Statistical Inference

Lecture 01a

ANU - RSFAS

Last Updated: Tue Feb 21 16:52:18 2017

What is Statistics?

- Statistics is the **science of learning from data**.
- Professor Jeff Wu in November 1997 gave a talk for his appointment to the H. C. Carver Professorship at the University of Michigan titled:

Statistics = Data Science?

http://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf

John Tukey (1915 - 2000)

"The best thing about being a statistician is that you get to play in everyone's backyard." — J. Tukey

– coined the terms 'bit' and 'software'.

Backyards that I Play In

- Assessing uncertainty in weather predication $\rightarrow$ Atmospheric Science.
- Developing a 'Health' index for streams $\rightarrow$ Environmental Science.
- Statistical models for game theoretic data $\rightarrow$ Political Science, Economics.
- Statistical models for network data $\rightarrow$ Sociology, Political Science, Economics, Biology.
- Statistical models for relating **gene marker data** to **genetic line data** to **phenotype data** $\rightarrow$ Biology.

Some Other Facts - American Statistical Association

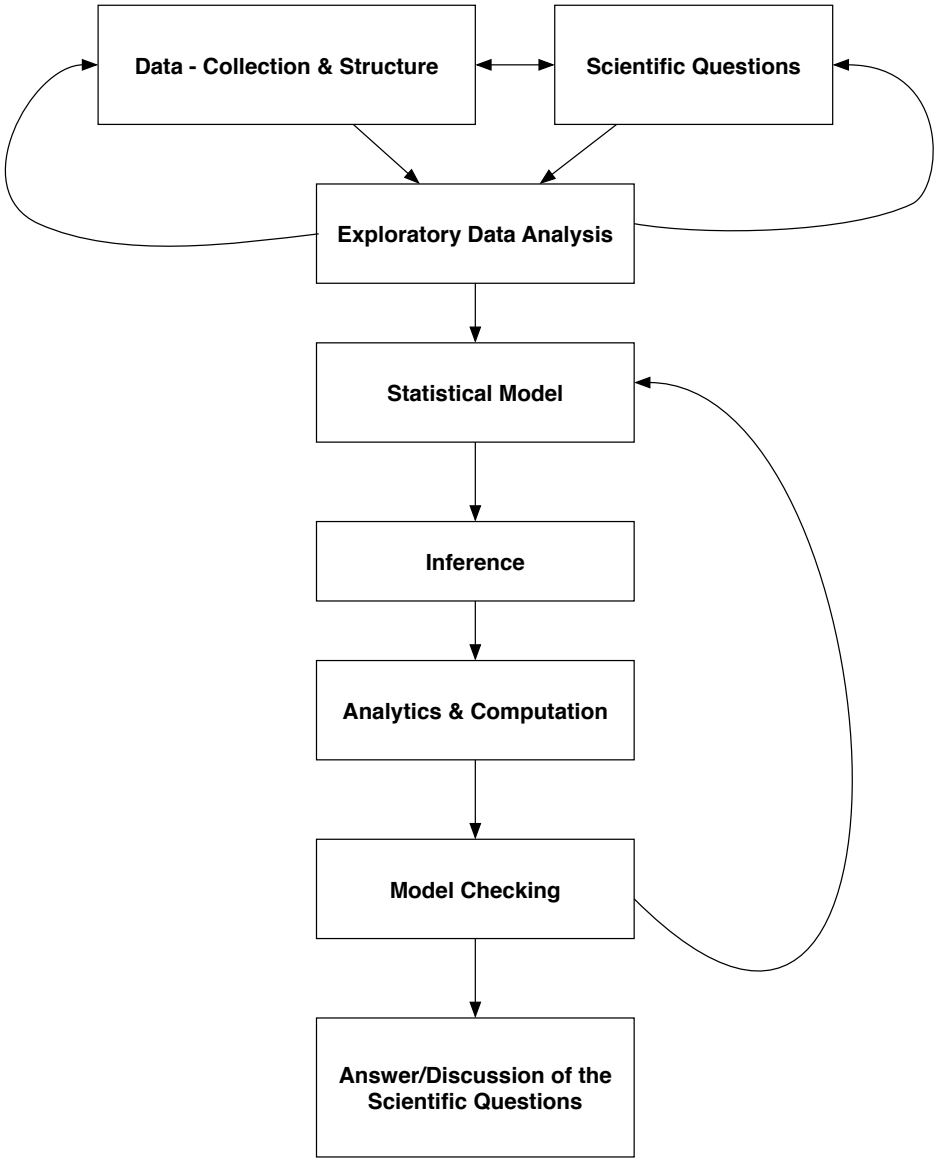
- Statistics is a big, important and growing field. In fact, it's a science . . . the science of learning from data. And as data has become more prevalent and important in our world, so has the field of statistics.

- Statistics offers tremendous opportunity for graduates. For instance, did you know?
  - Statistical analysis was the most in-demand skill in 2014 according to LinkedIn.
  - It is among the fastest-growing jobs? In fact, the McKinsey Global Institute predicts there will be a shortage of up to 190,000 people with the data analysis skills to work with Big Data by 2018.
  - Statistics also attracts a large number of women. More than 40% of undergraduate and graduate statistics degrees are awarded to women, the most of any of the quantitative sciences (e.g., math, engineering, computer science, physics, chemistry, . . .). Recently, CareerCast ranked statistician as a top job field for women.

