DEPARTMENT OF RENEWABLE RESOURCES

UNIVERSITY OF ALBERTA

REN R 690 - Multivariate Statistics for the Environmental Sciences

Winter 2024 Syllabus – http://tinyurl.com/mv690/syllabus

eClass shortcut – http://tinyurl.com/mv690/eclass

Instructor Andreas Hamann
Office: 733 General Services Building
Email: andreas.hamann@ualberta.ca
Zoom: http://tinyurl.com/zoom-andreas
Office Hours: http://tinyurl.com/schedule-andreas or open-door

Assistant Sarah Namiiro
Email: namiiro@ualberta.ca
Zoom: http://tinyurl.com/zoom-sn
Office Hours: http://tinyurl.com/schedule-sn

Location Computer lab: Rm 866 General Services Building (or on-line)

Class times We 9:00 am to 12:00 pm (first class: Jan 10, last class: April 10, no final exam)

No classes February 21 (Reading week)

Credits *3 credits

Delivery This course has simultaneous in-person and on-line sections. All participants are free to switch back and forth between attending in-person and on-line at will. As such, there are no in-person requirements for this class. Note that the in-person section is typically larger, and therefore receives more face-time from the instructor and TA. However, experience has shown that you can take this class exclusively on-line without much compromise to the overall learning experience.

Relevant dates Add-drop deadline: January 19; Withdrawal with partial refund deadline: Feb 7; Final withdrawal deadline: April 5.

(1) Course description

This course focuses on visualizing and analyzing complex biological or environmental data for the purpose of prediction and scientific hypothesis testing. Covers classical and modern approaches to ordination and classification, direct and indirect gradient analysis, and models of ecological and environmental interactions. Participants engage in problem-based learning by analyzing data from their thesis research project. Students without a suitable dataset should enroll in two or more ★1 modules from the REN R 581/582/585/586 options instead. Prerequisite: ★3 introductory statistics recommended.

(2) Target audience

Senior graduate students. Ideally, you should have a dataset for analysis from your thesis work, but it is not a strict requirement. Participants should understand univariate statistics and have good proficiency working with data tables, descriptive statistics, basic graphics, and general quantitative analysis. Some knowledge of R or programming of any kind will be an asset.

(3) Course format

The course has a weekly lecture followed by discussion and lab sessions. Problem-based learning is emphasized through individual student projects and labs developed by student groups. There are nine classes that cover important multivariate methods and techniques. This is followed by additional lectures requested by students in support of their projects, student-led technical seminars, presentations, peer-feedback, and hands-on technical support.

On-line resources developed in response to the Covid-19 pandemic remain available to you. Here are my suggestions on how to make best use of them, if you want to take this course fully or partially on-line, if you have to miss a lecture or lab due to other commitments, or for effective preparation for quizzes:

- ***Lecture Videos***: If you can't attend the in-person lecture at the beginning of the weekly meeting, you can watch a corresponding video asynchronously on eClass. This has advantages and disadvantages. The videos are clearer and shorter due to careful scripting and editing; you can pause and rewind; my virtual whiteboard organization will be superior to what I typically produce ad-hoc in class. On the downside, you can't ask questions right away.
- ***Labs via Zoom***: We will have a parallel Zoom session to in-person labs for Q&A, labs and project support during regular class times. Labs follow-up on lecture material, so for effective learning you need to attend the in-person introductory lecture, or view the video prior to the lab. The teaching assistant and me will alternate between the on-line and in-person lab for support, depending on demand (usually there are more in person participants and demand).
- ***Quiz Preparation***: There will be quizzes corresponding to the nine core lectures and labs covering the most important multivariate techniques. The quizzes will include difficult questions that are designed to probe the limits of your understanding. They are meant to provide useful feedback, even for the most advanced students. Note that quizzes will count for very little of your grade (10%) and are simply a mechanism for you to keep on track, and to discover what you didn't know or couldn't deduct. Average class scores for quizzes are around 60%, roughly corresponding to an average final grade between B+ and A- for orientation. The quizzes are completely open-book with no restrictions of any kind, with the exception of help from another person. The best preparation for quizzes, and also improved comprehension, will come from video review at 1.5x speed, after some time since first exposure has passed, according to: Learning in double time (2021) *Applied Cognitive Psychology* http://doi.org/10.1002/acp.3899. Quizzes will be held at 9am a week after the lecture, before we move on to the next class unit.

