# DATA 602 - Statistical Data Analysis Fall 2019 Additional Course Information

## Evaluation Methods

As indicated on the Data 602 Course Information Sheet, you course performance will be evaluated through the following assessment pieces:

| Method of Evaluation | Course Weighting | Dates |
|---|---|---|
| Assignments(5) | 55% | Sept.16; Sept.23; Sept.30; Oct.7; Oct.14 |
| Online Exercise : Sets (4) | 5% | To be Completed with WebWorK, Due on Sept.13 |
| Project Presentations | 40% | Tuesday, October 16th |

## Assignments

At the start of the term, you will be asked to completed a series of four exercises sets using the online homework system *WeBWorK*. These exercise sets have been created to enable you to practise the material appearing appearing in the **Online Module** folder (beneath **Contents**) in the DATA 602 D2L Page. These online assignments are worth a total of 10%, 2.5% for each. Each of these exercise sets go live on August 26th, at 8:00AM and you have until Friday, September 13th 11:59PM to complete. Excluding a handful of exercises, you have an unlimited number of attempts for each exercise set.

To access WeBWork, use this link (https://webwork.ucalgary.ca/webwork2/F2019DATA602L01). To login, use your ucalgary login and password.

You will be given a series of assignments (5) which are based on the material covered. These assignments are based on the lecture material and will be submitted via a Dropbox folder provided within the DATA 602 D2L page. Assignment problems can be worked on either individually or in groups. Your assignment write-ups should be your own work and must be completed in **R Markdown** and submitted as either an .html or .pdf file.

R Markdown is platform that can be used to communicate any piece of written work. It incorporates normal text, html, LaTeX, and "chunks" of R code in an all encompassing document. This document was created with R Markdown. For more information please consult the R Markdown page. (https://rmarkdown.rstudio.com/)

## Software

You will strictly using R Studio in this course. Please see the download instructions for R Studio (https://www.rstudio.com/products/RStudio/). Ensure you select the *Download R Studio Desktop* option from the *Open Source Edition* (left hand side of the webpage).

## Open Lab

There will be an additional opportunity for assistance on Friday eveings from 5:00pm - 7:00pm. ICT 517 will be staffed with your TA, allowing you to drop in and ask questions.

# Missed Components of Term Work

Shoud you be late in the submission of an assignment for reasons beyond your control, your final grade in the course will be assessed by a re-weighting the assignments that you have yet to complete. To be consistent and fair to all, this will apply to all students in the course. **There will be no makeup assignments.**

# Distribution of Final Grades

Please refer to the final percentage grade to final letter grade conversion scale appearing in the *provisional* Data 602 Course Information Sheet (http://people.ucalgary.ca/~jbstall/Fall2019/DATA_602_Fall_2019.pdf).

# Behaving…

Academic misconduct is a serious offense. Any student exhibiting behavior which is characterized as academic misconduct will be dealt with promptly and with great severity. See the 2019/2020 University Calender for more information. (https://www.ucalgary.ca/pubs/calendar/current/k-5-1.html)

# A Class-to-Class Look at DATA 602

Should you wish to read ahead to familiarized yourself with upcoming class material and prepare for the lectures, the following is an approximate class to class map, indexed with textbook sections that will correspond to the lecture material. Please note there is certain to be some flexibility here. I will notify you in class what sections of the text we are currently covering at that time.

**DATA 602 (9am/5pm) Fall 2019 Tentative Class to Class Guide**

**Day of**

September 5: Review of probability theory. Many examples will be completed in this first class that will encapsulate the probabilistic ideas reviewed - many of which you should have seen in the Statistics pre-requisite you satisfied to enter the program - in the five Online Modules. Exposure to R-studio and R Markdown.

September 10: Populations and Samples; Introduction to Statistical Inference. Data Visualization: Univariate Data Visualization (categorical = bar-graphs; numerical = histograms, density-plots, box-whisker plots),

September 12: Multi-variate Data Visualization (mixed variables, more elaborate data visualization methods with ggplot() to include more elaborate scatter diagrams, bar-graphs, density-plots, box-whisker plots). (During this time I will try to expose you to some data wrangling).

September 17: Exposure to various sample statistics (a 1-hour review through an interactive example to demonstrate the mean, median, variance, standard deviation, percentiles, quartiles, and proportions). Distribution of Sample statistics through SBI (mean, proportion, variance). Central Limit Theorem and application.
**Assignment 2 is due on September 18

September 19: Properties of Statistics (an unbiased compare to a biased statistic). Introduction to "condition free" inference - the Bootstrap Confidence Interval (mean, proportion, variance). Classical inference about the population mean, proportion through interval estimation.

September 24: Multi-population inferences and comparisons via (i) Bootstrap Intervals and (ii) 'classical' methods, emphasis on comparing two means and two proportions.

September 26: : Statistical modeling of data $X_i = \mu + e_i$ and introduction to single parameter hypothesis testing. Condition-free hypothesis testing of the mean and proportion through Permutation tests; classical hypothesis testing of the population mean and proportion. Exposure to Type I and Type II.

October 1: Multi-population inferences, Test of Independence for categorical variables. Exposure to Matched-Pairs Experimental Design.

October 3: Statistical Modeling of bivariate data $Y_i = \beta_0 + \beta_1 X_i + e_i$. Deterministic compared to probabilistic models. Least-squares estimation of the model. Model diagnostics. Inferences on the Regression Model, to include variance decomposition of the response variable.

October 8: Inferences on the Regression Model (continued). Prediction. Variance stabilization.

October 10: Infernece on the multiple-regression model.

October 15: Flex class (build in to ensure that I have covered all the topics above in their entirety).

October 17: Class Presentations.