FMS1204S: fraud, deception and data

Week 1

FMS1204S: fraud, deception and data

Time: Friday, 4-6pm

Venue: S16-05-98

Instructor: Chen Zehua

- ▶ Office: S16-07-108
- ▶ Email: stachenz@nus.edu.sg

Course Outline

The purpose of this seminar is to explore the relationship between fraud and deception and statistics. Very often misleading claims arise from an ignorance of basic statistical ideas, but statistical methods can also be abused knowingly in fraudulent behavior. On the other hand, statistical methods are also commonly used to detect and uncover fraud and dishonesty. This seminar will discuss the role of statistics in uncovering deception in areas such as:

- ▶ Misleading claims in health;
- ▶ Misleading surveys and opinion polls;
- ▶ Claims and counterclaims in environmental science;
- ▶ Fraud detection in the financial world;
- ▶ Authorship disputes and detecting plagiarism.

Course structure

Weekly workload

- ▶ 2 hours of seminars
- ▶ 8 hours of assignments, projects and preparatory work

Class format

- ▶ Each class has two parts.
    - (i) (75 minutes): 5 presentations. Each presentation consists of a 12-minute (max.) talk and a 3-mininute Q&A.
    - (ii) (15 minutes): Introduction of the topics for the next week.
- ▶ You are going to work in groups of three, each group will prepare and give a presentation based on the reading materials each week. The members in each group will take turns for the presentation.
- ▶ The presenter should upload the slides in the IVLE workbin in advance. The file name should include Group #, Week # and the name of the presenter.

Course structure (cont.)

Course Materials

- ▶ The reading materials include science news, popular science books, and scientific journal articles.
- ▶ The reading matrerials will be assigned by the instructor, but you are also welcome to find your own reading materials related the topics of the week.
- ▶ The assigned books can be found in the RBR of the Science library. Some assigned reading materials will be uploaded on the IVLE workbin.

Assessment

- A satisfactory/unsatisfactory final grade will be given. The grade will be based on a numerical score which assesses your performance in attendace, contribution to class discussion, final written report and class presentation.

- The final written report is to be handed in at the end of the semester. The report should be two pages long and include the topics you have learned or heard in the semester. The report should mimic the writing style of scientific journal articles (the structure, the wording, the references, etc.)

- The weightage of the score is as follows:
  - 10% for attendance
  - 10% for contribution to class discussion
  - 20% for your written report
  - 60% for your class presentations

Advantage of taking the FMS module

- ▶ Interactive, independent and peer-based learning.
- ▶ Relaxed and non-examination style.
- ▶ Ability to choose area of study and voice out opinions freely.
- ▶ Enables bonding among classmates.
- ▶ Personal development: skills that will help your future studies/career plans:
    - ▶ Presentation
    - ▶ Team working
    - ▶ Learn how to read scientific literature
    - ▶ Learn how to write scientifically
    - ▶ Critical thinking
    - ▶ Better common sense

Guidlines on presentation

- ▶ Know your audience.
- ▶ Understand everything you present.
- ▶ Provide an outline and organize your slides logically into sections.
- ▶ Make your slides vivid and impressive, use pictures and figures for summary or illustration while possible.
- ▶ Get the main ideas across, leave out the details.
- ▶ Talk about 30s to 2min per slide.
- ▶ Provide a take-home Message.

Group Membership

- Group 1: Celestine Foo Rui Ling; Chua Ming Yuan, Mervin; Leong Jun Yue.
- group 2: Crystal Wang; Hong Chuan Yin; Lieu Wei Zhi Ivan.
- group 3: Leow Gin Ee; Lim Ruiyuan James; Lim Zhi Zhong Walter.
- group 4: Somesh Dev s/o Mohan; Tan Shi Ying; Tan Yong Jia.
- Group 5: Foo Jia Yuan; Chong Hui Ping; Liew Xuan Qi.

During the break: Get to know each other

- ► Identify your group members.
- ► Self introduction.
- ► Conduct a mutual peer interview using the interview form provided.
- ► Exchange contact information after the class.

Discussion on dishonesty with data

Reference
Hand, D. (2007), Deception and dishonesty with data: fraud in science, *Significance*, Vol. 4, 22–25.

Discussion on dishonesty with data

Historical example 1

- In 2006, the biologist Woo Suk Hwang was accused of manufacturing data relating to his research.
- He had attracted worldwide attention for his claims to have cloned cows and a dog.
- An American collaborator, Gerald Schatten, raised ethical concerns about Hwang paying women for human egg donations.
- In subsequent investigations it transpired that in addition to manufacturing data he had also abused his research funds in various ways, such as buying his wife a car and giving gifts to politicians.
- Before his demise he was given the title "First Outstanding Korean Scientist" and even had a set of stamps issued in his honour.

