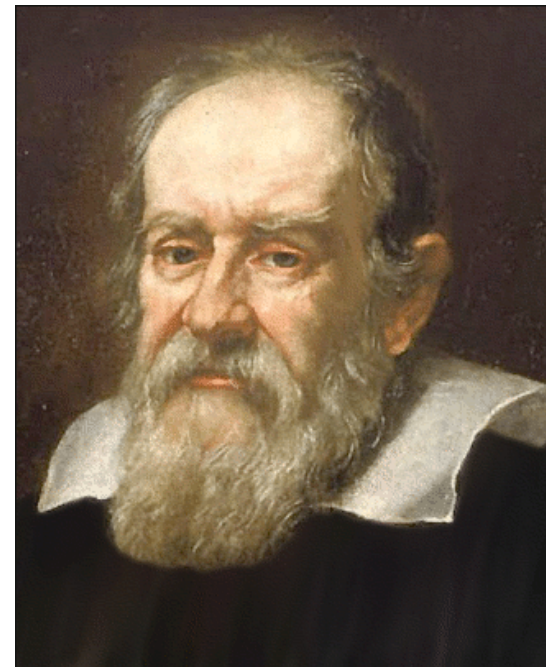
Professor Andrew Howes

Dr Ben Cowan

School of Computer Science
University of Birmingham

evidence versus authority

- "The leitmotif which I recognise in Galileo's work is the passionate fight against any kind of dogma based on authority. Only experience and careful reflection are accepted by him as criteria of truth" – Einstein.
- <http://www.guardian.co.uk/commentisfree/2011/nov/22/neutrino-revolutionary-image>



scientific claims are based on evidence

Here are some headlines from Science Daily (19th Dec 2011).

- Less knowledge, more power: uninformed can be vital to democracy, study finds.
- Traumatic experiences may make you tough.
- New strain of lab mice mimics human alcohol consumption pattern.
- Second-guessing ones decisions leads to unhappiness.

http://www.sciencedaily.com/news/mind_brain/psychology/

claims need evidence

- “E-mail Reveals Your Closest Friends” (Science, 2011)
- “We found that all measures of an ego network structure calculated from the self-reported data correlated significantly ($p < 0.05$, ...) with the email derived, ... networks.” (Wuchty & Uzzi, 2011)
- <http://news.sciencemag.org/sciencenow/2011/11/e-mail-reveals-your-closest-frie.html>

Aims of the module

- The aim of the module is to provide an introduction to the use of empirical scientific methods, including experimental design and statistics, for the purpose of investigating human interaction with computers.
- The module is targeted at computer scientists with an interest in
 - (i) building systems that support human activities (including Human-Computer Interaction),
 - (ii) building computational models of human behaviour, and/or,
 - (iii) understanding human behaviour as an inspiration for computational science (Machine Learning and Artificial Intelligence).

Outcomes

- On successful completion of this module, you should be able to:
- Identify and discuss research methodologies for investigating human behaviour.
- Recognise the appropriateness of statistical techniques in data analysis.
- Conduct and report a variety of statistical tests.
- Interpret research findings from a variety of statistical techniques to a high level.
- Discuss issues related to conducting research on human participants (sampling, recruitment etc).

Prerequisites

- There are no formal prerequisites.
- However, a willingness to engage enthusiastically with the mathematics and computation that underpin statistical reasoning will be essential.
- There will be a strong emphasis on individual and group practical work.

Lecturers

- Professor Andrew Howes
- Email: HowesA@bham.ac.uk
- Dr Ben Cowan
- b.r.cowan@cs.bham.ac.uk

What do we mean by human behavior?

- Perceptual-motor control, e.g. moving a mouse, driving a car, flying an airplane.
- Cognitive tasks, e.g. navigating the web, managing a diary.
- Social and collaborative tasks, e.g. maintaining social relationships, collaborating with colleagues.
- Economic tasks, e.g. managing businesses.

The scientific method

- The progress that has been made in understanding human behavior has been due to the application of quantitative scientific methods.
- Methods are used to test scientific theory.
- Methods specify the following components:
 - Constraints on the design of studies (e.g. control conditions, ethics etc.)
 - Tools for the analysis of data.
 - Conventions for reporting studies.
- The content of this module is largely concerned with these methods.

Course overview

- How to make an evidence-based argument. How to write scientific reports.
- Basic Statistics (mean, mode, standard deviation, central limit theorem, statistical distributions etc.)
- Experimental design (types of error, the Null Hypothesis, independent/dependent variables)
- Correlation (what does it mean for two variables to be correlated? how to calculate a correlation?)
- How to compare distributions (t-test, ANOVA, 2-way ANOVA).
- Questionnaires and interview methods.

Reading

- Discovering Statistics Using R, **Andy Field, Jeremy Miles, Zoe Field**. Sage, 2012.
- Statistical Methods for Psychology, **David C. Howell**. Duxbury, 2006.
- Statistics in R: An introduction using R, **Michael J. Crawley**. Wiley, 2005.
- Discovering statistics using SPSS, **Andy Field**, Wiley, 2009



Lecture I

1. Traditional authority: Science is NOT about trusting what scientists claim.
2. Science isn't about trusting what scientists claim are the facts.
 - quality of method?
 - quality of data.
 - mutated facts.
3. Science is about evidence based argumentation.
4. The form of a scientific argument: An example.



