**Advanced Statistics: Multivariate** **Fall 2024**

# COURSE SYLLABUS

## Course Information

Instructor: Laure Spake
Email: lspake@binghamton.edu
Office Hours: Mondays 3:00-5:00pm, or by appointment
Office: S1 219

## Course Description

This course is aimed at graduate students in anthropology who will need to analyze the increasingly large and complex datasets collected in our field. The course takes an applied approach to statistics: it prioritizes teaching theory, "how-to" skills, and critical thinking over teaching the math underlying the models. Students will leave this course equipped with the skills needed to perform data management and wrangling, transformations, and statistical modeling. Skills are taught using *R*, a free and open source statistical computing language. No prior experience with *R* or with programming is necessary, but willingness to learn these skills is expected.

This is a fast-paced course: you will learn how to write basic scripts in *R*, get familiar with data wrangling techniques, refresh your knowledge of introductory statistics, be introduced to modern statistical techniques, and learn about replicability and open science practices – all in 16 short weeks! Please expect to come to every class, to practice at home, and to reach out during office hours if you need extra support. In return for your hard work, you will learn powerful tools that will serve you well in research and industry alike.

## Learning Objectives

Upon the completion of this course students will be able to:

1. Define and use statistical terminology when speaking about data
2. Operationalize research questions into clear statistically testable hypotheses
3. Decide which statistical tests are appropriate to evaluate hypotheses
4. Describe problems with null-hypothesis significance testing and identify alternatives
5. Interpret the results of statistical models in statistical and scientific terms
6. Build better statistical models based on a critical understanding of model term selection
7. Appraise quality of statistical testing and strength of evidence in research literature
8. Report results in technical and scientific language using statistical models and data visualizations

**Readings**

There is no required textbook to purchase for this course. Assigned readings will be uploaded to or linked from Brightspace. All assigned readings and resources are open educational materials.

Much of the assigned readings will be from the following texts, which you should find useful for supporting you in learning *R*. All three books are made freely available online by the authors, and you may borrow my copies during office hours. You may choose to purchase a copy if you like, however you certainly do not need to do so. The books are:

1. Wickham and Grolemund. *R for Data Science.* https://r4ds.had.co.nz/
2. Healy. *Data Visualization.* https://socviz.co/
3. Navarro. *Learning Statistics with R*. https://learningstatisticswithr.com/lsr-0.6.pdf

You will also find the following text useful for supporting your learning of statistical techniques, as well as learning the associated *R* code. This textbook is comprehensive and written in a very accessible and funny style. Whether in its *R, Stata*, or *SPSS* versions, this book has helped many a graduate student learn statistics. The book itself is quite spendy, but a hard copy is on reserve at the library, and I strongly suggest it:

1. Field, Miles, and Field. *Discovering Statistics in R.*

**Course Content Delivery**

The course relies on *R*, and we will do most of our work on *Posit Cloud*. There will be a link to the *Posit Cloud* space posted to Brightspace, and this space will be where you will receive learning exercises and assignments. Lecture slides and readings will be posted to Brightspace. Assignments and term projects will be turned in through Brightspace.

**Learning Assessment**

| | |
|---|---|
| **Assignments** | **30%**<br>Students will complete 5 assignments that will be due throughout the term. These assignments will reinforce key concepts and skills learned throughout the semester |
| **Quiz** | **20%**<br>Students will complete one quiz during the course of the term. This quiz will be on paper. It will not require you to write code. The quiz is aimed at evaluating your understanding of core statistical concepts as conveyed in the course. This include key definitions, selecting models and interpreting outputs. |
| **Term Project** | **50%**<br>Students will write a report in which they analyze real data. Students will complete a pre-registration document (**March 26**, 20% of the term project mark), present their analysis to the class (**April 30** during class, 20% of the term project mark), and submit a final .Rmd document reporting on their analysis (**May 3**, 60% of term project mark) |

## COURSE POLICIES

### Course Readings

Throughout the course, you will be assigned readings and additional resource materials that could include tutorials, videos, and statistical blog posts. These readings are meant to support your learning – they are your first line of defense in case you are struggling to understand a topic. Everything you need to know to succeed in the course will be presented during lectures, however, your understanding of the course content will be greatly improved if you work through the readings.

### Coursework Expectations

This course is a 4-credit course, which means that in addition to the scheduled lectures/discussions, students are expected to do at least **9.5 hours of course-related work each week during the semester**. This includes things like: completing assigned readings, participating in lab sessions, studying for tests and examinations, preparing written assignments, completing internship or clinical placement requirements, and other tasks that must be completed to earn credit in the course. **Be aware that you are expected to complete any in-class exercises not completed in class on your own at home and to ask questions if you do not understand the material.** I do not give answers to assignments in class or in office hours, so you should use the in-class exercises to make sure you fully grasp concepts.

