2023-24 Course Syllabus:

Quantitative Research Methods for Social Sciences I (RES501)

1. Course title
Quantitative Research Methods for Social Sciences I (RES501)

2. Professor
Zachary Garfield

Assistant Professor at Mohammed VI Polytechnic University. Ph.D. (Anthropology, 2019, Washington State University).

3. Course Description

This course provides an in-depth introduction to quantitative research methods for students in the social sciences. Throughout the semester, students will gain a comprehensive understanding of the changing statistical landscape and the application of statistical techniques in solving real-world problems. The course covers topics such as linear regression, logistic regression, linear model selection, resampling techniques, fitting non-linear functions, tree-based methods, principal components analysis (PCA), clustering, and reproducible research. Practical application is emphasized through hands-on lab sessions using RStudio for data analysis and manipulation. By the end of the course, students will be equipped with essential tools and skills to conduct data-driven research and effectively analyze social science data.

Prerequisites: None.

4. Course Objectives and Learning Outcomes

The primary objective of RES501: Quantitative Research Methods for Social Sciences I is to provide students with a solid foundation in quantitative research methods and statistical analysis techniques. By the end of the course, students will be equipped with the necessary knowledge and skills to understand, apply, and critically evaluate statistical methodologies commonly used in social science research. Additionally, the course aims to foster a deep appreciation for the role of data analysis in informing evidence-based decision-making and hypothesis testing.

Learning Outcomes:

1. Comprehend Statistical Concepts: Students will demonstrate a comprehensive understanding of key statistical concepts, including linear regression, logistic regression, resampling methods, tree-based methods, principal components analysis (PCA), and clustering.
2. Analyze Data Using RStudio: Through hands-on lab sessions, students will gain proficiency in using RStudio to perform data analysis, data wrangling, and data visualization, enabling them to work with real-world social science datasets.
3. Apply Quantitative Techniques: Students will be able to apply appropriate quantitative techniques to analyze and interpret social science data, effectively addressing research questions and hypotheses.
4. Critically Evaluate Research Findings: Students will develop the ability to critically evaluate research findings and published studies that utilize quantitative methods, recognizing the strengths and limitations of different statistical approaches.
5. Implement Reproducible Research Practices: By the end of the course, students will understand the importance of reproducible research and be able to implement best practices for organizing, documenting, and sharing their data analysis workflows.
6. Problem-Solving Skills: Through various practical exercises and assignments, students will enhance their problem-solving skills, learning to identify appropriate statistical methods for different research scenarios and apply them effectively.

7. Communication of Results: Students will be able to communicate their data analysis results and research findings effectively, both in written reports and through verbal presentations.
8. Ethical Considerations: The course will emphasize ethical considerations related to data analysis, ensuring that students understand the importance of maintaining confidentiality, avoiding bias, and responsibly handling data in social science research.

By achieving these learning outcomes, students will be well-prepared to undertake further studies in quantitative research methods and apply their skills in various academic and professional settings across social science fields.

5. Pedagogical Arrangement of the Course

10 class sessions are planned to present theory, illustrations and examples. TD sessions as well as office hour sessions will be organized to allow a better understanding and application of theoretical concepts. Students will conduct a research project using a data set of their choice (further details to be discussed) which will include giving an in-class presentation and submitting a mansucript.

6. Main References

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning: With Applications in R. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-1418-1. Free PDF. (https://www.statlearning.com/) (Referred to as ISLR)

Grolemund, H. W. and G. (2023). R for Data Science. Second edition Free online version: https://r4ds.hadley.nz/ (Referred to as r4DS)

`swirl` online tutorials: http://swirlstats.com

7. Complementary References:

Assigned readings delivered as PDF

8. Assessment and Grading:

Evaluation	% of Total Course Grade
Participation and Attendance	**10%**
Presentations & Oral Work	**15%**
Research project	**15%**
Mid-Term Assessment	**20%**
Final Exam	**30%**

Assessments	Details
Mid-Term Assessment	Instructions delivered on **Week 5** In-class exam on content of all course before Week 7.
Final Exam	**3 Hours** during the Finals period

	In class exam on content of all course. Paper documents are allowed, but no electronics (computer, phone, calculator, watch, etc.)
Assessment 3	Case-study presentation (see additional activity 2)
Assessment 4	Case-study presentation (see additional activity 2)
Further Assessments	None

