STA305H1S/STA1004HS, Summer 2017
Design of Scientific Studies

Revised 12 July 2017: typo fixed on page 3

Instructor: Mark Ebden
Email: mark.ebden@utoronto.ca
Office: SS 6026C
Office hours: Tuesdays from 4-5 pm (SS 6026C), Tuesdays/Thursdays after class (MS3154), and at other times by appointment

Teaching Assistants: Xiucai Ding (counts as two), Boris Garbuzov, and Kenny Guo. Boris's office hours will be on Mondays 7-8 pm in SS1091

Course webpage: Accessible through the learning portal: https://portal.utoronto.ca
Public page: Later in the course, some information will be duplicated here: http://mebden.com/STA305

Classroom sessions: Tuesday 18:00-21:00 and Thursday 18:00-21:00, in BA1160
Twelve sessions from Tuesday 4 July to Thursday 10 August

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-plot designs: Split plot designs will be discussed as an example of restricted randomization in the design of experiments.

Textbooks

There are no required textbooks for this course. The following are optional references:

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

One copy of reference 1 is 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.

References 2 through 5 are available electronically through the UofT library (i.e., electronic copies of both these textbooks are available at no extra cost).

Evaluation

Undergraduate and graduate students will be evaluated according to the following marking scheme. There may be opportunities for minor bonuses from time to time.

	Weight	Dates	Time
Three hand-in assignments	7% + 6% + 7% = 20%	13 July, 20 July, 3 August	Submitted electronically by 3 pm on Portal
Term test	35%	25 July	18:15-19:45
Final exam	45%	Expected August 15-18	Scheduled by Faculty

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

Assignment Due Date and Time

Assignments are due by 3 pm on the due dates and must be submitted electronically on the UofT Learning Portal.

Late assignments will be accepted within 48 hours after the due date.

- Late penalty for assignments: 10% will be deducted from assignments submitted online within 24 hours after the due date; and 20% will be deducted from assignments submitted between 24 and 48 hours after the due date.

- Assignments submitted more than 48 hours after the due date will receive a grade of zero except for cases with documented reasons beyond the student's control. In the case of a medical reason this must be documented using the UofT Student Illness or injury form. If an assignment is missed for a documented reason then the weight for the assignment will be transferred to the weight for the final exam.

- Email submission of assignments will not be accepted. If an assignment is submitted by email then it will NOT be marked.

It is strongly recommended that you do not try to upload or submit your assignment one or two minutes before the deadline. Based on past experience there is a good chance that your assignment will be marked late. This will not be considered a valid reason for submitting your assignment late. If your assignment is processed by Portal a few seconds or minutes late then the software will flag it as late and you will lose marks.

Term test and exam

The test will be written during class time in EX100 on 25 July.

You are allowed a one-sided 8-1/2"x 11" (standard letter size) aid sheet on the term test and a two-sided aid sheet on the final exam. You must bring your student identification to the term test and the final exam.

You will not need to know R syntax on the tests and exam, but you will need to know how to interpret output from R.

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.

Missed Tests

- If a test is missed for a valid reason, you must submit documentation to the course instructor.
- If a test is missed for a valid medical reason, you must submit the University of Toronto Verification of Student Illness or Injury form to your instructor within one week of the test.
- The form will only be accepted as valid if the form is filled out according to the instructions on the form.
- Important: 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.
- If a test is missed for a valid reason then the final exam will be worth that much more of your final grade.
- Other reasons for missing a test will require prior approval by your instructor. If prior approval is not received for non-medical reasons then you will receive a term test grade of zero.

Computing

We will use R for all examples. R is freely available for download at http://cran.r-project.org for Windows, Mac, and Linux operating systems. For the test and exam, you will need to know how to interpret output from R. We will support the use of R to complete the assignments.

R Studio is a good integrated development environment to R. It is freely available at www.rstudio.com/products/rstudio/

If you wish to use R at UofT then you may 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

In this course it is assumed that students have never used R before. You will be provided with the R syntax for all examples in lecture, which should be sufficient for you to do your assignments.

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 term test and the final exam.

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, which affects your privacy. 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 highly 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/utoronto.ca/summer2017/sta305hs/home

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
- E-mail should only be used for emergencies or personal matters

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. Use your utoronto.ca e-mail account to ensure that your message doesn't automatically go to a Junk folder and include your full name and student number.

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, 2017 Summer term
Course Schedule

Lecture	Topics	Assignment Due Date
1. July 4	Introduction, review of mathematical statistics	
2. July 6	Comparing two groups via the randomization distribution	
3. July 11	Design of phase III clinical trials and comparing two groups designs	
4. July 13	Causal inference in randomized experiments	HW #1 due July 13, 3 pm
5. July 18	Design of observational studies	*(Around now: Exam date/time ann'nced)*
6. July 20	ANOVA - Comparing more than two groups	HW #2 due July 20, 3 pm
July 25	**Midterm test: 6:15-7:45 pm**	
7. July 27	ANOVA - Comparing more than two groups	*(July 31: Last day to drop course)*
8. August 1	Factorial designs at two levels	
9. August 3	Randomized block designs, Latin Squares, Balanced Incomplete Block Designs	HW #3 due August 3, 3 pm
10. August 8	Blocking in factorial designs, fractional factorial designs	
11. Aug 10	Restricted randomization and split-plot designs	

Tentative List of Contents, by Lecture 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. Comparing more than two entities
   - Sample size estimation
   - Contrasts and treatment means
   - Multiple comparisons
   - Checking Model Assumptions

8. 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

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

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

11. Split plot designs