

**STA220H1F - The Practice of Statistics - Summer 2015**  
**Course homepage accessible from the U of T Portal at <http://portal.utoronto.ca>**

This course (in combination with STA221) presents the basic statistical methodology used in many fields of application. It emphasizes conceptual thinking and the basic techniques that will be useful to students who seek to gain an understanding of the use of statistics in their own fields. Students seeking a more theoretical or mathematical development of the subject should consider taking STA 257+261. If you wish to pursue further studies in Statistics, you should follow up this course with STA255, which will provide some of the theory needed for such study.

**Tutorials - starting May 14**

Tutorials meet for two hours per week, on both Tuesday and Thursday 6-7pm (followed by 3 hour lectures – ugh!). The major purpose of tutorial is discussion and review of course material; no new material will be covered. Bring to tutorial your solutions for the current tutorial assignment, which will be posted at the course web page.

**Assignment #1 (to be posted at the web site) is due at your first tutorial.** These assignments will be discussed at tutorial.

**There will be some brief assessment grade at each tutorial (0 – 5 scale)** based on either a short quiz or classroom discussion/participation or submission of part of the current assignment. These are meant to be easy marks to encourage you to keep up with the work, and are not very meaningful feedback on your achievement level or preparedness for tests. At the end of term, your TA will submit a 0 -10 grade for you. We expect the majority of students who make a steady reasonable effort to receive at least 9 marks out of 10. Don't throw away these relatively easy marks, as sometimes they make the difference between passing and failing the course.

**The location of your tutorial will be posted on the course web page** prior to 10am, May 14. Records are kept by tutorial so please attend the one you are properly enrolled in, **to prevent missing marks.**

**If you miss a tutorial/quiz due to illness, late enrolment, etc., please discuss with your TA, and not your lecturer or the course coordinator.** Your TA has discretionary power to deal with, and adjust tutorial grades for, such things. If ill, bring some proof for your TA.

**Course Content**

Content, emphasis, etc. of the course is defined by means of the lecture - *not the texts*. It is important to attend all lectures, as there is normally no simple way to make up for missed lectures. There will also be many lecture examples using Minitab, the statistical software chosen for this course. *Important announcements, problem sets, handouts, practice tests, tutorial and aid centre info, etc., will be posted on the course homepage, or occasionally emailed out. So check the web site (via the portal) and your university mail account regularly.*

**Required Text**

**Stats: Data and Models, 1<sup>st</sup> Canadian Edition**, by DeVeaux, Velleman, et al (Pearson). Do not purchase U.S. editions of the text. A student solutions manual (odd exercises only) is also available, which sometimes adds additional details to the odd exercise solutions found in the back of the textbook. You can avoid some frustration, by ordering online from the bookstore: [www.uoftbookstore.com/online/](http://www.uoftbookstore.com/online/) Things move fast in the summer, so get ahold of the textbook asap. There are also some copies on reserve at various libraries on campus.

New copies of the textbook also give you an access code for the web site [www.mymathlab.com](http://www.mymathlab.com) which has additional learning aids such as online quizzes and exercises, interactive applets, podcasts, animations, and the above-mentioned student solutions manual. The required access code for this course will be posted at the course web page.

**Additional Sources of Help (starting May 13 or 14)**

Your primary source of help with difficulties is your TA in the scheduled tutorial, but additional assistance will be available at the **Statistics Aid Centre**, Room 1091, in Sidney Smith Hall and/or at the computing sites. A schedule will be posted at the course website. Your own TA will be available for help every week, but feel free to drop in on any Sta220 TA present.

#### Additional Help

There are electronic discussion forums at the course web page where you can communicate with other students in the course, as well as with the course lecturers (please use in place of email). For very general administrative queries, the Statistics departmental office can be reached at [stats@utstat.utoronto.ca](mailto:stats@utstat.utoronto.ca) or 416-978-3452.

#### Evaluation

Your tutorial grade is worth 10% of your final grade. The remaining portion of your grade will be equal to either  $\{0.35 \times \text{Term Test} + 0.55 \times \text{Final Exam}\}$ , or  $\{0.15 \times \text{Term Test} + 0.75 \times \text{Final Exam}\}$ , whichever is higher. This allows me to partially down-weight the term test if you show improvement by the time of the final exam. Consider it a second chance if you do very poorly on the term test.

