STA305H1S/STA1004HS, Design of Scientific Studies
LEC0101/0201, Winter 2019 Syllabus

Instructor: Prof Mark Ebden
Email: mark (dot) ebden (at) utoronto (dot) ca
Office: EP 102
Office hours: Mondays 2-3 pm (click here) and Tuesdays 3-4 pm (click here), by appointment.
Background: My experience has mainly been in gathering and analysing *scientific data,* with studies conducted in biomedical sciences, engineering, etc.

Teaching Assistants: Tianxiao, Michael, Vanessa, Galen, Alan, and Emad.
Office hours in SS B623, at times to be described on the 'Syllabus' page on Quercus.

Course webpage: Accessible through Quercus at q.utoronto.ca. Course materials provided on Quercus are for the use of students currently enrolled in this course only. Providing course materials to anyone outside of the course is unauthorized use.

Classroom sessions:

- Section 1: Monday 12-1 pm and Wed 11-1 pm, NF 003, from 7 January to 3 April
- Section 2: Tuesday 1-2 pm and Thurs 4-6 pm, ES 1050, from 8 January to 4 April

Course Content

This course will provide an introduction to the fundamental concepts of the design of scientific studies including the design of experiments and observational studies. Students will be become acquainted with statistical methods used to design and analyze experiments and observational studies. In particular this course will cover: experiments versus observational studies, clinical-trial design, comparing several groups using a completely randomized design, randomized blocks, Latin squares, incomplete block designs, factorial designs, causal inference in randomized studies and non-randomized studies, and adjusting for selection bias using propensity-score methods.

The learning objectives of this course are:

- Understand the ideas, principles, and considerations that are common to the design and analysis of scientific studies including the statistical design of experiments and observational studies.
- Develop a statistical toolbox of methods for the design and analysis of experiments and observational studies.
- Identify appropriate uses and interpretations of experimental designs, and observational studies, including their strengths and limitations.

Topics

Experiments, observational studies, and causal inference: Experiments versus observational studies, and causal inference in randomized experiments.

Selection Bias in Observational Studies: Causal inference in randomized experiments versus observational studies. Introduction to the propensity score and three ways to use the propensity score to adjust for selection bias: matching, sub classification, and direct regression adjustment.

Probability and Statistics: Mathematical statistics used in experimental design.

Comparing Several Groups: Comparing several groups in an experimental setting or observational setting and deciding whether differences that are found are likely to be real or due to chance.

Design of Clinical Trials: The design and analysis of clinical trials with continuous or binary variables will be introduced.

Blocking techniques: Blocked designs, Latin squares, randomized incomplete block designs.

Factorial Designs: Factorial, blocked factorial, and fractional factorial designs will be discussed.

Split-unit designs: Split-unit designs will be discussed as an example of restricted randomization in the design of experiments. *If time permits*

Textbooks

There are **no required textbooks** for this course. The following books in the U of T library system are listed in decreasing order of relevance:

1. *Statistics for Experimenters: Design, Innovation, and Discovery.* Box, G.E.P., Hunter, J.S., Hunter, W.G. Wiley 2nd Ed. 2005
2. *Experiments: planning, analysis, and optimization*. Wu, C.F.J., Hamada, M.S. Wiley, 2009, 2nd ed.: http://go.utlib.ca/cat/8598479
3. *Causal inference for statistics, social, and biomedical sciences*. Imbens and Rubin. Cambridge University Press, 2015. http://go.utlib.ca/cat/10127748
4. *Design and Analysis of Experiments*. Dean, A., and Voss, D. Springer. 1999. UofT link to electronic copy: http://go.utlib.ca/cat/2573215
5. *Design of Observational Studies*. Rosenbaum, P. R. Springer 2010. UofT link to electronic copy: http://go.utlib.ca/cat/7890274
6. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Gelman and Hill, 2007
7. *Clinical Trial Design: Bayesian and Frequentist Adaptive Methods*, G. Yin, 2012.

One copy of reference 1 will be on reserve at the Mathematical Sciences Library (Room 6141, 40 St George Street) while this course is running. A small number of copies may be available for purchase at the UofT Bookstore for about $160 each.

Most of references 2 through 7 are available electronically through the UofT library.

