MGST 754/PSYC 615
Linear Models I
Fall 2024

also taught Fall 2024 as "Analysis of Variance"
for the Department of Psychology

Instructor/Office	Dr. Andrew (Andy) F. Hayes (Scurfield 460)
Teaching Assistants/Office	Jonn Henke (Admin 135); Jenelle Morgan (Admin 143)
Telephone	Best to use my cell #, to be provided in class. Feel free to text if desired.
Email	andrew.hayes@ucalgary.ca
Office hours	By appointment or happenstance in person, or electronically via Zoom.
Web page	d2l.ucalcary.ca
Location	Mathison Hall 112 (lecture); Social Sciences 018 (lab)
Meeting times	Monday 13.00-15.45; Friday 13.30-16.15 (lab) September3 to December 6

Brief Course Description

Applications of the general linear model to research design and analysis. Topics include regression analysis, multiple and partial association, categorical predictors (ANOVA/ANCOVA), hypothesis testing and statistical inference.

Detailed Course Description

This course covers an introduction to the analysis of data using the general linear model. Topics include simple and multiple linear regression, partial association, multicategorical predictors, hypothesis testing and statistical inference, the interpretation of model parameters, and other topics in linear models as time allows. Focus is on conceptual understanding rather than mathematical computation. "Analysis of variance" (ANOVA) is just a special case of the general linear model. Thus, we will approach some of the topics often found in an ANOVA course from the perspective of multiple regression rather than as a distinct statistical method, which ANOVA is not.

Students will gain experience practicing their learning through various assignments and activities, often using statistical software. This course is the first of two required statistical methods courses in the Ph.D. program in Management and the graduate program in psychology and functions as the foundation course for further advanced study in applied statistical methods at the graduate level.

Course Objectives

By the end of this course, you will…

1. understand how linear models can be used to quantify association between a single dependent variable Y and a set of one or more regressor or predictor variables.
2. be able to use linear models to test and rule out various explanations for an association between two variables through the process of statistical control.
3. be familiar with how to quantify and test hypotheses about setwise association between two or more variables and a dependent variable while controlling for a set of one or more covariates.
4. know how to compare groups on a dependent variable that may (or may not) differ from each other on other variables correlated with the dependent variable.
5. see how analysis of variance (ANOVA) is just a special case of the linear model and why you don't much need a course on ANOVA if you understand linear models.
6. be able to apply the methods discussed in this course using readily available statistical software.
7. be prepared to tackle more advanced topics in data analysis in your education and career, almost all of which assume you already have familiarity with and understanding of linear modeling.

Textbook and/or Other Materials

The readings for this course will include a mixture of some chapters and sections of chapters from

Darlington, R. B., & Hayes, A. F. (2017). *Regression Analysis and Linear Models: Concepts, Applications, and Implementation.* New York, NY: The Guilford Press.

as well as from various journal articles now and then that are more for your interest than required reading. This book covers far more than we will be able to cover in this class, so consider it both a textbook for your learning as well as a future reference. The book and any journal articles are available online through the University of Calgary library. Some of the journal articles may be technical in nature, and due to differences in academic opinion, I will not always agree with everything published in articles that I provide. They are there for your interest to read as cursorily or as carefully as you desire.

If you choose to purchase Darlington and Hayes, the publisher usually has a sale during the Fall term for graduate students that includes free shipping. See https://www.guilford.com. The book is also available in electronic form from the publisher. You will likely also find some copies at the University of Calgary bookstore.

Darlington and Hayes is currently being revised. Now and then I may provide you with advance drafts of chapters from the second edition, depending on my writing progress during the academic term.

Prerequisite Knowledge

It is assumed you are already familiar with basic principles in the theory and practice of statistics, including measures of central tendency, variability, ideas in sampling, hypothesis testing, and inference. If you feel you do not have this background, please speak to the instructor. It may be better for you to do a review of the fundamentals before taking this class. Everyone in this class has access to online versions of some of my lectures in a graduate level introduction to statistics class that I taught at Ohio State University during the pandemic. See the D2L page for links to the videos and miscellaneous related content.

There is a good chance that you have been exposed to some of the ideas in this class in prior coursework as an undergraduate or graduate student. It has been my experience over 20 years that many students who think they know this material or feel that perhaps they don't need this class are overconfident, overstating their actual knowledge, and when put to the test don't actually know as much as they think they know. Furthermore, many instructors who teach material like we cover in this class are not experts and often convey information that is outdated or just plain wrong. The linear model is the foundation of so many more advanced procedures that a failure to grasp this material, or misunderstanding it, will put you at a serious disadvantage as you advance in your graduate education.