Some Other Facts - American Statistical Association

- The popularity of statistics also is growing on college campuses across the U.S.
- Fortune magazine recently reported that it's one of the fastest-growing STEM majors. Since 2003, the number of schools granting undergraduate statistics degrees has increased from 74 to more than 110 in 2013.
- But even with this growth, it's not enough to meet the outsize demand for people with data analytics skills.

Thoughts on Statistics & Science



Course Description

This course introduces students to the theory underlying the development and assessment of statistical techniques in the areas of:

- point estimation
- interval estimation
- hypothesis testing

Statistical Inference - the course is roughly broken into the following three sections

- Point estimation
  - Frequentist (maximum likelihood, method of moments, ...)
  - Bayesian
  - Non-Parametric (frequentist)
- Interval estimation
  - Frequentist
  - Bayesian
  - Non-Parametric (frequentist)
- Hypothesis testing
  - Frequentist
  - Bayesian
  - Non-Parametric (frequentist)

Format

- Lectures in CBE Bld LT 2
  - Tuesday 12:00 - 1:30
  - Thursday 3:30 - 5:00
- Tutorial (starting in the second week)
  - TBD

Texts I

- Prescribed Texts
  J. Rice
  Mathematical Statistics and Data Analysis (third edition)
  Brooks/Cole Cengage Learning
- Recommended Reading
  1. G. Casella and R. Berger
     Statistical Inference (second edition)
     Brooks/Cole Cengage Learning
  2. Professor Steven Stern's Notes 1-3
     ANU
  3. G. Givens and J. Hoeting
     Computational Statistics (second edition)
     Wiley

Texts II

4. J. Kadane
   Principles of Uncertainty
   http://uncertainty.stat.cmu.edu/wp-content/uploads/2011/05/principles-of-uncertainty.pdf
   CRC Press

5. C. Robert
   The Bayesian Choice (second edition)
   Springer

Texts for Revision

- D. Wackerly, W. Mendenhall, and R. Scheaffer Mathematical Statistics with Applications (seventh edition)
  Duxbury, Thomson, Brooks/Cole (WMS).
- R. Adams and C. Essex
  Calculus: A Complete Course (eigth edition)
  Pearson

Assesments

- Online quiz (0%)
- Final Examination (60% or 80%)
- Mid-Semester Examination (20% or 0%) (redeemable in favour of the final)
- Group Presentation/Project (15%)
- Weekly (most weeks) Tutorial Solutions (5%)

Tutorials

- Before most tutorials (see schedule) you should submit your answers to the tutorial questions online via Wattle.
- These will be graded weekly for "performance" (whether you made a solid attempt on the questions) and not whether you got the answer correct.
- Each week the "performance" will be graded as 0 or 100.
- Students may be asked to present their solutions during tutorial.

Group Presentation/Project I

- In groups of 2-4 (TBD), based on your cohort (STAT3013 or STAT4027/STAT8027), you will read and present an academic/scientific paper.
- In addition you will have to consider some type of "extension". This may be by simplifying the problem and considering another estimator and its properties, extending the inferential method, or even considering other data sets.
- Each presentation will last 15-20 minutes (TBD) and each member of the group must speak.
- Every presenter in a group will be given the same grade. Your presentation materials must be submitted on Wattle by the beginning of Week 12 (each member in the group should submit the same material).

Group Presentation/Project II

- All paper choices must be approved by me.
- You must attend the presentations of other groups. Failure to do so will results in 10% reduction in your final total grade.

STAT4027/STAT8027

You should choose a paper from a well known statistics journal, for example:

- The Annals of Applied Statistics
- The Annals of Statistics
- Australian & New Zealand Journal of Statistics
- Biometrika
- Canadian Journal of Statistics
- Environmetrics
- Journal of Agricultural, Biological, and Environmental Statistics
- Journal of the American Statistical Association
- Journal of Business & Economic Statistics
- Journal of the Royal Statistical Society (any series)
- Stat
- Statistics in Medicine
- Statistics and Public Policy

I will also consider other well known empirical journals such as:

- Econometrica
- Political Analysis
- Social Networks

I will allow much broader choices that have a strong data analytic and inferential component, so you may also consider:

- Chance
- Significance
  - "Does New York City really have as many rats as people?", Jonathan Auerbach. [pdf]

Examinations

- Mid-semester: Reading time: 15 minutes. Writing time: 90 minutes (I may give you the full 3 hours but the exam will be set to be 90 minutes).
- Final: Reading time: 15 minutes. Writing time: 3 hours.
- Some questions will differ between the STAT3013 & STAT4027/STAT8027 exams.

The permitted material for the mid-semester and final exams will be:

- A4 pages (Two sheets) with notes on both sides
- Paper-based dictionary, no approval required (must be clear of ALL annotations)
- Calculator (Any - programmable or not)