### Late Work

I understand that graduate students are often juggling work, other courses, family obligations, etc., in addition to this course. If you anticipate difficulty meeting a deadline due to any non-course happenings in your life, please notify me at least 24 hours in advance if possible. I will work with you to set an alternate deadline that works for both of us. **Work submitted late without prior notice will not be penalized if it is submitted before I begin grading. However, if it is submitted after I have started or completed grading, it will incur a 10% penalty**. I will generally work with you to help you make up missed work, but I expect that you will take the lead in co-ordinating a schedule for making up work.

### Attendance and Make-Up Assignments

You are expected to attend all classes and to be present in class by the start time. Late arrival and early departure disrupt the entire class. If you cannot make a class, you are expected to self-teach the material. This is especially important in the first half of the course as material build on previous weeks and it will be expected that you understand prior weeks' materials in your exercises and assignments.

### Academic Integrity

Binghamton University's Student Academic Honest Code policies apply for this course. Be aware that it is expected and encouraged that you work with your peers throughout the semester. Self and peer learning is the best way to deeply understand statistics and to learn to code. However, you should submit your own work in your own words. Any copying of others'

work will be considered cheating and will be taken seriously. Penalties can include a zero on the assignment, failure in the course, and/or reporting of the incident to university authorities.

**Students Needing Accommodations**

Please make an appointment with the University's Services for Students with Disabilities and meet with me to discuss specific needs and accommodations.

**Course Schedule**

I reserve the right to make reasonable changes to the schedule as necessary including changing topics, assigning additional readings, changing due dates, reorganizing timelines, etc. I will do my best to ensure that if things change, changes will be in the students' favour (e.g. deadlines being pushed back, assignments being cancelled, rather than causing additional workload) however I cannot guarantee this.

**Grading Distribution**

See below. Grades below the half point are rounded down, and grades equal or above half point are rounded up (e.g., 92.4 = 92 or A-, while 92.5 = 93 or A).

| | | | |
|---|---|---|---|
| **A** | 94 and above | **C+** | 77-79 |
| **A-** | 90-93 | **C** | 74-76 |
| **B+** | 87-89 | **C-** | 70-73 |
| **B** | 84-86 | **D** | 60-69 |
| **B-** | 80-83 | **F** | 59 and below |

**CARE Team Information**

If you are experiencing undue personal or academic stress at any time during the semester or need to talk to someone about a personal problem or situation, I encourage you to seek support as soon as possible. I am available to talk with you about stresses related to my class. I can also put in a CARE team referral for you, or assist you in reaching out to any of the resources available on campus, including:

Dean of Students' Office: 607-777-2804

Decker Student Health Services Center: 607-777-2221

University Police for On-Campus Emergency: 911

University Counseling Center: 607-777-2772

Interpersonal Violence Prevention: 607-777-3062

Harpur Advising: 607-777-6305

Office of International Student and Scholar Services: 607-777-2510

Ombudsman, Main Campus: 607-777-2388

Services for Students with Disabilities 607-777-2686

## COURSE SCHEDULE

Course topics are below. Please note that the syllabus is subject to change. If substantial changes are needed, you will be notified ahead of time. All due dates are by midnight on Thursday of that week (the day after our class meeting).

| Date | Topic | Due |
|---|---|---|
| Aug 26 | Course introduction and getting started with *R* | |
| Sept 2 | * *Labour Day holiday class does not meet* * | |
| Sept 9 | Data visualization and statistical terminology | |
| Sept 16 | Statistical basics in *R* and data manipulation | Term Project Topic |
| Sep 23 | Introduction to NHST | Assignment 1 |
| Sept 30 | Paired and independent sample t-tests | |
| Oct 7 | Comparing multiple groups | Assignment 2 |
| Oct 14 | Non-parametric alternatives for comparing groups | |
| Oct 21 | Correlation and linear regression | Assignment 3 |
| Oct 28 | Odds, odds ratios and logistic regression | Assignment 4 |
| Nov 4 | Midterm, catch up, project work | Project Pre-Registration |
| Nov 11 | Dimensionality reduction (TBC) | Assignment 5 |
| Nov 18 | Power calculations (TBC) | |
| Nov 25 | Survival analysis and GLMMs (TBC) | |
| Dec 2 | Project presentations | Term Project due Dec 9th |