9. Detailed Course Plan

Sessions	Contenu détaillé et évaluations
Week 1 (2 Hours)	Session Title The changing statistical landscape Session Details **Objectives** Lecture: Functions, the forward problem, and the inverse problem, statistifacl learning, assessing model accuracy. Lab: Introduction to RStudio, Fit a function to some data **Planning** • Introducing the changing statistical landscape, covering functions, the forward problem, the inverse problem, statistical learning, and assessing model accuracy. • Organize a lab session to introduce students to RStudio and guide them in fitting a function to some data. **Key outcomes** • Students will be introduced to the concept of functions and their role in statistical learning. • Students will begin to work RStudio and gain experience fitting functions to data Session Material Mandatory readings: Lecture: ISLR, Ch. 1, 2.1, 2.2 Lab: ISLR 2.3; r4DS, Ch. 1-3, 6, 8, 11.1, 11.2 Presentation paper: None
Week 2 (2 Hours)	Session Title Linear regression, part I Session Details **Objectives** Lecture: Linear regression, part I: simple linear regression, multiple linear regression

	Lab: Linear regression, part I **Planning** • Introduction to linear regression, focusing on simple linear regression and multiple linear regression. • Conduct a lab session to implement linear regression models and practice linear regression techniques. **Key outcomes** • Students will understand the principles of simple linear regression and multiple linear regression. • Students will be capable of applying linear regression models to analyze data. Session Material Mandatory reading: • Lecture: ISLR 3.1, 3.2, • Lab: ISLR 3.6.1 – 3.6.3; r4DS 4, 22, 23 Presentaiton paper: Tukey, J. W. (1980). We Need Both Exploratory and Confirmatory. The American Statistician, 34(1), 23–25.
Week 3 (2 Hours)	Session Title The Normal distribution: a mathematical model of “noise” Session Details **Objectives** In this week's session, we will explore the concept of the normal distribution as a mathematical model of "noise." The normal distribution plays a crucial role in statistics, representing a symmetric bell-shaped curve that describes the variation in many natural phenomena. We will start by delving into the historical background of its development when dealing with quantitative measurements in astronomical observations. The objectives are to understand how the normal distribution emerges from random processes, such as coin flips and dice rolls, and to explore the Central Limit Theorem, which reveals how the distribution of sample means converges to normal as the sample size increases. **Planning** • Understand the historical context and development of the normal distribution as a model of variability in astronomical measurements. • Simulate coin flips using R code to demonstrate the binomial distribution and how it approximates a bell-shaped curve with increasing trials. • Explore the convergence of the binomial distribution to a normal distribution as the number of trials increases in R code.

	• Introduce the Central Limit Theorem and its significance in statistical inference. • Illustrate the Central Limit Theorem using R code to calculate sample means of dice rolls and observe their distribution as the number of dice increases. **Key outcomes** • Comprehend the concept of the normal distribution and its relevance in modeling random variation. • Gain insight into the Central Limit Theorem and its role in statistical inference. • Develop the ability to use R code to simulate random processes and observe the emergence of normal distributions from such processes. Session Material Mandatory reading: • Lecture: • Stahl, S. (2006). The Evolution of the Normal Distribution. Mathematics Magazine, 79(2), 96–113. PDF. • Devore, J. L., & Berk, K. N. (2012). Statistics and Sampling Distributions. In J. L. Devore & K. N. Berk (Eds.), Modern Mathematical Statistics with Applications (pp. 284–330). New York, NY: Springer. PDF. • Lab: • R coder: Normal Distribution in R: https://r-coder.com/normal-distribution-r/ • Central limit theorem, Wikipedia, https://en.wikipedia.org/wiki/Central_limit_theorem • Presentation paper: Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual. The Sage Book of Quantitative Methodology for the Social Sciences, 391–408.
Week 4	Lab session Lab session dedicated to students working on their research projects with the instructor providing guidance and support as needed.
Week 5 (2 Hours)	Session Title Linear regression, part II Session Details **Objectives** Lecture: Linear regression, part II: Qualitative predictors, extensions, and potential problems. Lab: Linear regression, part II: Interations, data transformation **Planning**

