STA248H1S - Statistics for Computer Scientists - Winter 2015

Instructor: Olga Chilina
Office: SS6002
Office hours: Tue 12:00 - 1:00 pm and Thu 12:00 - 1:00 pm
Email: olgac@utstat.toronto.edu
Lecture time and location: Tue 10:00 am - 12:00 pm, Thu 11:00 am - 12:00 pm in SS2117

Course Objective

Statistics is the science of making sense of data. This course is a survey of statistical methodology. We will quickly cover techniques of data collection and description. Most of the course will focus on methods of analyzing observed data in order to make inferences. The topics covered include (but not limited) descriptive statistics, tests of significance and confidence intervals, power, linear regression and the analysis of variance, and count data.

Textbook

The textbook is *Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences*, 4th edition by J. Susan Milton and Jesse C. Arnold. We will be covering material from chapter 6 through 11, 13, and 15.

Evaluation

The grading scheme is the following (dates are tentative and will be confirmed):

Assignment 1 - 5% (Due on Thursday, February 12 at the beginning of lecture)
Assignment 2 - 5% (Due on Thursday, March 26 at the beginning of lecture)
Two-hour midterm test - 35%
Three-hour final exam - 55%

The assignments will involve both theoretical questions and data analysis projects for which you will use the R programming environment. Additional practice problems will also be assigned weekly and will be posted on the course website (blackboard). Note: 10% of the total marks for an assignment will be deducted for each day it is late.

Midterm Test/Final Exam

The midterm test is on February 24, between 10:00 am and 12:00 pm. **Programmable calculators are not permitted on tests and exam. You must bring your student identification to the term test as well as the final exam.** The day and time for the final exam will be announced later.

Missed Midterm Test

There are **no make-up tests**. Should you miss the term test due to illness, you must submit to your lecturer, within one week, completed by yourself and your doctor, **the 'U of T Student Medical Certificate'**, obtainable from your college registrar, the Office of the Faculty Registrar (SS1006), the Stats Dept. office, or the Koffler health service. The test's weight will then be shifted to the final exam. **If proper documentation is not received, your test mark will be zero.**

Computing

R will be used and it can be downloaded free-of-charge for any of Mac, Windows or Linux operating systems.
There is much documentation for R available at http://cran.r-project.org. You can look through An Introduction to R (available under Manuals) immediately.
The text from the R consoles used in lecture will be posted on the course website.

Additional help

- For continued discussion and questions outside of class, try posting on the Piazza discussion forums. The instructor and TAs will be monitoring them regularly.
- You can visit instructor (SS6002) and TAs (SS1091) during their office hours.
- There is a drop-in Statistics Aid Centre in New College: Wetmore Hall 68A. See http://www.utstat.toronto.edu/wordpress/?page id=154 for the schedule.
- E-mail should only be used for emergencies or personal matters.

Academic Offences

Academic offences are unacceptable, and harm everyone. Offenders are caught, and **sanctions can be severe** - zero in the course with annotation on the transcript for several years; suspension for a year; even expulsion. Various measures, announced and unannounced, will be taken throughout the year to reduce their incidence and to insure successful prosecution when they do occur (e.g. photocopying of students' tests, multiple versions of multiple choice exams). In addition, please note the following:

- ***Oversights in marking on a test paper*** (e.g. addition error, overlooked work) must be brought to the attention of the instructor **immediately** - during the class when test papers are returned
- ***Regrading requests*** will only be considered for ***term tests*** which are written in ***ink***

Coverage of topics

Topics covered will be selected from the corresponding chapters/sections in the textbook (Subject to change, and will update gradually). Note: Not everything we cover can be found in the textbook. You have to read lecture notes to do well on the test/exam.

Week 1: Review of important concepts. Introduction to Statistics.

Week 2. Summarising data: graphical and numerical description of data. (6.1-6.4)

Week 3: Point estimates and their properties. Likelihood function and its application. Method of moments. Maximum likelihood estimators and their properties. (7.1-7.2)

Week 4: Sampling distributions. Confidence intervals: concept, derivation and properties. CI for mean and variance of a population (7.3-7.4, 8.1-8.2)

Week 5: Estimating proportions. CI for proportions. (9.1)

Week 6: Hypothesis testing: concept, process and interpretation. Type I and II errors, power. Hypothesis and significance tests on the mean and proportion. Hypothesis tests on the variance. (8.3-8.6, 9.2)

Week 7: READING WEEK: NO CLASSES

Week 8: TERM TEST tentatively set for Feb 24 on weeks 1- 6 material

Week 9: Comparing two proportions. Comparing means: (i) variances equal (pooled test); (ii) variances unequal; (iii) paired data. Comparing variances: F distribution. (9.3-9.4, 10.1-10.5)

Week 10: Inference for non-normal population: sign test, Wilcoxon signed-rank test, Wilcoxon rank sum test. (8.7)

Week 11: Simple linear regression model. Least squares estimators: their properties and confidence intervals. Residuals. Correlation. General linear models. (11.1-11.3, 11.5)

Week 12: Introduction to analysis of variance. (13.1-13.3)

Week 13: Categorical data: Chi-squared goodness of fit test. Bootstrap. Review and final comments (15.1-15.3)