Discussion on dishonesty with data

Historical example 2

- ► The physicist Robert Millikan (winner of the 1923 Nobel prize for physics) is famous for an experiment in which he measured the charge of an electron. The experiment involved measurements on oil drops.

- ► In an article describing his experiment he stated about his published data that "this is not a selected group of drops but represents all of the drops experimented on during 60 consecutive days".

- ► Robert Holton in 1978 examined his notebooks which showed that many recorded observations were left out of his analysis (data were reported on 58 droplets but he recorded data for 175). His notebooks contained comments about the observations such as "publish this beautiful one", "very low, something wrong", "agreement poor, will not work out".

Discussion on dishonesty with data

Consequences

- ▶ We will see many more examples of dishonesty with data in this course.
- ▶ Such examples are very common in science, medicine and the financial world, and the consequences of dishonesty with data can be serious.
- ▶ In medicine, dishonesty with data can endanger lives, with inappropriate treatments being given to patients and research priorities and resources skewed based on misinformation.
- ▶ In the financial world, many people can lose their jobs and in very extreme cases even the stability of the financial system could be threatened.
- ▶ People can mislead with data both innocently and knowingly, and we will be concerned with both kinds of situations.

Discussion on dishonesty with data

Kinds of dishonesty

- ▶ Charles Babbage, one of the founders of the Royal Society, constructed a classification of different kinds of dishonesty with data.
  - ▶ Hoaxing - dishonesty which is intended ultimately to be revealed, often to ridicule those who are misled.
  - ▶ Forging - unlike hoaxing, the deception is intended to last a long time.
  - ▶ Trimming - "clipping off little bits here and there from those observations which differ most in excess from the mean, and in sticking them on to those which are too small".
  - ▶ Cooking - selecting the data to support an argument (the example involving Millikan is an illustration).

Discussion on dishonesty with data

Dishonesty with data: why?

- ▶ "I know my theory is right, and I am entitled to clean up the data to prove it." (Hand, 2007).
- ▶ Desire for promotion or to retain a job.
- ▶ Desire for peer or public recognition, status.
- ▶ The extent of the dishonesty frequently increases over time - dishonesty in a small way by trimming or selecting data may eventually lead to simply making data up.
- ▶ An unintended error with data may subsequently lead to intentional dishonesty (as someone attempts to cover up a mistake).

Dishonesty with data in the medical world

This is a particularly intriguing subject, both because of the many interesting historical examples and because lives can be at stake.

Groups and readings for next week

Group 1:

- ► Salsburg, David (2008). *The lady tasting tea : how statistics revolutionized science in the twentieth century*, (Chapter 18, Does smoking cause lung cancer?)
- ► Questions to address specifically from the reading:

  Does smoking cause lung cancer? Why was it controversial for a very long time?

Group 2:

- ▶ Goldacre, Ben (2008). *Bad Science*, Fourth Estate, London, pp. 28–62 (Chapter 4, Homeopathy).
- ▶ Questions to address specifically from the reading:
  What is homeopathy?
  What is the placebo effect?
  Explain the concepts of blinding, randomization and meta-analysis.

Groups and readings for next week

Group 3:

- ▶ Goldacre, Ben (2008). *Bad Science*, Fourth Estate, London, pp. 86–111 (Chapter 6, The Nonsense du Jour).
- ▶ Questions to address specifically from the reading:

    What are the "four key errors" with evidence identified by Goldacre as commonly made by so-called "nutritionists"? Elaborate on these errors for the case of health benefits of antioxidants.

Groups and readings for next week

Group 4:

- ▶ Wijlaars, L (2012), Side-effects in antidepressants: the drug or the disease? *Significance*, Volume 9, Issue 5, pages 10-13.
- ▶ Questions to address specifically from the reading:

  What is a challenge-dechallenge-rechallenge trial? What is it for?

  Is Prozac safe or not for adolescents and young adults?

Groups and readings for next week

Group 5:

- ▶ Rosenthal, J. (2005). Struck by lightning: The curious world of probabilities. London: Granta Publications, pp. 96–116 (Chapter 7, White Lab Coats).
- ▶ Questions to address specifically from the reading:
  What kinds of bias can afflict medical studies?
  What are the difficulties of determining cause/effect relationships?