

STA 304H1 F/1003H F L0201, FALL 2019
SURVEYS, SAMPLING AND OBSERVATIONAL DATA

Warning: This section is independent of section L0101. You cannot replace one with the other.

Time: Tu 12-2, place MC 102, Th 11-12, place PB B150, web-site: on Portal.

Instructor: Dragan Banjevic (dragan.banjevic@utoronto.ca), office BA8139, tel: 946-3939, office hours: Tuesday 3-5.

Piazza (if you join): <https://piazza.com/utoronto.ca/fall2019/sta3041003/home>

Textbook: Scheaffer, Mendenhall, Ott: Elementary Survey Sampling (Seventh ed.).

Useful but not required: Lohr: Sampling: Design and Analysis.

Marking scheme: First test (20%, October 17, time (tentatively) 7-8, place TBA), second test 20% (November 21, time (tentatively) 7-8, place TBA), final exam 60% (3h, in exam period, June 19-26). Final exam is cumulative. You will make your formula sheet (Test 1, one page, one side, Test 2, two pages, one-sided, or one page, two-sided, final, four pages, one-sided, or two pages, two-sided; the formula sheet is cumulative, that is, every time you can add new pages; you cannot include any worked examples, only theoretical formulas and definitions, no properties; highlighting is allowed; **hand-written only**). **There are no make-up tests.** For a missing test without a reason (U of T doctor's note) you receive a zero mark. With a valid reason your mark will be adjusted. If you miss the first test, the second test weight will be adjusted. If you miss the second test, the weight of the final will be adjusted (warning: difficulty increases from the first test to the final; final exam covers complete course).

Tutorials: There are no tutorials, but you can come for help to the Stats Aid Centre (as of this week moved from Sidney Smith Hall) to the Public Health Sciences (PHS) Building at 155 College Street, in an open area (381) with a table and surrounding cubicles for meeting with students, before tests: date and time TBA. Some extra office hours before the final will be available. Initial class slides and sample tests and finals will be posted on the web-site, as well as solutions to most of recommended exercises.

Calculation: The course includes a lot of numerical calculation. We will use R software to support the class and sampling from real populations (data sets provided by the textbook), but **for you using R is optional**. Understanding of R outputs is required, on the level explained in the class. Still, you will need a basic scientific hand-calculator, with statistical functions, and experience in working with it (**start using it from the first day**) for tests and final. Inability to work with it will not be an excuse. Programmable calculators are not allowed on tests and final exam. Don't forget this.

Course outline: Almost all of the course material is covered by the textbook. Related to the basic level of the textbook, some theoretical results will be considered in more detail. The following is a tentative schedule for the course:

1. Sampling problems and notions (Ch 2), recommended exercises: 1-7, 28.
2. Basic concepts (Ch 3; 3.3 will be covered in Ch 8, not here), exercises: 2-8, 21.
3. Simple random sampling (Ch 4; 4.6 is optional),
exercises: 1, 2, 14-17, 18a, 20, 21, 23-28, 36, 38, 41, 42.
4. Ratio, regression, and difference estimation (Ch 6; 6.5 will be covered with Ch 5, below),
exercises: 1, 2, 6, 9, 16, 23, 26, 27.

5. Stratified random sampling (Ch 5; 5.9, 5.10, 5.11 are not covered),
exercises: 1-3, 5-8, 12-17, 24, 26, 27.
6. Systematic sampling (Ch 7), exercises: 3, 4, 8, 21, 25, 27.
7. Cluster sampling (Ch 8; 8.8 is not covered),
exercises: 2-5, 8, 9, 16, 17, 20, 24, 25, 26, 27.
8. Two-stage cluster sampling (Ch 9; part of 9.6, on pp 301, 302, is not covered),
exercises: 2-4, 6, 9, 10, 14-16.
9. Supplemental topics, nonsampling errors (Ch 11.1, 11.4, 11.8 are covered, and maybe more, if
time allows), exercises: 1, 13, 14.