Multivariate Statistics - REMOTE

PSYX 522 – Autumn 2020

Course Location and Time

Skaggs Building 246
Thursdays 12:00 – 2:50pm

Instructor Information

Instructor: Daniel J. Denis, Ph.D.
Office: Skaggs Building 369
Phone: N/A
Email: daniel.denis@umontana.edu
Office hours (via **ZOOM**): Wed., 12-3 or by appointment.

Course Overview & Expectations

It is assumed that students entering this course have taken previous applied graduate statistics courses (or advanced undergraduate), and have a basic understanding of statistics and statistical inference from early concepts through to linear models such as ANOVA and multiple regression. Familiarity with calculus and matrix algebra is also encouraged, though not required. Prior exposure to statistical software (e.g., through independent study or completing assignments in a course, etc.) would also be an asset, though also not required.

Credits

3.0

Learning Outcomes

1. To provide the student with the opportunity to obtain knowledge of various classical and **modern multivariate statistical learning methods** as used in science in general, as well as approaches relatively unique to **psychometrics** and related areas. A heavy emphasis on the course is to provide a **foundation** for further learning via **problem-solving** through interpreting research in the field and **critically evaluating statements that you read** in an effort to discern the degree to which they are tenable given the quantitative evidence.

2. To gain experience in demonstrating that you are able to generate software output, as well as using computation as an aid to problem-solving and statistical and psychometric analyses.

Course Description

This course will survey topics which include requisite preparation in mathematics for multivariate statistics, regularization techniques in regression (e.g., ridge, lasso), discriminant analysis, logistic regression and support vector machines, principal components analysis, exploratory factor analysis, latent variable models, topics in psychometrics, cluster analysis, and the requisite philosophical reasoning to evaluate and assess the extent that these techniques are useful or helpful in drawing scientific conclusions. Understanding statistical methods (and the exercise of "obtaining software output") is one thing, **but understanding what you've computed and linking what you've produced to scientific statements and conclusions is quite another**. In this course, you will not only be exposed to these different methodologies (and learn how unified they are), you will also learn the benefits and limitations of using these techniques to draw scientific conclusions. **The wealth of quantitative methods in current existence could easily take many lifetimes to master in entirety.** This course will help instill the **fundamentals**, so that upon completing it you are in a good position to extend on your learning with virtually any techniques you wish to learn, now or in the future. Lacking such a foundation would make learning and interpreting (on a moderately deep level) new techniques virtually impossible, and you'd likely never see the wider forest.

Course Depth vs. Breadth

This course is necessarily a "breadth" course, as it is impossible to cover all of multivariate statistics or psychometric techniques in depth in the amount of time allotted for this course. For each of the methodologies that exist, there are many BOOKS written on these individual topics, and countless peer-reviewed journal articles. It is unreasonable to think that this course alone will make you an "expert" on any of the various multivariate techniques. Each research scenario, job, and data analysis is different (design issues are usually extremely difficult to figure out), and "cookbook" approaches to statistical analysis, even if somewhat helpful at times and having their place as a learning tool, can be dangerous if they are not used with caution. This course will introduce you to the underlying technical details of these procedures so that you have some background on the "anatomy" ("under the hood") of these structures before attempting to apply them to problems in research. **Most, if not all, of any statistical methods you will likely encounter in your career are based on the same fundamentals studied in this course. Seek to understand the fundamentals.** If you master the fundamentals and grasp the "big picture," your ability to learn new things in the future will be unstoppable. If you lack that foundation and resort to seeing all quantitative methods as distinct with no unifying foundations, you will forever be lost in the trees! Aim to see the forest and you will have a solid foundation for future learning and becoming a rigorous scientist. It's vital to understand and appreciate what you can vs. cannot conclude from statistics based on their application to data.

Required Texts & Sources

We will be drawing material primarily (but not exclusively) from the following 3 sources, either via lecture or through assignments. Additional class notes will be provided to fill some gaps that the instructor feels are necessary for a solid foundation (e.g., MANOVA in relation to discriminant analysis (including multivariate tests of significance), canonical correlation, and the Box-M test of covariances in MANOVA).

Mair, P. (2018). *Modern psychometrics with R*. Springer. (**PRIMARY for DEFINITIONS & CONCEPTS**)

James et al., (2013). *An introduction to statistical learning*. Springer. (**PRIMARY for DEFINITIONS & CONCEPTS**)

Irizarry, R. A. (2020). *Introduction to data science: Data analysis and prediction algorithms with R*. CRC Press.