The tentative date and time of the term test is: June 2, 6pm, rooms: TBA (followed by a lecture; sorry)

Though tests/exams place a minor emphasis on formulae, you are allowed a one-sided 8-1/2"x 11" (standard letter size) hand-written aid sheet on the term test (two-sided on final exam). **Programmable calculators are not permitted on the test or exam. Be sure to bring your student identification to the term test as well as the final exam.** The final exam is 3 hours long and will consist of 30 multiple-choice questions (5 choices each). A correct answer receives 3.33 marks. An unanswered question receives 0.67 marks. The term test is mostly not multiple-choice and will be 90-100 minutes long.

#### Missed Tests

There are **no make-up tests**. Should you miss the term test due to illness, you are required to submit, within one week the proper University of Toronto "Verification of Student Illness or Injury" form. For more information, go to <http://www.illnessverification.utoronto.ca>. *The test's weight will then be shifted to the exam. If this documentation is not received, your term test grade will be zero.*

#### Calculators

Hand calculators are cheap and useful. Any cheap one with a square root button and basic memory will do. Mean, standard deviation, sum, and sum of squares keys may save you a bit of time on occasion, but we do not recommend the purchase of expensive calculators to get keys with special statistical calculations. Tests and exams will be designed so that those calculators give no advantage. We emphasize the use of Minitab software for doing any tedious or complex calculations.

#### Software and Computing

Students will be using, on a weekly basis, the Minitab Statistical Computing Software, either at home (if you purchase or rent the software), or at the CQUEST or SS computing facilities at the University of Toronto. Here are your options:

- 1) Get a computing account at [www.cquest.utoronto.ca](http://www.cquest.utoronto.ca) and use Minitab at their rooms in RW building. Info is posted at the web site about accounts, rooms, hours of operation, etc.
- 2) Use the computing room on the basement level of SS, where Minitab is installed.
- 3) Rent Minitab Express at [www.onthehub.com/minitab](http://www.onthehub.com/minitab) and install on your Mac or PC. A free 30 day trial version is also available there. This is an abbreviated version of the full Minitab software, newly and specially designed for use in introductory Statistics courses. More info may be found at [www.minitab.com/academic](http://www.minitab.com/academic)
- 4) If you can run Windows software, rent Minitab (full professional version, release 17) at [www.onthehub.com/minitab](http://www.onthehub.com/minitab) A free 30 day trial version is also available. You may also be able to find

someone selling a used copy on CD of the abbreviated student version of release 14, which had been used in past years for this course. Any release 14+ is fine.

You will use this software to analyze data sets used in the text exercises. You will use Minitab on nearly every assignment. The data sets used in text exercises can be found on the DVD accompanying the textbook, or at the textbook's web site [http://wps.pearsoned.ca/ca\\_ph\\_deveaux/204/52457/13429053.cw/index.html](http://wps.pearsoned.ca/ca_ph_deveaux/204/52457/13429053.cw/index.html) or for those who have access, at [www.mymathlab.com](http://www.mymathlab.com). **Always bring to tutorial the full computer output**, along with your written answers. Computing problems frequently arise, so do your assignment early, and be prepared to come back another day or time, if the printer gets jammed, etc. Expect no special consideration by your TA for 'technical' problems'.

Students sometimes ask about using other software. Some homework problems may be especially designed for Minitab, so use Minitab on those occasions. At other times, it may well be possible to do the exercise using some other software, and this should be okay (check with your TA too). Excel is lousy, but internet based Statcrunch is not bad. Generally speaking, it is best if we all stick to the same software. And tests will include typical Minitab output.

### Academic Offences

**Academic offences are unacceptable** and harm everyone. E.g., some students have been known to alter a test paper after it is graded, and then re-submit it hoping for more marks - really! But, offenders are caught, and **sanctions can be severe** - zero in the course with annotation on the transcript for several years; suspension for a year; even expulsion. Various measures, announced and unannounced, will be taken throughout the year to reduce their incidence and to ensure successful prosecution (photocopying of graded tests, multiple versions of multiple choice exams, etc.). And please carefully note the following:

- ❖ Requests for a test remark may only be considered if you **write your test in ink**.
- ❖ **Grading oversights** such as addition errors and overlooked work must be reported to your TA *immediately* upon receiving your test paper at class. *So check it over as soon as you get it back.*

Feel free to discuss weekly assignments with others, but in the end, be sure to do all the computations, work and final write-up by yourself (i.e. plagiarism is unacceptable).