PART I:

Traditional authority: Science is NOT about trusting what scientists claim.

consider these headlines

- "Online networking 'harms health'" (BBC, 2009a)
- "Online risks: from cancer to autism?" (BBC, 2009b)
- "Facebook and Bebo risk 'infantilising' the human mind" from the Guardian (Wintour, 2009).

Professor Baroness Susan Greenfield (Newsnight, 2009)

"... one can look at the features of screen life and see that it is perhaps now mirrored in the behaviour of the upcoming generation if you like. One might argue a shorter attention span, an emphasis on process, on the experience of the moment rather than content, of an identity that needs to be bolstered up with twitter, and perhaps an increased recklessness."

‘infantilised’ mind

- According to Wintour (2009) she told the House of Lords that children's experiences on social networking sites,
- "are devoid of **cohesive narrative** and long-term significance. As a consequence, the mid-21st century mind might almost be infantilised, characterised by short attention spans, sensationalism, inability to empathise and a shaky sense of identity".

“erosion of our identity”

- In the same article Wintour reports that Susan Greenfield had said she found it strange we are "enthusiastically embracing" the possible erosion of our identity through social networking sites, since those that use such sites can lose a sense of where they themselves "finish and the outside world begins".

Claims from qualified scientists

- These articles are based on claims made by two well qualified scientists.
- Dr Aric Sigman is a Fellow of the Royal Society of Medicine, an Associate Fellow of the British Psychological Society, Member of the Institute of Biology and has received the Chartered Scientist award from the Science Council. (Sigman, 2009)
- Baroness Susan Adele Greenfield (CBE 2000, Baroness, 2001; Fullarian Professor of Physiology and Comparative Anatomy, Oxford, 1999-) and Director of the Royal Institution (Royal Institution, 2009).

However... these claims are not evidence based.

- Despite the eminent position of Baroness Greenfield, what she is doing is expressing an opinion.
- Science can be deployed to answer these questions. e.g. how evidence shows that Facebook has helped some build social capital (Ellison et al., 2007).

How can we come to an evidence-based claim?

Rational v Traditional authority

- We can choose to avoid relying on traditional authority and instead rely on rational authority.
- Driver, Newton, Osborne (2000), for example, have argued for doing so in the classroom.
- Scientists must try and make use of rational authority.
- If someone asks you why you believe something, the answer “because I read it on Wikipedia” or “because Newton said so” (traditional authority), or other sources, is a weak justification.



PART 2

Science is about high quality, quantitative data.

Quality of data

- **Claim:** Virtual Reality provides a better platform for e-commerce than traditional picture and text web sites.
- How might this claim be supported by data?
- **Verbal reports** e.g. 9 out of 10 people said that they preferred QTVR.
- **Performance data** e.g. In a recall study Howes et al. (2001) found that users of Virtual Reality had better memory for products.

as scientists we need to avoid
anecdotal evidence
everyone has a story...



getting the facts right...

- http://www.timesonline.co.uk/tol/comment/columnists/david_aaronovitch/article5834725.ece
- Some data is nothing of the sort. If you look on the web you will find the following statistic:
- “the average Brit is caught on security cameras some 300 times a day”
- For example, this claim was published in The Sunday Times almost exactly two years ago, and referred “to the results of a study by the Government’s privacy watchdog” (the Office of the Information Commissioner), which “found people were caught on a national network of 4.2 million CCTV cameras an average 300 times a day”.



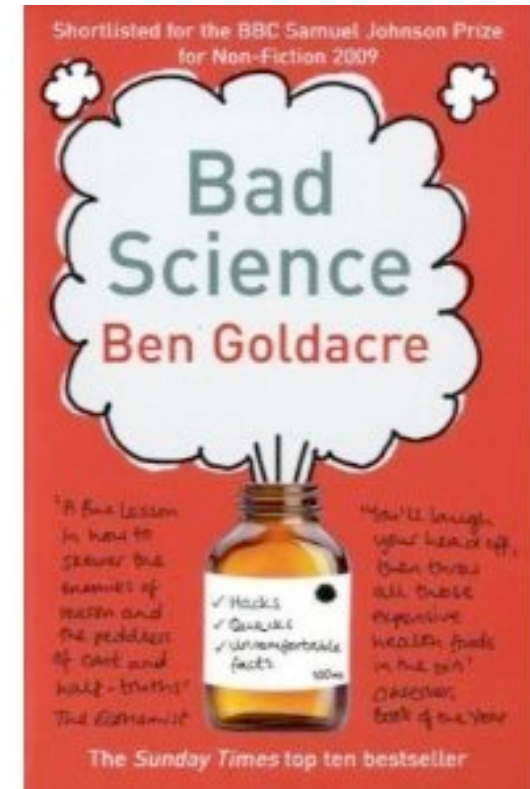
mutated facts...

