## Faculty of Science Course Syllabus

## Department of Mathematics and Statistics

## Statistical Methods for Data Analysis and Inference

## STAT2080 / MATH2080 / ECON2280

## Summer B 2020 (06 Jul – 24 Aug)

**Instructors:** Ethan Lawler lawlerem@dal.ca

Jonathan Babyn jonathan.babyn@dal.ca

**Lectures**: Lectures will be pre-recorded videos posted on the BrightSpace course page. Additional formats (code notebooks, lecture slides, etc.) may be used to supplement the video lectures.

---

### Course Description

This is the usual sequel to STAT 1060.03 or STAT 2060.03. This course introduces a number of techniques for data analysis and inference commonly used in the experimental sciences. Topics covered include linear model building, multiple regression, analysis of variance, factorial designs, analysis of covariance using the general techniques for linear models, two and three way tables, along with logistic regression. A natural sequel for this course is STAT 3340.03.

### Course Prerequisites

STAT 1060.03 or STAT 2060.03 or DISP

The materials you are expected to be familiar with are the following: computation and use of various measures of central tendency and variability; preparation and interpretation of graphical displays of data such as boxplots, histograms and scatterplots; normal and t distributions and the use of tables for these distributions; the difference between populations and samples, parameters and estimates; the concept of sampling distributions and why they are important; construction and interpretation of confidence intervals; elements of hypothesis testing; formation of null and alternative hypotheses and computation and interpretation of p-values.

### Course Objectives/Learning Outcomes

The main objective of this course is to provide a solid grounding in practical data analysis using common statistical methods encountered in scientific research. To this end the central emphasis of the course is on Analysis of Variance (ANOVA) and Regression. A secondary objective is to become comfortable analysing data using the R statistical software.

Outcomes:

• Full understanding of the statistical comparison of two means using both parametric and non-parametric methods.

• Full understanding of one-way and two-way analysis of variance (including assumptions, setup, calculations of key quantities, interpretation, and post-hoc diagnostics).

• Full understanding of correlation as a measure of dependence, including both parametric (Pearson's) and non-parametric (Spearman's) measures of correlation.

• Full understanding of simple linear regression (assumptions, key quantities and formulae, implementation, interpretation and graphical assessment via residuals).

• Basic understanding of multiple regression (assumptions, key quantities and formulae, implementation, interpretation and graphical assessment via residuals).

• Experience with the statistical analysis of categorical/count data in one-way and two-way tables (e.g. chi-squared tests and contingency tables).

• Ability to use modern statistical software (R).

## Course Materials

There is a BrightSpace site for the course. This is where assignment information and announcements will be posted. The Brightspace site also contains a link to the course space on the LON-CAPA（Learning Online Network with Computer-Assisted Personalized Approach）server, where class notes, statistical tables and assignments can be found.

Lecture videos will be posted on Panopto Recordings, and office hours will be hosted on Collaborate Ultra. Both of these resources will be available through the course BrightSpace page.

There is no required textbook for this course. However, textbooks used in STAT1060 (*Stats, Data and Models* by DeVeaux, Velleman and Bock) and STAT2060 (*Probability and Statistics* by J. Devore) are excellent resources. Another recommended book is *Statistics Explained: An Introductory Guide for Life Scientists* by Steve McKillup.

Students will be required to use statistical software as part of this course. The software used in the course will be the state-of-the-art open-source statistical package R. R is available from www.r-project.org for Mac OS, Windows, and Linux. An online environment for R is also available at rstudio.cloud.

The LON-CAPA e-learning software will be used for assessments (as well as for disseminating assessment marks). LON-CAPA can be accessed from the BrightSpace course space, or directly at capa.mathstat.dal.ca. Further details will be provided on the BrightSpace course space.

### Tutorial Groups and TAs

Students will be given the option to self-enrol in virtual tutorial groups with approximately 50 students per group. These tutorial groups are intended to foster a connection between the students in a group. Each group will have a dedicated TA to answer questions, hold additional office hours, etc. These will be organized through the course BrightSpace page.

## Course Assessment

**Component -- Weight (% of final grade) -- Date**

Assessment 1 10% Friday 10 July

Assessment 2 10% Friday 17 July

Assessment 3 10% Friday 24 July

Assessment 4 25% Friday 31 July (cumulative)

Assessment 5 10% Friday 07 August

Assessment 6 10% Friday 14 August

Assessment 7 25% Monday 24 August (cumulative)

**Conversion of numerical grades to Final Letter Grades follows the Dalhousie Common Grade Scale:**

**A+** (90-100) **B+** (77-79) **C+** (65-69) **D** (50-54)

**A** (85-89) **B** (73-76) **C** (60-64) **F** (<50)

**A-** (80-84) **B-** (70-72) **C-** (55-59)

**Course Policies**

This course follows the university policy on "missed or late academic requirements due to student absence" for assessments:

**https://www.dal.ca/dept/university_secretariat/policies/academic/missed-or-late-academic-requirements-due-to-student-absence.html**

Students experiencing a short-term absence which results in an assessment being missed must do the following:

- Contact the instructor(s) by e-mail prior to the scheduled due date of the assessment,
- Complete a Student Declaration of Absence Form within 3 calendar days of the last day of absence.

Note that a maximum of one Student Declaration of Absence will be accepted and no make-up assessment will be offered. Instead, the weight of the missed assessment will be transferred to the corresponding cumulative assessment (assessment 4 or 7), such that the cumulative assessment will then be worth 35%. If a student misses an assessment and does not submit a Student Declaration of Absence, the mark for the missed assessment will be zero. If the first cumulative assessment is missed, the weight will be put on the second cumulative assessment whose weight will then be 50%. The second cumulative assessment is not eligible to be excused with a Student Declaration of Absence; conflicts with this assessment should be discussed with the instructor.

If there is a legitimate conflict with the time of a cumulative assessment students must inform the professors of this at least 3 weeks in advance of the assessment and provide details of the conflict.

There will be seven assessments. These will be online assignments delivered using the LON- CAPA software (see http://capa.mathstat.dal.ca). Late assignments are not accepted.

**Course Content**

Listed below in roughly chronological order are the topics to be covered. Note that these may be altered slightly as the term progresses.

- Study design, causal inference and inference to population
- The central limit theorem; hypothesis testing and confidence intervals
- Comparison of two means - paired samples and independent samples
- Comparison of two means - permutation test, Wilcoxon rank-sum test
- One-way analysis of variance
- Bonferroni method for multiple comparisons
- Assessing the model assumptions - residual plot
- Non-parametric one-way ANOVA - Kruskal-Wallis test
- Two-way ANOVA without interaction
- Two-way ANOVA, with interaction, Randomized block design, Post-hoc comparisons of means
- Categorical data, multinomial distribution and goodness of fit test
- Chi-square tests and contingency tables
- Scatterplots, Pearson's correlation, Spearman's rank correlation
- Regression and least squares estimators
- Coefficient of determination, Residual plots, remedies and transformation
- Inference in regression
- Multiple regression basics, hypothesis testing and inference
- Issues in multiple regression
- ANOVA using regression
- Special topics and review