Articles Reading List (select articles will be posted on Moodle)

Some or all of the following will be assigned during the course, or will be featured in student seminars/presentations.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annu. Rev. Psychol*, *53*, 605-34.

Bollen, K. A. & Noble, M. D. (2011). Structural equation models and the quantification of behavior. *PNAS*, *108*, 15639-15646.

Denis, D. & Legerski, J. (2006). Causal Modeling and the Origins of Path Analysis. *Theory & Science*.

Friendly, M., Monette, G. & Fox, J. (2013). Elliptical insights: Understanding statistical methods through elliptical geometry. *Statistical Science*, *28*, 1-39.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, *88*, 355-383.

Mulaik, S. A. (1987). A Brief History of the Philosophical Foundations of Exploratory Factor Analysis, *3*, 267-305.

Schervish, M. J. (1996). P-values: What they are and what they are not. *The American Statistician*, *50*, 203-206.

Wang, K. & Qin, X. (2014). Use of structural equation modeling to measure severity of single-vehicle crashes. *Journal of the Transportation Research Board*, 17-25.

Optional Texts & Resources

Johnson, R. A. & Wichern, D. W. (2007). *Applied multivariate statistical analysis*. New Jersey: Prentice Hall.

Rencher, A. C. & Christensen, W. F. (2012). *Methods of Multivariate Analysis*. New York: Wiley.

Tabachnick, B. G. & Fidell, L. S. (2000). *Using multivariate statistics*. Allyn & Bacon.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research*. Sage publications: London.

Hays, W. L. (1994). *Statistics*, 5th ed. Wadsworth Publishing Company, Belmont CA.

Kirk, R. E. (2008). *Statistics: An introduction*. Thomson/Wadsworth: Belmont, CA.

Field, A. (2009). *Discovering statistics using SPSS*. Sage Publications: California.

Field, A. & Miles, J. (2012). *Discovering statistics using R*. Sage Publications: California.

Upton, G., & Cook, I. (2006). *Oxford Dictionary of Statistics*. Oxford University Press. New York.

Morgan, G.A., Leech, N. L., Gloeckner, G. W. & Barrett, K. C. (2011). *IBM SPSS for Introductory Statistics: Use and Interpretation, 4th ed*. Routledge: New York.

Leech, N. L., Barrett, K. C. & Morgan, G. A. (2011). *IBM SPSS for Intermediate Statistics: Use and Interpretation, 4th ed*. Routledge: New York.

Office Hours

Office hours are held weekly. You are also strongly encouraged to e-mail questions to the instructor as they arise. Writing your question out in an e-mail, as clearly as you can (even if very long) is an **excellent** way to clarify what you do not understand, and often, you achieve a deeper understanding of the topic itself. **Please be as detailed and specific as you can in your e-mail** so I know how to frame my response to best suit your needs.

Evaluation

Your final grade will be based on the following:

1. **Assignments/Participation (25%)** – 1/2 Theory, 1/2 Application (Oral Discussion, Binary Grading)
2. **Student Seminar/Paper (25%)** *** Major Course Component
3. **Final Exam (50%)** 1/2 Theory / 1/2 Application

Student Seminar and Paper Submission (20 PAGES SINGLE-SPACED)

Each seminar will be approximately 15 minutes. You will submit the accompanying paper featured in your seminar summary. Seminars are an EXCELENT way to learn new topics, which you'll need to do for the rest of your career if you continue on in research or quantitative sciences. Seminars will be evaluated using the following criteria. **Plagiarism is not permitted**.