Lecture and Lab Components

This course has a lecture and a lab component. The lectures will be held in Mathison Hall 112 on Monday afternoon and focus on concepts, theory, and some practical computational matters. The lab component of the course is held on Friday afternoon in SS018. This component of the course will focus primarily on hands-on training using statistical software. The lab component is the primary place where you will get your hands dirty learning how to write statistical code in various computing languages under the guidance of the TAs. It is also a good place to have discussions about concepts. The course TAs are also graduate students and have knowledge and advice pertinent to the course and graduate school in general that will benefit you in one way or another. Use the TAs as a resource.

Computer Software

In this class you will get exposure to and practice with statistical programming and data management using SPSS, R, and Stata. All of these programs are installed in the computer lab in SS018. You are encouraged to download SPSS and R and install each of these programs on your personal laptop or other computer so you will have access to them outside of the lab. SPSS is freely available to students through a university site license. See https://ucalgary.service-now.com/it. R can be downloaded at no charge from http://www.r-project.org. Stata is not available for free installation on personal machines at this university, but you can get a good student price on it from https://www.stata.com/. I recommend making a small investment in your future by purchasing it or at least getting a 6- or 12-month license for this year and giving it a try.

All of this software can also be found installed in the PhD-student/CCRAM computing lab on the 4th floor of Scurfield Hall. Note that this space is *not* a

university computing lab like SS018. It is open to those affiliated with the Haskayne School of Business, but students from anywhere in the university enrolled in this class can feel free to use it when taking this class. It is open only during business hours and will typically be locked on weekends and after 4.30pm except to students affiliated with the Haskayne School of Business. **Note that this lab shares space with graduate student offices so please respect their privacy and limit noise when using the computers and software housed here.**

There is no required text or reading pertinent to the use of these programs. You are encouraged to seek out various books and web resources for additional information about these programs. On D2L you will find "**Using SPSS**," and "**Using R**" documents that will be helpful. There is also a "Using SAS" document, but SAS is not being taught in this class for a variety of reasons related to access and operating system requirements. The UCLA page referenced below has some good resources on the use of all of these programs.

For additional guidance, consider consulting one of many resources on the use of computer software for data analysis. Some suggestions are below, **some of which are available electronically** through the University of Calgary libraries.

Material on the web:

- https://stats.oarc.ucla.edu/ (look under the "Software" tab)
  This is an outstanding resource for all matters of statistical computing and modeling using popular software.
- https://www.princeton.edu/~otorres/Stata/ (learning about Stata)

Books on SPSS Syntax

- Grotenhuis, M. T., & Visscher, C. (2014). *How to use SPSS syntax: An overview of common commands*. Sage Publications.
- Collier, J. (2010). *Using SPSS syntax: A beginner's guide*. Thousand Oaks, CA: Sage Publications.
- Einspruch, E. L. (2004). *Next steps with SPSS*. Sage Publications.
- Boslaugh, S. (2005). *An intermediate guide to SPSS programming: Using syntax for data management*. Sage Publications.

Books on R

- Jones, E. (2023). *The R book* (3rd Edition). John Wiley and Sons.
- Davies, T. M. (2016). *The book of R*. No Starch Press.
- Zuur, A. F., Leno, E. N., & Meesters, E. H. W. G. (2009). *A beginner's guide to R*. New York: Springer.
- De Vries, A., & Meys, J. (2012). *R for dummies*. John Wiley & Sons.

Books on Stata

- Mitchell, M. N. (2015). *Stata for the behavioral sciences*. Stata Press.
- Longest, K. C. (2020). *Using Stata for quantitative analysis*. Sage Publications.
- Pevalin, D., & Robson, K. (2009). *The Stata survivor manual*. McGraw-Hill.

Course Delivery and Class Preparation

Course content will be delivered face-to face. There is no online or hybrid component to this course and the course content will not be available on recordings. In the course schedule you will find a list of recommended readings. Students should be prepared to pivot to online learning should current government or University protocols change. Please see other pages in this outline for additional details and schedules.

Course Workload

Generally, it is understood that students should spend two hours per week outside of class time for every hour of lecture. This means that for each course, students should expect to spend approximately 9 hours per week total on course work and lectures. This may vary by week depending on both the assessment schedule and on students' ability to manage their time.

Personal Computing and In-Class Laptop Use

This course is designed to be hands-on and operates at times much like a workshop. You will benefit most if you have a laptop or desktop computer with SPSS, Stata, and/or R installed so that you can follow along with computing lessons, though this is not required. A PDF of the course slides will be available for note taking and viewing on your computer during lectures (or for printing prior to class), as will data files that you can use to work through the examples presented in class.

