CIS6660 Summer 2017

Statistical Methods for Computer Science

contact

Dr. Daniel Gillis
Reynolds 316

dgillis@uoguelph.ca
http://danielgillis.ca
twitter:@drdanielgillis

course description

This course will provide students with a review of basic statistical methods and more advanced parametric and non-parametric methods, including, but not limited to: Wilcoxon, Kruskal Wallis and Friedman tests, ANOVA, linear regression, and generalized linear regression.

lecture, lab & final exam times

While this is a reading course, we will have regular meetings to discuss the topics and explore areas of statistics for use by computer scientists. Meeting time and location will be determined.

prerequisites & restrictions

There are no prerequisites or restrictions for this course. It is assumed that students will have an undergraduate level of knowledge of probabilities and some familiarity with simple statistical concepts. Regardless, these will be reviewed in the first part of the course.

texts

To be determined.

grading scheme

The following grading scheme will be used to assess your progress in the course. Extensions for course deliverables will not be provided. As such, it is in your best interest to submit whatever you have completed by the due date. Failure to submit will result in an automatic grade of 0 being assigned. Your assignments should be submitted using LaTeX, and in PDF format unless otherwise stated by the instructor. If you are having difficulty with an assignment, please discuss this with the instructor in advance of the due date. Do not wait until the last day to do so.

Assignment 1	**Due Beginning of Week 3** Will cover topics covered in weeks 1-2	10%
Assignment 2	**Due Beginning of Week 5** Will cover topics covered in weeks 3-4, with some reference to weeks 1-2	10%
Assignment 3	**Due Beginning of Week 7** Will cover topics covered in weeks 5-6, with some reference to weeks 1-4	10%
Assignment 4	**Due Beginning of Week 9** Will cover topics covered in weeks 7-8, with some reference to weeks 1-6	10%
Final Presentation	**Due Week 12**	20%
Final Paper	**Due First Week of Exams**	40%

CIS6660 Summer 2017

Statistical Methods for Computer Science

learning outcomes

Learning outcomes for the course include:

- The ability to critically evaluate scientific literature: the learner will compare and contrast statistical methods in peer reviewed literature.
- The ability to select appropriate statistical methods given constraints of experimental design, data types, data quality, and potential correlation structures: the learner will recommend statistical tests for at least 3 different studies.
- The ability to clean data and manipulate data using R: the learner will develop code to manipulate and clean a data set that has at least 3 different types of data errors (e.g. missingness, data type inconsistencies, etc.)
- The ability to generate data visualizations using R: the learner will be required to create histograms, scatterplots, bar plots using the R statistical software.
- The ability to summarize data using R: the learner will create at least 10 aggregated tables of data that summarize large data sets.
- The ability to implement parametric and non-parametric tests in R: the learner will develop code to implement parametric and non-parametric tests (e.g. Wilcoxon rank sum, chi-square, t-tests, etc.)
- The ability to implement linear regression, ANOVA, and generalized linear models in R: the learner will develop code to implement appropriate regression and factor analyses using R.
- The ability to interpret statistical tests, and p-values: the learner will interpret and discuss findings from at least 5 statistical tests.
- The ability to implement and interpret Bayesian models in R (using JAGS): the learner will develop code to implement a Bayesian analysis for a simple linear model using R and JAGS.

CIS3750Fall 2016

Systems Analysis & Design in Application

topics covered

Topics covered in the course are listed below. Order of presentation of the topics may vary.

Week 01-02 Review of data types, hypotheses, probability rules, distributions

Data summaries & visualizations in R: mean, median, variance, correlations (Pearson & Spearman), histograms, scatterplots, table summaries

Assignment 1

Week 03-04 Parametric methods in R (one, two, or more samples)

Introduction to simulations

Introduction to LaTeX

Assignment 2

Week 05-06 Non parametric methods in R (one, two, or more samples)

Assignment 3

Week 07-08 Linear Regression in R (interactions, and connections to ANOVA)

Introduction to Bayesian Methods, and MCMC (Metropolis Hastings, Gibbs)

Assignment 4

Week 09-12 Logistic & Poisson Regression

Other topics (possibly Data Mining, Machine Learning, etc.)

Critical evaluation of analyses in literature

Final presentations

Final reports