IVEY PhD
Business School
WESTERN UNIVERSITY · CANADA

COURSE SYLLABUS

Business 9702A – Multivariate Analysis
Fall 2023

Lauren Cipriano
Office No. 2361
Tel. Number 519-661-4181
lcipriano@ivey.uwo.ca

Faculty Assistant: Allison Bowes
Office No. 2319
Tel. Number 519-661-2111 (85497)
abowes@ivey.ca

Wednesdays 9:00am – 12:00pm
Location: TBD
(12 Sessions)

INTRODUCTION

Pre-requisites

- Probability at the approximate level of Ross S. A First Course in Probability. Pearson Education.
- Calculus at the level of a first year university course: single variable differential calculus and integration.

COURSE DESCRIPTION

Modern statistical concepts and methods developed in a mathematical framework: statistical inference, point and interval estimation, hypothesis testing, maximum likelihood estimation, large sample theory. This course introduces statistical regression models (ANOVA, linear regression, non- linear regression) with an applied focus. Students will learn the basic concepts behind linear and nonlinear statistical models, and apply them to the analysis of real data sets from various fields. A significant component in the course will be data analysis in R.

This course is an introduction to statistical theory, methods, and applications. The course should serve as a background preparation for many intermediate courses in statistical methods and applications.

We will cover the following topics:

- Limit theorems and sampling distributions
- Point estimation, confidence intervals, and fitting probability distributions
- Hypothesis testing
- Goodness of fit tests
- ANOVA
- Linear regression models
- Hypothesis testing and parameter estimation in linear models
- Model building and variable selection

- Non-linear models

LEARNING OUTCOMES/OBJECTIVES

At the end of this course, students will be able to:

- Understand the purpose, strengths, and limitations of various data collection and sampling strategies
- Recognize, describe, and calculate the measures of central tendency (mean, median, and mode), measures of location of data (quartiles, percentiles), and measures of spread of data (variance, standard deviation, and range).
- Apply and interpret the central limit theorem in the context of developing univariate statistics and confidence intervals
- Calculate the sample size required to estimate a population mean and a population proportion given a desired confidence level or margin of error
- Conduct and interpret hypothesis tests for a single sample including goodness of fit tests
- Conduct and interpret hypothesis tests for two or more populations, including matched or paired samples, repeated measures
- Perform, critically assess, and interpret simple and multiple linear regression with continuous, categorical, and interaction predictor variables
- Demonstrate an ability to evaluate the underlying assumptions of linear regression in order to evaluate the appropriateness of categorizing continuous variables, incorporating interaction variables, and/or transforming the dependent or independent variables
- Demonstrate an ability to evaluate outliers, including measures of influence and leverage on the properties and interpretability of regression
- Perform, critically assess, and interpret simple and multiple logistic regression with continuous and categorical predictor variables
- Demonstrate the ability to critically develop models and perform model selection procedures aligned with the purpose of the statistical model

METHODS OF EVALUATION

Weekly assignments (72%)

There will be **twelve weekly homework assignments**. Unless otherwise specified on the assignment, homework should be completed as an individual effort. Group discussion of assignment questions is permitted, but sharing of complete solutions or software code is not. If you worked with classmates on the assignment, please list their names at the top of your assignment.

All assignments must be completed. Homework assignments are due as digital submissions of a single PDF file at **8 am on Wednesday** (at the beginning of class). Late assignments will not be accepted.

Final project (28%)

The final project will be assigned in mid-November. The project is exclusively an individual effort and must be completed as such. The **final project is due as a single PDF file uploaded to LEARN Thursday December 21, 2023 at 3pm** but can be submitted earlier.

MATERIALS/REQUIRED READING

1. [Rice] Rice JA. Mathematical Statistics and Data Analysis. 3rd edition. Duxbury Press.
2. [C&H] Chatterjee D, Hadi AS. Regression Analysis by Example. 4th edition. John Wiley & Sons, Inc.
3. Additional readings or links to readings may be posted by the professor on the course LEARN website

Instruction will be done using the statistical software R (https://www.r-project.org/). The software is free and compatible with Windows, Mac, and Linux/Unix. The software is a standard for users both in and outside of statistics, particularly those that need to do programming of new methods.

