UNIVERSITY OF CALGARY

FACULTY OF SCIENCE
Department of Mathematics and Statistics

Data Science 305 Computation Statistical Modelling

(see Course Descriptions under the year applicable: http://www.ucalgary.ca/pubs/calendar/)

Syllabus

Topic	**Number of hours**
Review of introductory probability theory: $A \cup B, A \cap B, A^c$, mutually exclusive events, independent events, conditional probability and Bayes Theorem	3
Random variables: Discrete random variable and its behavior, expected value and variance/standard deviation. Discrete Probability models: Binomial, Negative Binomial, Geometric, Hypergeometric, Poisson	**4**
Continuous Random Variables and their Probability Models: Uniform, Normal, Gamma (Chi-Square and Exponential (including relationship to Poisson), Pareto, Lognormal.	**6**
Populations, Samples, and Working with Data: Parameters and statistics. Univariate Data. Fitting data to models. Histograms and Density-curves. Summary statistics of data: $\bar{X}, \tilde{X}, S^2, S, Q1, Q3, \hat{p}$. Data transformations	**3**
Central Limit Theorem: Distribution of the sample mean and the sample proportion	**1**
Estimation: Point estimation and unbiased/biased estimation. Pivotal Quantities and Interval estimation. Maximum likelihood estimation	**4**
Hypothesis testing: Introduction. Type I, Type II, and Power	**2**
Simulation Based Inference: Bootstrap and Permutation tests	**2**
Parametric Hypothesis Testing: Small and large sample testing about the population mean, proportion, and variance/standard deviation	**3**
Parametric Hypothesis Testing: From independent random samples testing of (i) two population means (ii) two population proportions (iii) two population variances (Levene's test)	**4**
Bivariate Data and Simple Linear Regression: Scatter diagrams, correlation, and least squares estimation of the linear model with one predictor variable; model diagnostics and inferences on the linear model	**4**
Total Hours	**36**

Data Science 305
Course Outcomes

- Recognition of quantification of random events through the creation of a random variable; employment of probability foundations to design a probability model of a random variable.

- Differentiate between when to apply the various probability models covered in the course (Bernoulli, Binomial, Negative Binomial, Geometric, Hypergeometric, Poisson, Normal, Gamma and is special cases (Chi-square and Exponential)). In addition, demonstrate application of such probability models to compute probabilities with R.

- Statement and application of the Central Limit Theorem to both the sample mean and the sample proportion in order to consider the probable (and improbable) values of these statistics.

- Derive the probability distribution of a statistic via computational simulation and compute both its mean, is variance/standard deviation, and its bias.

- Distinction between a parameter and a statistic. Use simulation based methods as a basis for parameter estimation. Employment of pivotal quantities and their distributions as a parallel means for parameter estimation.

- Comprehend the scientific method of statistical hypothesis testing. This is to include the derivation of a statistical hypotheses, identification and subsequent application of a statistical test and the computation and interpretation of a P-value.

- Derivation of maximum likelihood estimators through simulation and computation.

- Model the existing synergy between two variables that are either numerical or categorical, through the employment of (i) least-squares estimation, resulting in the creation of a statistical model that predicts one variable based on the value of another or (ii) test of independence. Distinguish between numerical and categorical variables.

- Conduct a statistical hypothesis on the appropriate of the simple linear model with both the t-test and F-test. Awareness of the conditions of the linear model as well as diagnosis of their satisfaction. Confidence interval estimation of both the mean and an individual value of the response variable.