In addition to video lectures and prepared labs, there will be substantial problem-based learning components for this senior-level graduate class:

- ***Projects***: You carry out an individual course project that you present as a website. You can view previous projects soon dating back two decades here: http://tinyurl.com/mv690/projects. Guidelines for carrying out your class projects are here: http://tinyurl.com/mv690/guidelines.
- ***Presentations***: For your draft project submission at approximately mid-term, you will create a 5-minute summary video in the format of a narrated PowerPoint slideshow. You can view examples from previous classes here: http://tinyurl.com/mv690/presentations. For the final project submission, you have a choice of in-person presentations or a video submission.

- ***Peer review***: You will submit a draft of your project around mid-term, and receive feedback from me and your peers, which will closely emulate the scientific peer review process. One difference though: this will be a group activity for you as reviewers, which will be a much improved experience: http://tinyurl.com/mv690/reviews and http://tinyurl.com/mv690/revisions.
- ***Seminars***: A second problem-based learning activity will be for you to figure out a quantitative method of any kind yourself. This can be an individual or group effort, with a 10-15 minute recorded video and brief lab that you share with the class: http://tinyurl.com/mv690/seminars.

(4) Textbooks

There are no required or recommended textbooks, and all course material will be provided.

(5) Computing and software requirements

This course does not require powerful computing hardware (unless your individual course project does). However, I recommend a proper workstation set-up, meaning a reasonably modern computer or laptop (PC or Mac) with one or two medium to large size monitors, full-size keyboard and mouse. In the past, I observed that students who fall behind in this class often try to do this course on marginal hardware or on small laptops. This will not work for this class.

If you need a proper workstation at home or in your office on campus, I can loan out Departmental equipment. Place your request here: http://tinyurl.com/renr-computers (log in with your CCID, ignore the stated RENR restriction and indicate RENR 690 in the comment field at the end).

On an experimental basis, you can also access a general purpose virtual computer, and one specialized deep learning workstation through eClass here: http://tinyurl.com/mv690/aws. However, this is experimental, so back your data and results up to your own computer.

Software that is needed (R, R Studio) or that may be useful for your course projects (e.g., ArcGIS or QGIS, InkScape, Microsoft Office 365, Camtasia, Adobe) is either open-source or it can be obtained as free (or low-cost) student version here: https://ualberta.onthehub.com.

(6) Marking and Grading

Your grade will be determined based on the course rubrics listed below. The final letter grade will primarily be based on the relative ranking of your overall percentage value. The class median for RENR 690 typically sits between a B+ and an A-. However, this may vary from year to year, as I also factor in what the class has achieved, relative to past cohorts of REN R 690.

I will let you know your relative class standing and your projected letter grade half-way through the course, after the draft projects have been graded. You can then decide if an extra effort is required to get you where you want to be in terms of grading.