1. R. The R Project for Statistical Computing. Available at: http://cran.utstat.utoronto.ca/ Select: Download R for Windows
2. R Studio. Available at: https://www.rstudio.com/products/rstudio/download/ Select: Install RStudio Desktop

It is expected that all students will be responsible for getting up to speed on the basics of the R package (or whatever statistical package they choose) on their own within the first week or so.

If you prefer to use a different statistical software or programming language/environment, you are free to do so. You will be responsible for seeing that you have sufficient access to software tools and technical assistance for the various topics covered in the course.

COURSE TIMELINE AND FORMAT

In-person class sessions on **Wednesdays** unless otherwise noted.

Schedule of activities

Week	Date	Session Topic and Readings
1	**Sept 13** *HW1 is due*	**Introduction to statistics, descriptive statistics, central limit theorem, and methods of sampling** *Readings:* • Rice, Chapter 10 (Summarizing data) • Rice, Chapter 5 (Limit Theorems) • Rice, Chapter 6 (Distributions Derived from the Normal Distribution) • Rice, Chapter 7 (Survey Sampling) *Help with R:* • Dalgaard Chapter 4 (Descriptive statistics and graphics) • Dalgaard Chapter 3 (Probability and distributions) *DataCamp tutorials (Optional support activities)* • Exploratory Data Analysis in R

2	**Sept 20** *HW2 is due*	**Properties of Point Estimators, Method of Moments, and Maximum Likelihood Estimation** *Recall:* • Binomial distribution (Rice 2.1.2) • Poisson distribution (Rice 2.1.5 and Rice 4.1 Example C) • Expected values and MGF (Rice Chapter 4, sections 4.1-4.5) *Reading:* • Rice, Chapter 8.1-8.5 (Estimation of parameters and fitting of probability distributions)
3	**Sept 27** *HW3 is due*	**Elements of a Statistical Test & Introduction to Hypothesis Testing** *Reading:* • Rice, Chapter 8.7-8.9 (Estimation of parameters and fitting of probability distributions) • Rice, Chapter 9 (Testing Hypotheses and Assessing Goodness of Fit) • Rice, Chapter 11 (Comparing two samples) • Sainani KL. Interpreting “null” results. PM & R. 2013. 5:520-523. • Baker M. Statisticians issue warning on P values. Nature. 2016; 531;151. • Benjamin DJ, et al. Redefine statistical significance. Nature Human Behaviour. 2018. 2:6-10. • Siegel S. Nonparametric Statistics. The American Statistician. 1957; 11(3):13-19. • Zimmerman DW. A note on preliminary tests of equality of variances. British Journal of Mathematical and Statistical Psychology. 2004; 57:171-181. • Sainani KL. Dealing with non-normal data. PM& R. 2012. 4:1001-1005. *Help with R:* • Dalgaard Chapter 5 (One- and two-sample tests) • Dalgaard Chapter 9 (Power and the computation of sample size) • Crawley, Chapter 8 (Classical Tests) *DataCamp tutorials (Optional support activities)* • Foundations of Inference in R • Introduction to A/B Testing in R • Inference for Numerical Data in R • Inference for Categorical Data in R

