

STA 248H1 S - Statistics for Computer Scientists
Winter 2016 (January 12 to April 29)

Lectures: Tuesdays 10:10-12pm and Thursdays 11:10-12pm in **BR 200**

Instructor: Dr. Shivon Sue-Chee (**E-mail:** shivon.sue.chee@utoronto.ca)
Office hours: Tuesdays and Thursdays 12:10-1pm 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) 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.

Evaluation

	Weight	Date	Time	Location
Assignment #1	7.5%	Thursday, Feb. 11	11:10am	In BR 200
Term Test	30%	Tuesday, Mar. 1	10:10-12:00pm	TBA
Assignment #2	7.5%	Thursday, Mar. 24	11:10am	In BR 200
Final Exam	55%	Between Apr. 12-29	(3 hrs)	TBA [@]

[@]See Faculty of Arts and Science Final examination timetable posting on February 12.

The assignments will involve both theoretical questions and data analysis projects requiring the use of R. **Additional practice questions will be assigned weekly** and will be posted on the course website.

Assignment, Test and Exam Policies

Late assignments are not encouraged; 10% of marks will be deducted for each hour an assignment is late.

Only non-programmable calculators are 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 within one week of the test. If documentation is not received in time, your test mark will be zero. If the test is missed for a valid reason, its weight 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 2 weeks after the initial test return date. 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?

- Our online discussion forum via Piazza. Forums will be moderated by TAs and instructor regularly.
- TA office hours in Statistics Aid Centre, Room 1091, in Sidney Smith Hall from the week of January 18. Schedules will be posted at the course web site.
- Instructor office hours in SS6026.
- Experienced TAs at the New College Stats Aid Centre at NC 68A (see schedule posted at http://www.utstat.toronto.edu/wordpress/?page_id=154).

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, tutorials and Instructor and/or TA office hours.

E-mail is appropriate for personal matters only. 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 two business days.

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. 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. **READING WEEK: NO CLASSES**
- Week 7. 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 8. **TERM TEST** tentatively set for Mar 1 on weeks 1- 7 materials.
- 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. (15.1-15.3)