Grade Scale

The Haskayne School of Business endeavors to ensure consistency of final grades across courses. Variations in distribution will always be considered by the instructor where called for by the performance in each individual class. Students do not have any 'right' to a certain grade but, instead, are responsible for earning grades. The instructor has unfettered discretion to evaluate student performance and assign all grades.

Your grade will be based on the weighted total percentage of points possible that are earned, with a maximum value being 100, in accordance with the table below.

Grade		Percentage	Grade Point Value	Description
A+	≥	95.0	4.0	Outstanding
A	≥	90.0	4.0	Excellent
A-	≥	85.0	3.7	Very good
B+	≥	80.0	3.3	Good
B	≥	75.0	3.0	Satisfactory
B-	≥	70.0	2.7	Minimum pass
C+	≥	65.0	2.3	Grades below B- indicate failure at the graduate level and cannot be counted toward degree requirements
C	≥	60.0	2.0	
C-	≥	55.0	1.7	
D+	≥	52.0	1.3	
D	≥	50.0	1.0	
F	<	50	0	

Grade Distribution

Date due	Assessment	Weight	Course Outcomes assessed
ongoing	Attendance/participation	20%	1 to 7
See below	Take home exercises	80%	1 to 7

Missed Assessment Policy

Graded in-class activities are provided only during class and cannot be made up if you are absent, so it is important that you attend class regularly. Attendance, of course, cannot be made up. You do not have to have a perfect attendance record to receive full attendance credit.

Assessment of Writing

Writing skills are not exclusive to English courses and, in fact, should cross all disciplines. The University supports the belief that throughout their University careers, students should be taught how to write well so that when they graduate their writing abilities will be far above the minimal standards required at entrance. Consistent with this belief, students are expected to do a substantial amount of writing in their University courses and, where appropriate, members of faculty can and should use writing and the grading thereof as a factor in the evaluation of student work. The services provided by the Writing Support, part of the Student Success Centre, can be accessed by all undergraduate and graduate students who feel they require further assistance. In this course, on each of the graded components, you will be required to convey your understanding of the concepts. You will be evaluated based on what you said in writing, not what you intended to say but did not communicate effectively.

Late Policy

In class and lab activities, when administrated, must be turned in during class time when the instructor says time is up. Assignments are due by the day and time provided on the assignment/project. An electronic copy should also be uploaded to the drop box on D2L by this date and time.

Unless permission has been given to submit an assignment or activity late, the penalty for late submissions in 1% for every business hour late, where a business hour is defined as any hour between 9.00am and 5.00pm Monday through Friday. Each 1% will be deducted from your mark from the assessment had it been turned in on time.

Assessment, Assignments, and Grading Weights

Your course grade will be derived from your performance on a variety of assignments, lab activities, and attendance, with each component weighted as per the schedule and description below.

<u>Lab and Lecture Attendance and Participation (20%)</u>

You are expected to attend class and labs, participate with your own questions when you have them, and contemplate the questions of others and my answers. At least some material not included in the book, or my lectures will be delivered in response to questions, and you will be benefit from hearing those questions and answers. Thus, you will learn merely by attending class regularly even if you are just a passive observer of others most of the time. Attendance will be taken, typically at the beginning of class as well as in labs. You do not need to have a

perfect attendance record to receive full attendance credit. Attendance also consists of participating in the labs in the form of completing various small concept and software-related assignments you will be given during labs now and then.

Take-Home Assignments (80%)
At four points during the term, you will receive a take home assignment to complete. The assignment will be distributed at least one week prior to the due date. The due dates are

Assignment #1: September 23
Assignment #2: October 21
Assignment #3: November 4
Assignment #4: December 2

Assignments are due at the beginning of class on the due date. Due dates may be adjusted depending on the pacing of the material in class. In no circumstance will an assignment be due earlier than the date listed, but the due date may be pushed back if the course gets behind schedule or it is otherwise warranted. Unless you are told otherwise, you *may* work as a team with one other student enrolled in this class when working through these graded assignments. In this case, you will turn in one response to the assignment with each person's name on it, and you will each receive the grade allocated to your response. It is academic misconduct to collaborate on the assignments with anyone who is *not* a part of your team (other than the course instructor or the course TAs) prior to the time and date the assignment is due. Such collaboration includes exchanging answers, electronically or otherwise, or other forms of casual or formal conversation related to the content of the assignment.