- Topic Knowledge & Expertise (30%)
- Level of Difficulty, Complexity and Depth (30%)
- Presence and Clarity of Exposition (20%)
- Organization, Delivery, and Thought Process (20%)

Student Seminar Topic Choices (Pick one of the following):

1. Support Vector Machines (James)
2. Bagging, Random Forrests, Boosting (James)
3. Correspondence Analysis (Mair)
4. Modeling Trajectories and Time Series. (Mair)
5. Preference Modeling. (Mair)
6. Networks. (Mair)
7. Regression Splines (James)
8. Decision Trees (James)

Letter Grade Distribution

Percentage	Grade	Percentage	Grade	Percentage	Grade
100	A	79	B +	59	D +
99	A	78	B +	58	D +
98	A	77	B +	57	D +
97	A	76	B	56	D
96	A	75	B	55	D
95	A	74	B	54	D
94	A	73	B	53	D
93	A	72	B -	52	D -
92	A	71	B -	51	D -
91	A	70	B -	50	D -
90	A	69	C +	< 50	F
89	A -	68	C +		
88	A -	67	C +		
87	A -	66	C		
86	A -	65	C		
85	A -	64	C		
84	A -	63	C		
83	A -	62	C -		
82	A -	61	C -		
81	A -	60	C -		
80	A -				

Course Guidelines & Policies

Disability Modifications

The University of Montana assures equal access to instruction through collaboration between students with disabilities, instructors, and Disability Services for Students. If you have a disability adversely affecting your academic performance, and you have not already registered with Disability Services, please contact Disability Services in Lommasson Center 154 or call 406-243-2243. I will work with you and Disability Services to provide an appropriate modification.

Academic Misconduct

You are expected to adhere to the university's Student Conduct Code with regard to academic integrity. Academic misconduct in this course will not be tolerated and will result in an academic penalty. **If you are suspected of cheating on a test or exam, you will receive zero on that test or exam and be asked to leave the class permanently**. In short, even if you do not know the answer to a question, you're much better off guessing than risking the chance of getting caught cheating.

Incompletes

Departmental and university policies regarding incompletes do not allow one to change "incomplete" grades after 1 year has passed since the "I" was granted.

Tentative Course Schedule (Subject to Change)

Date	Topic	Statistics	Psychometrics ("Lead up")	ASSIGNMENTS
20 Aug.	Introductions/Syllabus	Fox Appendices	SEM: Bollen (2002)	#1 ASSIGNED
27 Aug.	Foundations for Multivariate	Fox Appendices		
3 Sept.	Foundations for Multivariate	Fox Appendices		#1 DUE
10 Sept.	Intro to Statistical Learning (and Regression)	James, Chapters 2, 3	SEM: Bollen & Noble (2011)	
17 Sept.	Classification	James, Chapter 4		
24 Sept.	Resampling	James, Chapter 5	MODEL SEL: Denis (2020) CAUSAL MODELS: Denis (2006)	
1 Oct.	Linear Model Selection and Regularization	James, Chapter 6		
8 Oct.	Principal Components Analysis	James, Chapter 10	Classical Test Theory: Mair (2020) – Chapter 1	
15 Oct.	Exploratory Factor Analysis		Mair, Chapter 2	
22 Oct.	Path Analysis and SEM/Confirmatory Factor Analysis		Mair, Chapter 3	
29 Oct.	Item-Response Models		Mair, Chapter 4	
5 Nov.	Cluster Analysis & Mixture Regression	James, Chapter 10	Mair, Chapter 12	
12 Nov.	Student Seminars (8 @ 15mins)			
19 Nov.	**Final Exam (Take Home)**	**Due on or before Wednesday, November 25 @ 5:00PM E-mail to daniel.denis@umontana.edu**		

* Many of the topics in the course schedule above are also covered in Irizarry (2020), though mostly coding only with minimal theory or explanation. The following are some examples (there are many more) of cross-referenced sections (some of these materials may appear on assignments and can also be used in preparing your seminar) in the latter chapters. You can use these sections for coding ideas, but refer to the other two books (and class discussion) for theory and definitions, etc. Use Irizarry as a "catalogue" of coding options and "workbook" now and into the future and a "spattering" of topics to relate to what you've already learned, are currently learning, or will learn in the future.

Sections I – IV = Mostly Topics from Univariate Statistics, Data Analysis, and Visualization (and many coding/programming options as you may need them).

Section V

Chapter 29 – Cross-Validation and Bootstrap cross-references with Resampling Methods (James, Chapter 5)

Chapter 31 – Linear Regression cross-references with Chapter 3 (James)

Chapter 31 – Logistic Regression and Discriminant Analysis (including Quadratic Discriminant Analysis) cross-references with Chapter 4 (James)

Chapter 31 – Classification Trees and Random Forests cross-references with Chapter 8 (tree-based methods) in James.

Chapter 33 – Regularization cross-references with Chapter 6 of James.