

## STA221H1S - The Practice of Statistics II - Winter 2015

Instructor: Augustin Vukov SS6024, 416-9784722, Office Hrs: M5-6pm, T3-4pm

This course is a continuation of STA220H, with emphasis on the basic statistical methodologies needed in a broad variety of fields: regression, experimental design and analysis of variance, chi-square and non-parametric procedures. The emphasis is on understanding the concepts and careful application of the basic techniques, using realistic data sets. The stats software of choice for this course is Minitab (available at the CQUEST and Sidney Smith computing labs on campus; more below).

### Tutorials (weekly, starting Jan 14)

Tutorials begin January 14 and meet every Wed 4-5pm. Rooms: TBA at the course website. Weekly problem assignments (usually requiring Minitab or other software) will be posted at the portal course website. They are due at tutorial, for discussion and quizzing. *There is an assignment due at the first tutorial.*

### Texts/Software:

**Required: The Practice of Statistics: Course Notes for STA221, by A. Vukov (Canadian Scholar's Press)**  
**Exercises will also be assigned from your STA220 textbook: Stats: Data and Models, First Canadian Edition, by DeVeaux, Velleman, et al (Pearson)**

The former is a complete set of course notes based on the lectures by A Vukov, including relevant Minitab commands, some statistical tables, and sample exam questions. I will reference *Stats: Data and Models* mainly for its exercises and data sets (also downloadable from [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), and also for supplementary readings. Both texts are available at the campus bookstore - Order online at [www.uoftbookstore.com/online](http://www.uoftbookstore.com/online).

Statistical software will be an *extremely important tool* for this course. Minitab is a top choice for academia as well as certain industries, and my choice for this course, due to both its simplicity and its power. You can either use computing facilities on campus at the CQUEST site in RW (go to [www.cquest.utoronto.ca](http://www.cquest.utoronto.ca) to sign up for a free account) or at the computing rooms in SS, or rent the software at [www.onthehub.com/minitab](http://www.onthehub.com/minitab), where you can also try it free for 30 days. Only Minitab Express works with the Macintosh OS. For Windows, you can choose either the full release 17 or Minitab Express. Minitab Express is a new product from Minitab, designed for introductory stats courses. However, you may use other software for assigned homework exercises if you wish. There may still be around some copies of the abbreviated student version of Minitab (release 14), and that is fine too.

### Statistics Aid Centre (from Jan 12)

Your primary source of help with difficulties is your TA in the scheduled tutorial, but additional assistance can be obtained at the Statistics Aid Centre, room 1091 in Sidney Smith Hall. (and also at the aid centre in New College) The course TAs will be on duty in the SS room at hours to be posted at the course website *For further assistance, try using the course online discussion forum.*

### Evaluation

Your final course grade will contain a tutorial/assignment component worth 10%. There will be a term test & final exam worth either 35% & 55% respectively, or 15% & 75% respectively, depending on which calculation gives you a higher course grade. In other words, I will down-weight the test if your performance improves at the final exam.

**The term test will be either on Feb 23 or March 2 at 3-5pm – date will be finalized soon and posted.**

**Programmable calculators are not permitted on tests and exam.** A one-sided 8-1/2"x 11" aid sheet, hand-written, is allowed on the midterm test (two-sided on final exam). **You must bring your student identification to the term test as well as the final exam.** The final exam will consist of 30 multiple choice questions (5 choices each) - correct answers get 3.33 marks, unanswered questions get 0.67 marks.

### Missed Tests

There are **no make-up tests**. Should you miss the term test due to illness, you must submit to your lecturer or to SS6018 (Stats office), within one week the proper documentation as required by the University. The test's weight will then be shifted to the final exam. **If this documentation is not received, your test mark will be zero.**

### Academic Offences

**Academic offences are unacceptable**, and harm everyone. 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 insure successful prosecution when they do occur (e.g. photocopying of students' tests, multiple versions of multiple choice exams). In addition, please note the following:

(i) **Oversights in marking on a test paper** (e.g. addition error, overlooked work) must be brought to the attention of the T.A. **immediately** - during the tutorial class when test papers are returned.

(ii) **Regrading requests** will only be considered for **term tests** which are written in **ink**

### Tentative Lecture Outline

[Course Notes *page references* in brackets] (SDM *chapter references* in parentheses)

**Week 1:** The chi-square goodness of fit test, for categorical data. The Poisson distribution. Testing fit of a hypothesized distribution. Chi-square test for association between two variables. Relation to z-tests. (ch26) [pp1-25]

**Week 2:** Regression with one explanatory variable: the regression model, estimating the coefficients by least-squares, the ANOVA and F-test. Inference for parameters. Residual plots. Linearizing transformations. (ch27) [pp.26-45]

**Week 3:** Prediction and Confidence Intervals at specified x. Testing lack of fit with replicates at some x's. Inference for the correlation. Introduction to regression with several explanatory variables. (ch27/30) [pp34-60]

**Week 4:** The multiple regression model. ANOVA table, R<sup>2</sup>, global F-test, residual plots, inference for individual parameters, prediction/confidence intervals for response. Interaction and polynomial models. The General Linear Model (GLM) and transforming to GLM (e.g. exponential models). Testing portions of a model via the Extra SS principle. (ch30/31) [pp57-80]

**Week 5:** Multicollinearity. Use of dummy variables. Stepwise regression & Model building. (ch31) [pp80-96]

**Week 6:** The completely randomized design and one-way ANOVA. Variation between and within samples, ANOVA table, F-test, CI's. Minitab output. Residual plots and transformations. Relation of one-way ANOVA to the two-sample t-test. (ch28) [pp.97-106]

**Week 7:** Planned comparisons (contrasts) among the means: t-test & F-test. Orthogonality and decomposing the Between Group SS. [pp106-116]

**Week 8:** Analyzing a quantitative factor with orthogonal polynomials. Post-hoc comparisons (Bonferroni, Fisher, Tukey, Scheffe methods). (ch28) [pp116-129]

**Week 9:** Orthogonal comparisons for a 2x2 factorial design. The general two-factor factorial arrangement. Interaction and additivity. ANOVA summary table, tests, CIs. Planned/post-hoc comparisons for the factors or treatments. (ch29) [pp130-153]

**Week 10:** Three factor experiments. The GLM and missing observations. Randomized block design. Factorial experiments run in complete blocks. Latin square design. [pp153-168]

**Week 11:** Analysis of Covariance. Fixed & random factors, crossing & nesting. Expected MS's and F-ratios. The split-plot and other designs. [pp169-187]

**Week 12:** Introduction to some nonparametric and computer intensive (re-sampling) approaches: the sign test, Wilcoxon tests and multi-sample extensions, Spearman's rank correlation; Bootstrapping standard errors and confidence limits. (ch32/33) [pp188-205]