In some cases, answers will be right or wrong, but in other cases there is room for subjective grading based on presentation, thoroughness, and so forth. Writing quality will matter when your assignments are graded. Be specific, precise, attentive to detail, and careful in how you phrase your answers, as you will be graded based on your actual answer, not what you intended to say or said awkwardly. Submit something you will be proud to submit, not something to just get you by until the next deadline. Do not wait until the last minute to start the assignments, as procrastination will show in the quality of your work. Use Word or a comparable word processing program to complete assignments. Use the symbol font for Greek symbols when needed, and **learn to use Microsoft's Equation editor** or some other system for generating clean, crisp mathematical expressions. Be careful in your formatting of mathematical equations, and be aware of order of operations rules. Submit something presented neatly and that you will be proud to claim is a product of your thinking.

You are expected to turn in a hard copy of your assignment with all sheets stapled together, as well as upload an electronic copy to D2L. An assignment is determined to be late if it is not delivered by the date and time the assignment is due.

The answers for each question will be provided soon after the assignment is due, but not before everyone has turned in a response. It is up to you to check your responses with the official answer sheet. If you do not understand any inconsistencies between the official answers and your own, you may contact me for assistance. Frequently, we will discuss the assignments in class or lab after the due date has passed and everyone has turned in their assignment.

Attendance

Showing up is a big and important part of life. You are expected to attend regularly, and some of your course grade will be based on your attendance record, with another component of your grade determined by performance on activities that require you to be there to receive credit. Not attending class regularly is a very bad idea, as some of the examined material will be presented only during lecture, and many of the statistical computing techniques are not always easily found in the documentation or other readings. As a general rule, subjective decisions about grading on assignments are less likely to go in your favour if you appear not be putting in the effort to learn by regularly attending class.

Mathematics Anxiety

Often one of the student's greatest barriers to mastering material in statistics courses is fear of mathematics. Many students lock up with anxiety when they are asked to do any computation and this anxiety typically interferes with the ultimate goal of conceptual understanding. I hope you will not let this happen to you. In this class most of the computations will be done by computer, although during lecture some basic computations cannot be avoided. You will be shown formulas and expect to understand them. But you need not understand the mathematics of the formula so much as you need to understand how they are conceptually used. To be sure, you need to be comfortable with basic mathematical operations. This is graduate school, and you have chosen the scientific study of psychology, management, or whatever your field. You will have to think analytically and quantitatively throughout your days as a graduate student at this university. If this is something you do not feel up to, you probably don't belong here. You will be challenged in this course, but there is no reason why everyone can't do well. The best thing that you can do to enhance your likelihood of success is discarding all the baggage that you may be bringing with you into the course—fear, anxiety, a belief that you are no good with numbers, or that you are destined to fail.

With these words of encouragement, at the same time remember that this is a graduate-level course. I admonish Master's students with less experience dealing with the intensity and pace of graduate school, and even Ph.D. students with a Master's degree from another university, not to approach this course as if it were an undergraduate course. You will not succeed if you don't dedicate time and energy to reading and contemplating the material. You will probably find yourself working harder during your first years of graduate school than you have ever worked before.

Email and Communication with Instructor

There may be occasions where I will need to get in touch with you outside of regular class hours or you may need to get in touch with me. Email will usually be the first means by which contact will be initiated. However, you are free to wander by my office in Scurfield Hall. If the door is open, you can assume I am available for an impromptu meeting. It is important that you check your university email account regularly. If you do not use your university email address as your primary email account, please arrange to have your university email forwarded to your preferred account. We can also arrange a private Zoom meeting at a time of mutual agreement.

Academic Integrity and Rigor

Academic integrity and rigor are critical components of a university degree. Academic integrity is the foundation of the acquisition of knowledge and is based on values of honesty, trust, responsibility, and respect. The University of Calgary values ethical leadership and personal integrity and expects its faculty, staff, and students to live these values. In the online environment, certain additional measures will be put in place to help safeguard the integrity of assessments and the intellectual property of others, including the instructor.

Unless otherwise granted permission, assignments, projects, and other assessments are generally not collaborative activities and must be completed independently, without the assistance of others in the class, colleagues, personal statistical consultants, or the use of generative AI tools such as ChatGPT (except under the conditions documented in the "Use of Artificial Intellection" section below).

Use of Artificial Intelligence (e.g., ChatGPT)

If you use ChatGPT or other forms of generative artificial intelligence in this class, you may only do so on assignments or other grade assessments in which it is not prohibited, and you must disclose its use by highlighting in yellow sections of assignments, take home assessments or projects, or other written material that were generated either whole or in part by or with assistance of these tools. A failure to do so will be considered academic misconduct. In general, I recommend avoiding these tools and submitting your own work rather than work generated with the assistance of artificial intelligence. All other things being equal, material written with the assistance of AI will not be graded as favourably as content generated entirely on your own. And especially in statistical methods, AI tools often get things wrong and their output is hard for a someone learning about those methods to evaluate for accuracy.