The book *Fundamentals of Statistical Experimental Design and* Analysis by Robert Easterling (Wiley, 2015) tries to provide a folksy overview of key concepts; it's not in the U of T library system, despite a recent request.

Evaluation

Undergraduate and graduate students will be evaluated according to the following marking scheme.

Assessment	Weight	Notes
Assignments	6%, 6%, and 6%	Not to be discussed during office hours
Midterm test	34%	Occurring during class time
Final exam	48%	Expected 6-30 April

Graduate students will be evaluated at the graduate level according to the University Assessment and Grading Practices Policy.

You are not evaluated on attendance or similar forms of participation. But, generally the more you put into a course the more you'll get out of it.

Assessments

You are allowed a one-sided 8-1/2"x 11" (standard letter size) aid sheet on the midterm test, and a two-sided aid sheet on the final exam. You must bring your student identification to all assessments. You can keep your aid sheet after every assessment, including the exam.

You will need to know how to interpret output from R on the test and exam.

Marking concerns

Any requests to have marked work re-evaluated must be made in writing within *one week* of the date the work was returned. The request must contain a justification for consideration. Send your request to sta305me@gmail.com, with your assessment name (e.g. HW1) or date (e.g. "Feb 27") in the subject line.

Submitting your assignments

• Assignments are submitted electronically on Crowdmark, never via email messages.
• If you leave your submission to the last minute and an overburdened Crowdmark records a late timestamp while processing multiple requests, this is your responsibility. If you're not *feeling lucky*, submit at least 20-30 minutes early.
• Assignments submitted late by less than 24 hours will be penalized 10%.
• Those submitted late by less than 48 hours will be penalized 20% total.
• Those submitted more than 48 hours late will receive a grade of zero except in the case of medical reasons documented properly (see below section). The weight for the assignment will be transferred to the weight for the *final exam*.
• After you have submitted your work, staff may use *Turnitin* software to help consider the class's assignment submissions.

What if I miss the test or an assignment?

- If it is missed for a valid medical reason, you can submit the proper documentation (see below section), and the *final exam* will be worth more.

- Other reasons for missing an assessment will require prior approval by your instructor. If prior approval is not received for non-medical reasons then you will receive a grade of zero for that assessment.

Submitting medical documentation:

- Submit the University of Toronto Verification of Student Illness or Injury form soon after the assessment occurred. (Come to any lecture or office hours.)
- The form will only be accepted as valid if the original paper form is submitted, filled out according to the instructions on it.
- The form must indicate that the degree of incapacitation on academic functioning is moderate, serious, or severe in order to be considered a valid medical reason for missing the term test. If the form indicates that the degree of incapacitation on academic functioning is negligible or mild then this will *not* be considered a valid medical reason.

Computing

We will make use of the R language (and RStudio). The main advantages of R are that it is freeware and that there is a lot of help available online.

R is freely available for download at cran.r-project.org for Windows, Mac, and Linux operating systems. *RStudio* is a good integrated development environment to R. It is freely available at www.rstudio.com/products/rstudio You may also like to sign up for a CQUEST account. To get an account and find out more information about using CQUEST go to www.cquest.utoronto.ca

If you need help installing R and RStudio, and learning the basic syntax of R, a helpful document is here. Another good online R reference is here, and there is a downloadable book called *Introduction to R* by William Venables and David Smith. The alternative reference textbooks above, and websites such as datacamp.com, are also good to help you pick up R.

I'll give you a quick introduction to R and provide you with R source code for the examples in lecture. Note that there are many graphics options available to produce plots, but we'll focus on the basics, sufficient for this course. You may like to use *Rmarkdown* to type up your work.

Calculators

You will need a calculator. Any calculator that has logarithmic functions will be sufficient. Calculators on phones or other devices equipped to communicate with the outside world (for example, through the internet or cellular or satellite phone networks) will not be permitted during the paper-based assessments.

You will see from old midterms that manual calculations are a part of this course.

Online Discussion Board

This term you will have the option to use Piazza for class discussion. If you decide not to use Piazza it will not disadvantage you in any way, and will not affect official University outcomes (e.g. grades and learning opportunities). If you choose not to opt-into Piazza then you can ask questions or discuss course material with the instructor or TAs during office hours.

Please read Piazza's Privacy Policy and Terms of Use, taking time to understand and be comfortable with them. They provide for substantial sharing and disclosure of your personal information held by Piazza. If you decide to participate in Piazza, only provide content that you are comfortable sharing under the terms of the Privacy Policy and Terms of Use.

