

DSE I1030: Applied Statistics
Prof. Irina Gladkova

Course Description:

This course will examine real data sets from a variety of domains, examine multiple models for these data sets, assess the validity of modeling assumptions, and determine the strength of conclusions that can be drawn. The course will include project-based learning and use a statistical programming language R. A strong emphasis will be placed on the critical analysis of modeling assumptions in real-world settings.

Prerequisite(s):

Intro to programming (CSc 102/103) or equivalent, probability and statistics (CSc 217), calculus (MATH 202), linear algebra (MATH 346), discrete mathematics (CSc 104).

Grades:

H/W assignments (from datacamp) (20%), 2 individual projects (20%), group project (20%), Midterm (20%) and Final (20%); Extra credits for in-class participation.

Topics covered:

- Inferential statistics (such as hypothesis testing and estimation in parametric and nonparametric settings, conditional inference, resampling methods, cross-validation, and multiple hypothesis testing);
- Regression and prediction (such as elements of linear and nonparametric regression, assessment of variable importance, introduction to causal inference).
- Experimental design (analysis of variance)
- Bayesian statistics (such as prior distributions, posterior and predictive inference, and Bayesian model comparison);

Academic Integrity:

CUNY policy on academic integrity can be found at

<http://www2.cuny.edu/wp-content/uploads/sites/4/page-assets/about/administration/offices/legal-affairs/policies-procedures/Academic-Integrity-Policy.pdf>

In particular, according to the policy, Academic Dishonesty is prohibited in The City University of New York and is punishable by penalties, including failing grades, suspension and expulsion.

Textbooks:

1. **IS:** David Diez “OpenIntro Statistics”:
https://www.openintro.org/download.php?file=os3_tablet
2. **AS:** David Dalpiaz “Applied Statistics with R”:
<https://daviddalpiaz.github.io/appliedstats/>
3. **SL:** Gareth James “An Introduction to Statistical Learning with Applications in R”:
https://www-bcf.usc.edu/~gareth/ISL/ISLR_First_Printing.pdf

Online Courses:

1. **SRS:** Statistics with R Specialization
<https://www.coursera.org/specializations/statistics?authMode=login>
2. **FDA:** *Foundations of Data Analysis*
<https://courses.edx.org/courses/course-v1:UTAustinX+UT.7.21x+2T2017/course/#block-v1:UTAustinX+UT.7.21x+2T2017+type@chapter+block@9ff7c7be54f242d98ec9b8fc35bc2ca0>

R programming resources:

- R for Reproducible Scientific Analysis --
<https://swcarpentry.github.io/r-novice-gapminder/>
- RStudio Cheat Sheets -- <https://www.rstudio.com/resources/cheatsheets/>
- R Markdown -- <http://www.stat.cmu.edu/~cshalizi/rmarkdown/>
- The R Inferno -- http://www.burns-stat.com/pages/Tutor/R_inferno.pdf
- **RIS:** Using R for Introductory Statistics --
<https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>

Calendar

Week 1 - Intro to R

August 28 (Tuesday)

August 30 (Thursday)

Week 2 - Review of Prob/Stat fundamentals

September 4 (Tuesday)

Topics: Probability and random variables

Readings: IS (Chapter 2)

September 6 (Thursday)

Topics: Probability distributions, Central Limit Theorem

Readings: IS (Chapter 3)

Week 3 - Review of Prob/Stat fundamentals

September 13 (Thursday)

Topics: Review: Parameter Estimations (method of moments, maximum likelihood);
Confidence interval

Readings: IS (Sections 4.1-4.2), RIS (Section 9)

Project 1 (confidence intervals, residuals, and graph plotting; R markdown)

Week 4 - Hypothesis Testing

September 20 (Thursday)

Topics: P-value, Standard Normal null distribution, Z-test

Readings: IS (Sections 4.3 - 4.5), RIS (Section 9)

Week 5 - Hypothesis Testing

September 25 (Tuesday)

Topics: Errors in testing; Type I&II, level of significance; Alpha and critical values

Readings: IS (Sections 4.3 - 4.5), RIS (Section 10)

September 27 (Thursday)

Topics: T-test; Student Distribution;

Readings: IS (Sections 5.1 - 5.2), AS (Sect.5.2), RIS (Sections 10 & 11)

Week 6 - Hypothesis Testing

October 2 (Tuesday)

Topics: Inference for categorical value

Readings: IS (Sections 6.1 - 6.2)

October 4 (Thursday)

Topics: Chi-square test; Testing independence (based on chi-square)
Readings: IS (Section 6.3), RIS (Section 12)

Week 7 - Regression

October 9 (Tuesday)

Topics: Correlation; R squared

Readings: IS (Section 7.1), AS (Section 7.1)

October 11 (Thursday)

Topics: Least Squares Line

Readings: IS (Section 7.2), AS (Section 7.2), RIS (Section 13)

Week 8 - Regression

October 16 (Tuesday)

Topics: Inference for Linear Regression; Outliers, Influential points

Readings: IS (Sections 7.3 & 7.4), AS (Chapter 8)

October 18 (Thursday)

Topics: Multiple Predictors; Adjusted R squared

Readings: IS (Chapter 8), AS (Sections 9.1-9.3)

Week 9 - Model evaluation / Midterm

October 23 (Tuesday)

Topics: Evaluating Statistical Model: Error & Inference

Readings: AS (Chapter 10)

October 25 (Thursday)

Topics: Model Selection

Readings: AS (Chapter 16), SL (Chapter 6)

Week 10 - Model evaluation/ Midterm

October 30 (Tuesday)

Topics: Cross-Validation; Residuals Plot

Readings: AS (Chapter 13), SL (Chapter 6)

November 1 (Thursday)

Midterm

Project 2

Week 11 - Assessing errors in statistical estimation (Resampling techniques)

November 6 (Tuesday)

Topics: Resampling Methods, Cross-validation (Leave-one-out, k-fold)

Readings: SL (Section 5.1)
November 8 (Thursday)
Topics: Bootstrap estimate of bias; Parametric, non-parametric bootstrap
Readings: SL (Section 5.2)

Week 12 - Analysis of Variation

November 13 (Tuesday)
Topics: Difference of two means; One-way ANOVA
Readings: IS (Sections 5.3 - 5.5), AS(Sections 12.1 - 12.3)
November 15 (Thursday)
Topics: Two-way ANOVA
Readings: AS(Chapter 12.5)

Week 13 - Design of Experiments

November 20 (Tuesday)
Topics: Permutation Test (shuffle method)

Group Project

Week 14 - Bayesian inference

November 27 (Tuesday)
Topics: Prior and posterior; Bayesian estimation
November 29 (Thursday)
Topics: Bayesian hypothesis testing
Data Camp Course:
<https://www.datacamp.com/courses/fundamentals-of-bayesian-data-analysis-in-r>

Week 15 - Bayesian inference

December 4 (Tuesday)
Topics: Markov Chain Monte Carlo
<https://github.com/bwv988/mcmc-tutorial>

December 6 (Thursday)
Group Presentations

Week 16 - Wrap-up, Group Presentations

December 11 (Tuesday)
Group Presentations

Final Exam