	Percent	***Due Date***
Quizzes (worst quiz score or one missed quiz removed from grading)	10	TBA
--- Last date to discuss a course project idea ---		Feb 2
Draft Project (draft website, 5-minute presentation)	20	**Feb 27, 6 pm**
--- Last date to schedule a volunteer seminar ---		Mar 6
Your feedback on draft projects by your peers (group activity)	15	**Mar 8, 6 pm**
--- Receive feedback and projected course grade ---		Mar 19
Participation (attendance, engagement, seminar)	15	NA
Final Project (final website and 5-minute presentation)	40	**Apr 14, 6 pm**

(7) Course policies

Late submission of draft project: The most important deadline of this course is the draft submission deadline. Although I call it "draft", a complete project is expected based on the analytical tools you have acquired within the first month (standard graphics and ordinations). This is enough for you to understand your data and write-up a report for a scientific analysis from beginning to end. Your draft should include a brief background, rationale, objectives, results, interpretation and conclusions that correspond to the the objectives. By not meeting the draft submission deadline, you essentially compromise your ability to revise your project. If you miss this deadline by a significant margin, you will not be able to receive feedback from your peers, but I will provide late feedback at any time.

Late submission of peer feedback: This deadline is also important, so that you don't compromise other student's time to make revisions. You may receive deductions depending on the reason for the delay. If you have trouble meeting this deadline, you should communicate this to me right away so that we can make alternate arrangements.

Course policy for missed quizzes: If you miss a quiz, it should not be a big disaster as each individual quiz will only account for 1-2% of the grade. Furthermore, the quiz with your worst score (e.g., 0% for a missed quiz) will be dropped from the grading. As such, it does not make sense for me to spend my finite time to design an alternative quiz, should you miss one.

Course policy for technical issues: For on-line participation, you need stable internet, a reliable computer, and functional webcam (camera-on is required). I'm happy to make accommodations for lost work due to crashed hard-drives, etc. However, when the course ends, I need to give a grade on what was submitted. I recommend that you make frequent backups to the cloud or to an external hard drive to secure your ongoing work on the course project, especially toward the end.

Individual accommodations: if you are registered with University of Alberta Accessibility Resources (AR) for a disability or health condition, please let me know at your earliest convenience so that I can implement extra time for quizzes. There is no need to register individual quizzes directly with AR, I will use eClass for blanket accommodations.

Emergencies: If you have a major issue that prevent you from keeping up with schedules and deadlines, such as family emergencies, medical issues, etc., communicate this to me right away and we will determine the best course of action together. ***I do not need*** a Doctor's note or supporting documentation for any reason. You don't even need to share the exact reason with me. Just keep me in the loop on your schedule so that I can advise.

Use of AI-based tools: You are welcome to use Artificial Intelligence (AI) to help you with writing and coding for your course project, or even for help to answer quiz questions if you like. Note that you need to repeatedly critique the AI's output to get better than mediocre suggestions for bacground, rationale, objectives or hypothesis statements (but it can be useful to get inspired). You are largely on your own interpreting your data and drawing the correct conclusions from your analysis. Note that my course policy may be different from future UofA policies for theses, or journal policies for publications.

(8) General policies of the University of Alberta

Disabilities and health conditions: The University of Alberta is committed to creating work and learning communities that inspire and enable all people to reach their full potential. Accessibility Resources promotes an accessible, inclusive, and universally designed environment. For general information to register for services visit the University of Alberta Accessibility Resources webpage.

Recording of lectures: Audio or video recording, digital or otherwise, of lectures, labs, seminars or any other teaching environment by students is allowed only with the prior written consent of the instructor or as a part of an approved accommodation plan. Student or instructor content, digital or otherwise, created and/or used within the context of the course is to be used solely for personal study, and is not to be used or distributed for any other purpose without prior written consent from the content author(s).

***Plagiarism and Cheating*:** The University of Alberta is committed to the highest standards of academic integrity and honesty. Students are expected to be familiar with these standards regarding academic honesty and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behaviour (online at www.governance.ualberta. ca) and avoid any behaviour which could potentially result in suspicions of cheating, plagiarism, misrepresentation of facts and/or participation in an offence. Academic dishonesty is a serious offence and can result in suspension or expulsion from the University.