- Aaronovitch (2009) shows how the “300 times” had become viral: “It now occurs all over the place, and is the standard statistic used for the number of times Britons may or will be captured by CCTV cameras daily.”
- He also shows the tendency for the statistic to mutate, as in the transformation from “can be captured” to the completely different “the average Briton is captured”. A British boy can have a baby at 13. That is clearly not the average age of first fatherhood.
- A New Statesman columnist had it as the “average Londoner going about his or her business... may be monitored by 300 cameras each day”, and a Daily Mail report that “it has been calculated that each person is caught on camera an average of 300 times daily”.

where does the statistic originate...

- The source was a book “The Maximum Surveillance Society”, published in 1999, by two academics, including a C. Norris.
- It wasn’t a fact at all, but a fiction. Norris had imagined a “day in the life of Thomas Reams”
- “Reams is a City type who, rather unusually, lives on a drug-infested estate. He manages to visit two schools, the maternity wing of a hospital, goes to work, shops, is caught speeding in his car, crosses a level-crossing, parks in several car parks before switching to public transport. He goes to Heathrow airport, then a football match at Chelsea, after which he drives through London’s most notorious red-light district (by mistake, I hasten to reassure the fictional Mrs Reams).”
- by the end of the day the fictional Mr Reams has been observed 300 times.

- This is the study of just one statistic.
- There is plenty of evidence of sloppy reporting and of bad science (e.g. see www.badsience.net for examples).
- The same problems can be found in the academic literature.
- A key contribution of this module will be to help provide you with the intellectual tools to stop you falling into similar traps.





PART 3

Science is about evidence based
argumentation.

The problem

- argumentation is difficult.
 - Arguments are subject to confirmation biases.
 - They are inattentive to opposing positions (Kuhn, 2007).

The Layout of Arguments

what not to do...

- consider this example...

“What are the characteristics of a good manager?

A good manager in my view must possess charisma; an individual without charisma is definitely not going to become a good manager. So why is charisma such an important attribute towards a good manager? Well management and leadership are very closely linked; although not the same usually a good manager can easily be seen as a leader. Looking at good leaders through the years it is hard to come across a good leader who was not portrayed as a charismatic character; for example Winston Churchill; Martin Luther King and Barack Obama were or are seen to be charismatic. When we consider the skills and attitudes which create charisma; it becomes very easy to see why charisma plays such a vital role in the formulation of a good manager. “

what not to do..

- This is a more subtle example.
 - It is a highly cited paper by a respected author in a scientific conference.
 - The citation appears to offer evidence for a claim but the nature of this evidence is not described and may not be present at all.
 - Instead a metaphor follows the citation.
- Brignull & Rogers (2003). Enticing people to interact with large public displays in public spaces.. Proceedings of Interact, 3, 17-24.

Brignull and Rogers (2003):

“Social embarrassment has been identified as a key factor, especially in determining whether people will interact with a public display in front of an audience (Rogers & Brignull, 2002). We draw an analogy here with a street performer in a public place, who invites a participant from the audience to ‘help out’ with their show. Such a person can often be wary of volunteering, not knowing what exactly will be required from them, especially if it entails making them look foolish in the eyes of the on-looking audience.”

What is a good argument?

- Toulmin (1958) provides a framework that can be used to help distinguish well-formed from poorly formed or incomplete arguments.
- If you make a claim that is challenged then you will need to make an argument to support the claim. What form should that argument take?

Claim (C) Data (D) Warrant (W)

- Claim: Virtual Reality could increase online sales.
- Data: Howes et al. (2001) observed that people tend to remember more about the range of available products when using virtual reality.
- Warrant: People who remember more about the contents of a store are more likely to return and therefore more likely to purchase more.

Exercise

- identify the claim, data, and warrant in the following paragraph.

Claim, data, warrant

- The internet has beneficial effects for social connectivity. The Pew internet surveys between 2000 and 2003 asked hundreds of people about the role of email in family communication and a majority of respondents said that it increased frequency of communication. Generally it is thought that increased frequency of contact is associated with higher social connectivity.

Claim, data, warrant

- The internet has beneficial effects for social connectivity. The Pew internet surveys between 2000 and 2003 asked hundreds of people about the role of email in family communication and a majority of respondents said that it increased frequency of communication. Increased frequency of contact is associated with higher social connectivity.

Claim, data, qualification, warrant

- The internet may have beneficial effects for social connectivity. The Pew internet surveys between 2000 and 2003 asked hundreds of people about the role of email in family communication and a majority of respondents said that it increased frequency of communication. Generally it is thought that increased frequency of contact is associated with higher social connectivity.

Practical

- practical class in the R programming language starts next week.
-

The R programming language

- R is a programming language for statistical computing.
- It was first implemented in the 1990s and has millions of users.
- It is possible to get started using R with a simple command-line syntax.
- For example,

```
> 1+2
[1] 3
> data <- c(1,2,3,4)
> mean(data)
[1] 2.5
>
```

end.