4	**Oct 4** *HW4 is due*	**ANOVA / Hypothesis testing with k ≥ 3 groups** *Readings:* • Rice, Chapter 12 (The Analysis of Variance) • Sainani KL. Misleading comparisons: The fallacy of comparing statistical significance. PM& R. 2010. 2:559-562. • Sainani KL. Dealing with longitudinal data. PM& R. 2015. 7:649-653. *Help with R:* • Dalgaard Chapter 7 (Analysis of variance and the Kruskal-Wallis test) • Crawley, Chapter 11 (Analysis of variance)
5	**Oct 11** *HW5 is due*	**Introduction to multivariable regression** *Readings:* • Rice, Chapter 14 (Linear Least Squares) • C&H, Chapter 2 (Simple Linear Regression) • C&H, Chapter 3 (Multiple Linear Regression) • Sainani KL. Understanding Linear Regression. PM& R. 2013. 5:1063-1068. *Help with R:* • Crawley, Chapter 10.1 (Regression) • Dalgaard, Chapter 6 (Regression and correlation) *DataCamp tutorials (Optional support activities)* • Introduction to Regression in R
6	**Oct 18** *HW6 is due*	**Model Structure: Confounding, Mediation, Moderation, Fixed vs. Random Effects** *Readings:* • C&H, Chapter 5 (Qualitative Variables as Predictors) • Theobald T. Students Are Rarely Independent: When, Why, and How to Use Random Effects in Discipline-Based Education Research CBE Life Sci Educ. 2018 Fall; 17(3): rm2. doi: 10.1187/cbe.17-12-0280 • Bates D, Mächler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models Using lme4. • Hayes, A. F., & Rockwood, N. J., (2016). Regression based statistical mediation and moderation analysis in clinical research: Observations, recommendations and implementation. Behaviour Research and Therapy, 1-19. http://dx.doi.org/10.1016/j.brat.2016.11.001

		• Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. Social Psychological and Personality Science, 8, 379-386. DOI: 10.1177/1948550617715068 *DataCamp tutorials (Optional support activities)* • Intermediate Regression in R • Hierarchical and Mixed Effects Models in R
	Oct 25	*Note: No class this week*
7	**Nov 1** *HW7 is due*	**Model Assumptions and diagnostics** *Readings:* • C&H, Chapter 4 (Regression Diagnostics: Detection of model violations) • Williams, M. N., Gomez Grajales, C. A., & Kurkiewicz, D. (2013). Assumptions of multiple regression. Correcting two misconceptions. Practical Assessment, Research & Evaluation, 18(11). *Help with R:* • Linear regression Assumptions and Diagnostics in R: Essentials. Statistical tools for high-throughput data analysis. Available: http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-assumptions-and-diagnostics-in-r-essentials/ *DataCamp tutorials (Optional support activities)* • Inference for Linear Regression in R • Handling Missing Data with Imputations in R
8	**Nov 8** *HW8 is due*	**Remedial measures** *Readings:* • C&H, Chapter 6 (Transformation of Variables) • C&H, Chapter 7 (Weighted Least Squares) • C&H, Chapter 8 (The Problem of Correlated Errors) • C&H, Chapter 9 (Analysis of Collinear Data)
9	**THURS Nov 16** *HW9 is due*	**Model selection procedures** *Readings:* • C&H, Chapter 10 (Biased Estimation of Regression Coefficients) • C&H, Chapter 11 (Variable Selection Procedures)

		• Sainani KL. Multivariate Regression: The Pitfalls of Automated Variable Selection. PM& R. 2013. 5:791-794. • Sainani KL. Explanatory versus predictive modeling. PM& R. 2014. 6:841-844. *Help with R:* • Dalgaard, Chapter 11 (Multiple regression)
10	**Nov 22** *HW10 is due*	**Categorical dependent variables** *Readings:* • Rice, Chapter 13 (The Analysis of Categorical Data) • Sainani KL. Understanding Odds Ratios. PM& R. 2011. 3:263-267. • Sainani KL. The importance of accounting for correlated observations. PM& R. 2010.2:858-861. *Help with R:* • Crawley, Chapter 16 (Proportion data)
11	**Nov 29** *HW11 is due*	**Logistic Regression** *Readings:* • C&H, Chapter 12 (Logistic Regression) • Sainani KL. Logistic Regression. PM& R. 2014. 6:1157-1162. *Help with R:* • Logistic Regression Essentials in R. Statistical tools for high-throughput data analysis. Available: http://www.sthda.com/english/articles/36-classification-methods-essentials/151-logistic-regression-essentials-in-r/ • DataCamp: Logistic Regression in R Tutorial (Article). Available at: https://www.datacamp.com/tutorial/logistic-regression-R
12	**Dec 6** *HW12 is due*	**GLM, Poisson Regression, or a special topic of class choosing** *Readings:* • C&H, Chapter 13 (Further Topics) *Help with R:* • Dalgaard, Chapter 15 (Rates and Poisson regression) *DataCamp tutorials (Optional support activities)* • Analyzing Survey Data in R

