

STA 248H1 S - Statistics for Computer Scientists
Summer 2016 (June 27 to August 15)

Lectures: Mondays and Wednesdays 6:10-9pm in **SS 1073**

Instructor: Dr. Shivon Sue-Chee (**E-mail:** shivon.sue.chee@utoronto.ca)
Office hours: Mondays and Wednesdays 5:10-6pm in **SS 6026**

Course website: Available through <https://portal.utoronto.ca> (UT Blackboard)

TA office hours: (To be announced (TBA))

Course description

This course is a survey of statistical methodology with emphasis on data analysis and applications. The topics covered include (but not limited to) descriptive statistics, data collection and design of experiments, tests of significance and confidence intervals, power, linear regression and the analysis of variance, and count data. Students will use R for statistical computing as part of this course.

Pre-requisites

STA247H1 or STA257H1, CSC108H1 or CSC148H1

Required Textbook

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 materials of chapters 6 through 11, 13, and 15. Answers to selected problems are included in the text's Appendix. A student solution manual is available for purchase from bookstores.

Supplementary Textbook

Mathematical Statistics and Data Analysis, 3rd edition by John A. Rice.

Evaluation

	Weight	(Due) Date	Time	Location
Assignments [3]	20%	Jul. 6, Jul. 20, Aug. 3	At 6:10pm	In class
Term Test	30%	Mon., Jul. 18	6:10-8:00pm	TBA
Final Exam	50%	Between Aug. 9-15	(3 hrs)	TBA [®]

[®]See Faculty of Arts and Science Final examination timetable posting on July 13.

The assignments will involve both theoretical questions and data analysis projects requiring the use of R. **Additional practice questions will be assigned at the end of each class and will be posted in the course notes.**

Assignment, Test and Exam Policies

Assignments are to be submitted to the instructor, on or before the respective due date. Late assignments or electronic submissions will not be accepted.

Only non-programmable calculators will be permitted on the test and final. You must bring your student identification to the term test as well as the final exam. The final exam is cumulative.

If the midterm test is missed for a valid reason, you must submit appropriate documentation, in person and in original form, within one week of the test to the instructor. If documentation is not received in time, your test mark will be zero. If the test is missed for a valid reason, a makeup online or oral test will be arranged. The makeup test will be worth 15% and the remaining 15% will be shifted to the final exam.

Requests for test remarking must be submitted at the time the test is returned back to you and no later than 1 week after the test is return to the class. The request must contain a justification and will only be considered for tests which were written in ink.

Where to get help outside of class?

- TA office hours in Statistics Aid Centre, Room 1091, in Sidney Smith Hall from the week of July 4. Schedules will be posted at the course web site.
- Instructor office hours in SS6026.

Computing

This course requires statistical computing. We will use the R computing package for all examples and provide sample code that would be sufficient for you to complete homeworks.

R can be downloaded free-of charge for any Mac, Windows or Linux operating systems. Software, documentation and manuals are available at <http://cran.r-project.org>.

If you wish to use R through UofT, you may do so via CQUEST lab services. To find out more, go to <http://www.cquest.utoronto.ca>.

Communication

In general, I am not able to answer questions about the course material by e-mail. Students are encouraged to attend lectures, and Instructor and/or TA office hours.

Before you send an e-mail, make sure that you are not asking for information that is already on the course website. E-mail is appropriate for personal matters. Use your utoronto.ca or mail.utoronto.ca account and write a proper email including the addressee, your name and student number. I will generally answer e-mail within one business day.

Announcements will be posted on Blackboard. Please check there regularly. If an urgent matter arises, I may contact the entire class by e-mail. In order to receive these messages, please make sure that your ROSI account has your utoronto.ca e-mail.

The TAs and instructor are here to help you. Ask questions and let me know if there are any concerns.

Accessibility Needs

The University of Toronto is committed to accessibility. If you require accommodations for a disability, or have any accessibility concerns about the course, the classroom, or course materials, please contact Accessibility Services as soon as possible at accessibility.services@utoronto.ca or <http://www.accessibility.utoronto.ca>.

Academic Integrity

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters at <http://www.artsci.utoronto.ca/osai/students>. If you have any questions about what is or is not permitted in this course, please do not hesitate to contact me.

Your responsibility

The classroom sessions for this class are designed to actively engage you in the course material. I hope you'll find them interesting, challenging, and fun, and an excellent opportunity to truly learn the material. In order for these sessions to be effective, coming prepared, by learning about the week's concepts through the textbook, is essential.

Tentative Course Schedule

- Week 1. Review of important concepts. (Chps. 1-5) Introduction to Statistics.
Summarising data: graphical and numerical description of data. (6.1-6.4)
- Week 2. Point estimates and their properties. Likelihood function and its application. Method of moments. Maximum likelihood estimators and their properties. (7.1-7.2)
Sampling distributions. Bootstrapping. Confidence intervals: concept, derivation and properties. CI for mean and variance of a population (7.3-7.4, 8.1-8.2)
- Week 3. Estimating proportions. CI for proportions. (9.1)
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 4. **TERM TEST** on July 18 on weeks 1- 3 materials.
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 5. Inference for non-normal population: sign test, Wilcoxon signed-rank test, Wilcoxon rank sum test. (8.7)
Simple linear regression model. Least squares estimators: their properties and confidence intervals. Residuals. Correlation. General linear models. (11.1-11.3, 11.5)
- Week 6. Introduction to analysis of variance. (13.1-13.3)
Categorical data: Chi-squared goodness of fit test. Review. (15.1-15.3)