Roles of the Teaching Assistant

Graduate teaching assistants are responsible for the lab component of the course, some of the grading, and helping you master the topics. Although teaching assistants will do their best to respond to your concerns and questions in a timely fashion, keep in mind that they are also students and have their own demands and schedules that may not always mesh with yours. So please be patient if they are not available to respond to your needs immediately.

Intellectual Property

Generally speaking, course materials created by professor(s) (including presentations and posted notes, labs, case studies, assignments, and exams)

remain the intellectual property of the professor(s). These materials may not be reproduced, redistributed, or copied without the explicit consent of the professor. **Posting of course materials to third-party websites such as note-sharing sites without the written permission of the instructor is prohibited.**

Tentative Nature of this Syllabus

Events that transpire over the term may require me to modify the administration of this course and therefore the syllabus. In the event I need to modify the syllabus, I will announce the modification in class, through email, and/or through D2L. Ultimately, it is your responsibility to keep up with any such modifications and be aware of current policies, deadlines, etc.

Course Outline Part B

The Course Outline Part B contains more generalized information for Haskayne and the University. You are responsible for reading and understanding all content in both parts of the outline. Part B can be found here

Tentative Course Schedule

The pace of my classes is governed not by the clock or calendar but instead by the needs and interests of the students and how the conversation evolves during the presentation of material. We complete a topic when we have covered it in sufficient detail and your questions have been answered and comments conveyed and addressed. Nevertheless, a very tentative schedule of topics is provided below for those who prefer to have some structure and would like to see roughly where we are headed and when. The dates below, if any, are "approximate" to allow flexibility. It is likely we will fall behind or, rarely, get ahead as the term progresses, and material may be added or cut as needed so that when the end of the term comes, we will have covered as much of the material below as possible, and perhaps even more.

The course is divided up into seven units, with the time dedicated to each unit being flexible and determined in part by the pace of discussion and questions asked during lectures. We will use the assignment due dates as a rough guide to scheduling. You will be told when we are transitioning into the next unit.

I recommend you read the readings for each unit several times as we work through the unit, for your understanding will grow by this repetition, and after concepts that may have confused you at first are clarified during lecture. You are advised to set aside time each day to read what you have not, and reread what you have.

UNIT 1: Regression Analysis Fundamentals

This unit introduces the fundamentals of linear regression analysis using the simple regression model. Topics include the least squares criterion, residuals, estimation and interpretation of model parameters, the correlation versus the regression coefficient.

Reading: Darlington and Hayes (2017) Chapters 1 and 2

UNIT 2: Multivariate Association

This unit addresses the use of the linear regression model for generating estimates of one variable from a set of regressors, with a focus on measures of multivariate association.

Reading: Darlington and Hayes (2017) Chapters 3, 7 (mostly section 7.4), and 8. (Notice that the readings for this unit are the same as for the unit on partial association. These chapters blend material from these three units. We will cover topics on partial association in the next unit).

UNIT 3: Partial Association

This unit covers the use and interpretation of regression coefficients, partial and semipartial correlation, and the concept of statistical control.

Reading: Darlington and Hayes (2017) Chapters 3, 7, and 8, Chapter 6 for your interest (optional)

UNIT 4: Statistical Inference

This unit addresses statistical inference, including inference about multiple correlation as well as for individual variables in a model. Also included are such topics as collinearity, misspecification, bias, and power.

Reading: Darlington and Hayes (2017) Chapter 4, Section 17.1

UNIT 5: Setwise Partial Association and Inference

This unit further develops your understanding of the fundamentals of linear regression analysis, including setwise multivariate and partial association and inference, including model comparisons.

Reading: Darlington and Hayes (2017) Chapter 5 (section 5.3)

Unit 6: Comparing Groups

Coding systems for representing two or more groups and the interpretation of regression coefficients when using different coding strategies. Omnibus inference about differences between groups ("ANOVA"). Also addressed is the comparison of groups when adjusting for other variables, adjusted means, and the equivalence between linear regression analysis and analysis of covariance ("ANCOVA").

Reading: Darlington and Hayes (2017) Chapter 5 (section 5.1), Chapters 9, 10

UNIT 7: Miscellaneous Topics

This unit will cover various miscellaneous topics as time allows. Some possible topics, among others, include effect size, measurement error, the hunt for influential or "irregular cases"; statistical assumptions underlying the use and interpretation of linear regression analysis; a preview of what's to come in MGST 799.26 "Multivariate Analysis".

Reading: to be announced