LEARNING TOOLS AND RESOURCES

R help resources:

1. An Introduction to R. http://cran.r-project.org/doc/manuals/R-intro.html
2. R Software Reference. http://r-dir.com
3. Quick-R. Accessing the power of R. http://www.statmethods.net/
4. Crawley MJ. The R Book. 3rd edition John Wiley and Sons. 2022.
5. Black K. R tutorial. http://www.cyclismo.org/tutorial/R/
6. Dalgaard P. Introductory Statistics with R. 2nd edition. Springer.
7. An Introduction to R. http://cran.r-project.org/doc/manuals/R-intro.html
8. Kerns GJ. Introduction to Probability and Statistics Using R. First edition. Download here
9. Each other.

ENROLLMENT RESTRICTIONS

Enrollment in this course is restricted to graduate students in the Ivey PhD Program, as well as any student that has obtained special permission to enroll in this course from the course instructor as well as the Graduate Chair (or equivalent) from the student's home program.

ATTENDANCE

Students are expected to attend class, arrive on time, participate substantially in class discussions, behave professionally, and to always exhibit mutual respect.

On-time attendance in all sessions is mandatory. If absenteeism has reached 25 percent (i.e., more than 6 classes), your absences will be reported to the Dean's Designate, the PhD Program Director. You may not be eligible to write the final exam or submit the final project.

This UWO policy is outlined at:
http://www.uwo.ca/univsec/pdf/academic_policies/exam/attendance.pdf

See also Western's Policy on Accommodation for Medical Illness at:
www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_medical.pdf

In the event of an illness requiring medical documentation, please see the PhD Program office for specific instructions. Note that medical documentation must meet Western's requirements and be submitted to the PhD Program office, not the course instructor. Any non-medical absences from assignments, reports, and examinations must be approved by the PhD Program office and accommodation for such absences will only be granted under extenuating circumstances.

NOTICE OF ABSENCE
If you are unable to attend class, please email the professor in advance. Submitting assignments on time remains your responsibility.

ACADEMIC OFFENCES: PLAGIARISM AND ACADEMIC INTEGRITY

Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at http://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_grad.pdf.

All required papers may be subject to submission for textual similarity review to the commercial plagiarism-detection software under license to the University for the detection of plagiarism. All papers submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (http://www.turnitin.com).

HEALTH AND WELLNESS SERVICES

As part of a successful graduate student experience at Western, we encourage students to make their health and wellness a priority. Western provides several on campus health-related services to help you achieve optimum health and engage in healthy living while pursuing your graduate degree. See https://www.uwo.ca/health.

Students who are in emotional/mental distress should refer to Mental Health Support at https://www.uwo.ca/health/psych/index.html for a complete list of options about how to obtain help. Additionally, students seeking help regarding mental health concerns are advised to speak to someone they feel comfortable confiding in, such as their faculty supervisor, their program director or program coordinator.

ACCESSIBLE EDUCATION WESTERN

Western is committed to achieving barrier-free accessibility for all its members, including graduate students. As part of this commitment, Western provides a variety of services devoted to promoting, advocating, and accommodating persons with disabilities in their respective graduate program.

Graduate students with disabilities (for example, chronic illnesses, mental health conditions, mobility impairments) are strongly encouraged to register with Accessible Education Western (AEW), a confidential service designed to support graduate and undergraduate students through their academic program. With the appropriate documentation, the student will work with both AEW and their graduate programs (normally their Graduate Chair and/or Course instructor) to ensure that appropriate academic accommodations to program requirements are arranged. These accommodations include individual counselling, alternative formatted literature, accessible campus transportation, learning strategy instruction, writing exams and assistive technology instruction.

A FINAL WELCOME AND REQUEST OF STUDENTS

I welcome individuals of all ages, backgrounds, beliefs, ethnicities, genders, gender identities, gender expressions, national origins, religious affiliations, sexual orientations, ability, and other visible and nonvisible differences. I consider this classroom to be a place where you will be treated with respect. All members of this class are expected to contribute to a respectful, welcoming and inclusive environment for every other member of the class. If it is appropriate to our learning and you feel comfortable doing so, I ask that you share your unique point of view as we explore the course content.