### Accessibility Needs

The University of Toronto is committed to accessibility. If you require accommodations for a disability, or have any accessibility concerns, please contact Accessibility Services as soon as possible: [disability.services@utoronto.ca](mailto:disability.services@utoronto.ca) or <http://studentlife.utoronto.ca/accessibility>

### Final Comments

For most students, Statistics is a new subject. Like most new subjects, it seems difficult at first but simple in retrospect. Many new concepts will be introduced and built upon. The course will be far easier and more useful if you master concepts and terminology as they are introduced - *similar to learning a new language*. Doing the exercises is not like calisthenics or jogging, which are good for you no matter how your mind wanders as long as you go through the motions. *Think about the objectives of the study, why a particular method was used, and what has been learned about the original questions that motivated the study!*

The professor and TA's are there to help. Ask questions! If you have any comments or suggestions about the course, please tell us. We want to know what is good; we want to know when problems are developing. Remember that it takes time for information to travel and for action to be taken. Therefore, if you become aware of any problems, mention this to your instructors right away.

## STATISTICS 220F (SUMMER 2014) LECTURE GUIDE – TENTATIVE

Textbook (SDM 1<sup>st</sup> Cdn edition by DeVeaux, et al) chapter references are in parentheses.

*Do many of the odd problems in SDM for practice (the answers for these are at the back of the text, with longer solutions in the students solutions manual)*

**Lecture 1:** Introduction to course, overview. Cases and variables, categorical & quantitative variables. Worksheet organization of data. Relative frequency distribution. Bar charts & Pie Charts for categorical variables. Two way tables; marginal and conditional distributions. Simpson's paradox. Introduction to Minitab. Quantitative variables: dotplots and histograms. Shapes of distributions, outliers. (2-4)

**Lecture 2:** Stem and leaf plots. Timeplots. Summation notation. Sample mean, median, mode (modal class), trimmed mean. Resistant measures. Standard deviation. The empirical rule. Calculating mean & standard deviation for grouped data. Percentiles. 5-number summary & boxplots. Comparing distributions. (4 -5)

**Lecture 3:** z-scores. Linear and non-linear transformations of data. Density curves and the normal distribution. Normal probability (quantile) plot. Bivariate data: Scatterplots & correlation. Cautions. (6-7)

**Lecture 4:** Regression and the least-squares line, coefficient of determination, residual plots, outliers & influential observations; lurking variables, association & causality. (7-9)

**Lecture 5:** Collecting data: Observational studies and randomized experiments. Designing experiments: comparison, randomization, blocking, factorial design. Sample vs population characteristics. Random samples: SRS, stratified, systematic, cluster & multistage. Sampling and non-sampling errors. Bias. (12, 13)

**Lecture 6:** Probability and relative frequency. Sample space, outcomes, and events. Rules of calculation for probabilities. Conditional probability (with a Bayes example via tree diagram). Statistical independence. Estimating probabilities by computer simulation. Two useful counting rules (combinations, sequential operations). (11, 14, 15)

**Lecture 7:** Discrete random variables and probability functions. Expectation (mean) and variance of random variables and linear combinations (e.g. sums, differences, averages). The binomial distribution, applications, mean and variance (using sums of Bernoulli variables), the sample proportion. (16-17)

**Lecture 8:** Continuous random variables and density functions. Normal approximation of binomial distribution. Sampling distributions. Distribution of the sample mean (and total). The Central Limit Theorem. (17-18)

**Lecture 9:** Margin of error, Confidence Intervals for  $\mu$  ( $\sigma$  known / large  $n$ ) and  $p$  (large sample), and sample size. Exact and Plus 4 confidence intervals for  $p$ . Statistical tests of hypothesis: observed and fixed levels of significance (P-value,  $\alpha$ -level). The z-test for  $p$  and for  $\mu$  ( $\sigma$  known or large sample). (19-20)

**Lecture 10:** Decision errors - types I & II, and power (with a z-test calculation). Testing hypotheses via confidence intervals. Statistical vs. practical significance. Multiple tests and data snooping. Test & C.I. for  $\mu$  with unknown variance: the Student t distribution. Robustness of t procedures (via simulations). Power of the t-test (using graphs or Minitab). (21, 23)

**Lecture 11:** Two independent samples: large sample z-test & C.I. for comparing the means. Small sample t-tests comparing two population means: pooled & unpooled variance procedures. Paired comparison t-test, and a nonparametric (distribution-free) approach - the sign test. Comparing designs: paired comparison vs. independent groups. Robustness, transformations, power (via Minitab). (24-25)

**Lecture 12:** Comparing two proportions with z-test. Inference for regression (time permitting). Final case study (video: the AZT trials). Overview of inferential procedures. (22, 26)