All students at the University of Alberta are subject to the Code of Student Behaviour, as outlined at: University Governance > Code of Student Behaviour. Please familiarize yourself with it and ensure that you do not participate in any inappropriate behavior as defined by the Code. Key components of the code include the following statements. 30.3.2(1) No Student shall submit the words, ideas, images or data of another person as the Student's own in any academic writing, essay, thesis, project, assignment, presentation or poster in a course or program of study. 30.3.2(2)c. No Student shall represent another's substantial editorial or compositional assistance on an assignment as the Student's own work." Students should speak with the course instructor about any questions or concerns about the code. Students should be particularly aware of the code as it pertains to internet and library research, use of previous class notes, reclamation plans of former students and interviews or discussions with others.

(9) Course Objectives

(1) Develop a useful, toolbox of analytical methods, using the R programming environment.

Objective / Technique	Reduce Complexity	Understand Relationships (multiple response variables) (in-) direct	Understand Relationships (multiple response variables) constrained	Analyze Treatment Effects	Classification	Prediction (single response variable)
Rotation or metric (scale of variable matters)	PCA CANDISC + clever graphics	PCA CANDISC + 2nd set of vectors	CANCOR[l] CCA[l] RDA[l]	MANOVA	DISCRIM (what differentiates classes?)	DISCRIM[c] LM[n] GLM[nb] [RSF]
Distance , non-metric (scale doesn't matter)	NMDS + clever graphics	NMDS + 2nd set of vectors	dbCCA[u] dbRDA[l] MRT[lu]	MRPP db/per-MANOVA	CLUSTER (develop classification system)	CART[mcb] [RF]

[l]) linear, [u]) unimodal relationships

[n]) numeric, [c]) class, [b]) binary variable

(2) Apply the methods to do some real science. There are a number of other options, but ideally, your course project should be part of your thesis. Make it more than just an exercise for a class.

(3) Have fun working with and learning from your peers.

(10) Lecture Schedule, Activities & Deadlines

Jan 10 – Unit 1: Course syllabus and overview, introduction to R
Syllabus, software essentials, data management, exploratory graphics

Jan 17 – Unit 2: Multivariate fundamentals: Rotation
Visualization and scaling of multivariate data, PCA

Jan 24 – Unit 3: Rotation continued, visualization of multivariate data
Rotation wrap-up, FACTOR analysis, graphics for visualization of multivariate data

Jan 31 – Unit 4: Working with predetermined groups of multivariate observations
Inferential statistics with class variables: CANDISC, LDA, MANOVA

Tell me about your class project idea and show me your data during the labs above or by individual appointment, by **Friday, Feb 2.**

Feb 7 – Unit 5: Multivariate fundamentals: Distances
Grouping observations using multivariate data: CLUSTER, NMDS, PCoA

Feb 14 – Individual project support meetings
Individual meetings to provide technical project support (no formal class on Feb 14).

Feb 21 – [Reading Week – No classes]

Draft projects are due on **Tuesday, Feb 27, 6pm.**
You will receive peer review assignments on or before **Wednesday, Feb 28.**

Feb 28 – Unit 6: Inferential multivariate statistics with distance metrics
Concepts of multivariate statistical testing: MRPP**,** perMANOVA

Peer feedback is due **Friday, March 8, 6pm**. You will receive the editor's (i.e. my) review and summary on or before **Tuesday Mar 19**

Mar 6 – Unit 7: Direct and indirect gradient analysis with two datasets
Analyzing dependencies among two variable sets: CANCOR, RDA, CCA

Let me know if you want to give a seminar on or before **Wednesday, March 6**

Mar 13 – Unit 8: Gradient analysis continued, "shallow" machine learning
Classification and regression trees: CART, MRT, RandomForest

Mar 20 – Unit 9: Machine learning with deep neural networks
Deep neural networks: Base R & Python Tensorflow / Keras

Mar 27, Apr 3, Apr 10 – Extra lectures, student-led seminars & labs, project support, or an optional in-person project symposium TBA

Final project revisions due on **Friday, Apr 14, 6pm.**