The Piazza system is catered to getting you help quickly and efficiently from classmates, the TA, and the lecturers. Rather than emailing questions to the teaching staff, we encourage you to post your questions on Piazza. To sign up for the discussion forum, click on the link: https://piazza.com/class/jpcwtgs37dd3js

Additional help

Need extra help with the coursework? Here are some options:

- For continued class discussion and questions outside of class, try posting on the discussion forums. The instructor and TAs will be monitoring them

- You can visit the instructor or teaching assistants during their office hours
- You may choose to join (or create) an STA305 Recognized Study Group: http://www.artsci.utoronto.ca/current/recognized-study-groups/
- E-mail should only be used for emergencies or personal matters

Questions concerning the content/format of upcoming assessments won't generally be addressed in office hours. Please raise your hand in class, or ask on Piazza, so that everyone may hear the answer.

How to communicate with your instructor

Questions about course material such as:

- How do I do question 3.7 in this textbook?
- What is standard deviation?
- When is the midterm?

can be posted on the discussion forums. If you are shy, questions can be posted anonymously (so that the author is anonymous to other students but not to the instructors).

For private communication, such as "I missed the test because I was ill," e-mail your instructor.

You may post entirely anonymous feedback here. Nobody will know who you are. In past courses this has helped sort out small problems and big ones.

Academic integrity

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters at www.governingcouncil.utoronto.ca/policies/behaveac.htm. If you have any questions about what is or is not permitted in this course, please do not hesitate to contact your instructor.

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: accessibility.services@utoronto.ca or http://accessibility.utoronto.ca.

Your responsibilities

The classroom sessions for this class are designed to actively engage you in the course material. We hope you'll find them interesting, challenging, fun, and an excellent opportunity to truly learn the material.

Design of Scientific Studies
STA305H1S/1004HS, LEC0101 / LEC0201
2019 Winter term Course Schedule

Assignment deadlines

- HW 1: Sunday 27 January, 4 pm
- HW 2: Sunday 10 February, 4 pm
- HW 3: Sunday 17 March, 4 pm

Tentative List of Lecture Contents, by Week Number

1. Overview of experimental design and review of some mathematical statistics
   - Hotelling's weighing problem
   - Introduction to Experiments, Observational Studies, randomization, replication, blocking
   - Review of the Normal distribution (univariate, bivariate), hypothesis testing type I, II errors, statistical significance, linear regression

2. Hypothesis testing via the randomization Distribution
   - Hypothesis testing via the randomization distribution
   - Comparing two means using the randomization distribution

3. Comparing two groups – Design of Phase III Randomized Clinical Trials
   - Clinical trials terminology
   - Statistical design of phase III randomized clinical trials (RCT)
   - Two-sample t-test versus randomization test
   - Power and sample size in the design and planning of phase III RCT
   - Continuous, binary, and time-to-event endpoints

4. Causal inference in randomized experiments
   - Potential responses under alternative treatments
   - Covariates and outcomes
   - Possible treatment assignments and randomization
   - Interference between units
   - Testing the null hypothesis of no treatment effect

5. Two Simple Models for Observational Studies
   - Population before matching
   - The ideal matching
   - Naïve model
   - Propensity score
   - Balancing property of the propensity score
   - Propensity score and ignorable treatment assignment

6. Comparing more than two entities
   - Comparing more than two entities using ANOVA
   - Sample size estimation
   - Contrasts and treatment means
   - ANOVA as a linear model
   - ANOVA in R

7. **Test** tentatively scheduled for 27 & 28 February, in your section's class

8. Comparing more than two entities
   - Sample size estimation
   - Contrasts and treatment means
   - Multiple comparisons
   - Checking Model Assumptions

9. Factorial designs at two levels
   - 2^2 factorial designs
   - Cube plots
   - Linear model representation of a factorial design
   - Analyzing data from factorial design in R

10. Blocking Designs
   - Randomized Blocked Designs
   - Latin squares
   - Balanced incomplete block designs

11. Fractional Factorial Designs
   - Normal and Lenth Plots
   - Blocking the 2^k factorial designs
   - Fractional Factorial

12. Split-unit designs
   - If time permits

April: **Final exam** at a date to be determined