

STAT 545 - Data Analysis

Term 1, 2005-2006

Mon, Wed. 13:30-15:00, LSK 301 (starting Sept. 7)

3 credits

<http://hajek.stat.ubc.ca/gustaf/stat545>

Paul Gustafson (gustaf@stat.ubc.ca)

OFFICIAL CALENDAR DESCRIPTION AND PREREQUISITE: This is very outdated and should be ignored. The prereq. is not being enforced. Recent versions of this course (taught sporadically over the last decade) have been more in line with the following.

DESCRIPTION: This is an overview course in “modern applied statistics.” Goals include:

- Exposing students to a broad range of statistical methods and concepts. Both core techniques (linear models, generalized linear models) and more modern and flexible techniques will be discussed.
- Exposing students to datasets and scientific problems that exemplify modern statistical practice and thinking.
- Exposing students to modern software (R/S) for data analysis and statistical research.

Note that the emphasis is on understanding concepts and statistical principles, while the ‘mechanical’ side of getting software to implement various procedures is secondary.

AUDIENCE:

Statistics graduate students: This course is regarded as an overview of applied statistics that would be useful for say a first-year M.Sc. Statistics student, though some more senior Statistics students may find the course to be of interest.

Graduate students from other departments who are particularly interested in quantitative methods are welcome. The instructor will endeavour to present material in as self-contained manner as possible, though students with less exposure to undergraduate statistics and mathematics may have to undertake some background reading at times. It is hard to pinpoint exactly what a suitable background for the course is, but some combination of:

- some undergraduate math and statistics,
- general quantitative and mathematical aptitude/maturity, including comfort (or ability to quickly get comfortable) with some basic mathematical/statistical notation,
- some experience using statistical methods,
- ability to read-up on ‘missing’ background independently,

should prove useful.

To try to say more about this, if your work requires use of ‘sophisticated’ statistical methods (more than t-tests and ANOVA, for instance) and/or you deal with ‘complex’ data, then 545 may

well be useful to you. If you are looking to acquire particular strengths on the quantitative and methodological side of your own discipline, then 545 may well be useful to you. If you are simply looking for a ‘second course’ in statistics than goes only modestly beyond an intro stats course, then 545 is overkill for you! The real acid test of course is to attend the first few lectures and see if you understand and like what is going on!

TEXTBOOK: Venables and Ripley (2002). Modern Applied Statistics with S (4th ed.). Springer.

COURSEWORK AND EVALUATION: Not finalized, but will involve assignments with somewhat open-ended tasks. There will not be a final exam.

LECTURE TOPICS:

1. Statistical Principles I (estimation: ad-hoc, likelihood, Bayesian).
2. Statistical Principles II (testing: philosophy and criticisms; the pharmacy shelf example).
3. Software and some example tasks (pairs plot, kernel density estimate, permutation test).
4. Linear models I (linear in what? ubiquity of design matrix and least-squares; "model syntax")
5. Linear models II (testing; diagnostics).
6. Logistic regression (how extendible are linear model ideas?).
7. Generalized linear models (general formulation; Poisson regression).
8. Further GLM issues.
9. Contingency tables.
10. Bootstrapping.
11. Model comparison (AIC, BIC versus formal testing).
12. Cross-validation (versus other model comparison techniques).
13. Mode-finding; non-linear models.
14. EM algorithm.
15. Simulation studies.
16. Bayesian computation (realistic???)
17. Random effects I.
18. Random effects II.
19. Fitting curves to data.
20. More curves, and generalized additive models.
21. Tree models.
22. Power calculations.
23. Robustness/outliers.