	• Continuing linear regression, covering qualitative predictors, extensions, and potential problems. • Lab session to work on interactions and data transformation in linear regression models using R. **Key outcomes** • Students will comprehend how to handle qualitative predictors and potential issues in linear regression. • Students will be skilled in performing interactions and data transformation in linear regression models. Session Material Mandatory readings: • Lecture: Reading: ISLR 3.3 – 3.5 • Lab: Reading: ISLR 3.6.4 – 3.6.7; r4DS 5 • Presentation paper: Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155–159.
Week 6 (2 Hours)	Session Title Classification, part I: logistic regression Session Details **Objectives** Lecture: Classification, part I: logistic regression, making predictions. Lab: Classification, part I: logistic regression; Exploratory data analysis (EDA), GLMs, **Planning** • Introducing classification, focusing on logistic regression and making predictions. • Lab session on classification, part I, including logistic regression, exploratory data analysis (EDA), and Generalized Linear Models (GLMs). **Key outcomes** Session Material Mandatory readings: Lecture: Reading: ISLR 4.1 – 4.3 Lab: Reading: ISLR 4.6.1, 4.6.2; r4DS 7 Presentation reading: Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. European Journal of Epidemiology, 31(4), 337–350.
READING WEEK NO COURSES	
Week 7	MID-TERM EXAM
Week 8 (2 Hours)	Session Title Classification, part II: LDA

	Session Details **Objectives** Lecture: Classification, part II: LDA, generative models, classification methods. Lab: Classification, part II: LDA ; Data wrangling **Planning** • Continuing classification, introducing Linear Discriminant Analysis (LDA), generative models, and classification methods. • Lab session on classification, part II, implementing LDA and focusing on data wrangling techniques using R. **Key outcomes** • Students will understand Linear Discriminant Analysis (LDA) as a classification method and generative models. • Students will apply LDA and gain expertise in data wrangling using R. Session Material Mandatory reading: • Lecture: ISLR 4.4, 4.5 • Lab: ISLR 4.6.3 – 4.6.6; r4DS 9, 12 • Presentation reading: Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspectives on Psychological Science, 9(6), 641–651.
Week 9 (2 Hours)	Session Title Resampling: cross-validation and the bootstrap Session Details **Objectives** Lecture: Resampling: cross-validation and the bootstrap Lab: Resampling **Planning** • Introducing resampling techniques, specifically cross-validation and the bootstrap. • Lab session on resampling, allowing students to apply cross-validation and bootstrap methods. **Key outcomes** • Students will be proficient in using cross-validation and bootstrap techniques for data analysis. • Students will understand the importance of resampling methods for model evaluation. Session Material Mandatory readings: • Lecture: ISLR 5.1, 5.2

	• Lab: ISLR 5.3, r4DS 13 • Presentation reading: Meehl, P. E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. Philosophy of Science, 34(2), 103–115.
Week 10 (2 Hours)	Session Title Linear model selection and regularization Session Details **Objectives** Lecture: Linear model selection and regularization, subset selection, shrinkage methods, high dimensional data. Lab: Model selection and regularization, String manipulation, ridge and lasso penalties. **Planning** • Lecture on linear model selection and regularization, covering subset selection, shrinkage methods, and handling high-dimensional data. • Lab session on model selection and regularization, focusing on string manipulation and ridge and lasso penalties using R. **Key outcomes** • Students will grasp the principles of linear model selection and regularization, including subset selection and shrinkage methods. • Students will apply regularization techniques and perform string manipulation using R. Session Material Mandatory readings: • Lecture: ISLR 6.1, 6.2, 6.4 • Lab: ISLR 6.5, r4DS 14 • Presentation reading: Cohen, J. (1994). The earth is round ($p < .05$). American Psychologist, 49(12), 997–1003.
Week 11 (3 Hours)	Session Title Dimensionality reduction Session Details **Objectives** Lecture: Principal components analysis (PCA), Clustering Lab: Principal components analysis (PCA), Functions in R, Clustering, Vectors **Planning**

	- Introducing dimensionality reduction techniques, such as Principal Components Analysis (PCA) and Clustering. - Lab session on dimensionality reduction, including implementing PCA, working with functions in R, and applying clustering methods using vectors. **Key outcomes** - Students will understand the concept of dimensionality reduction and its practical applications. - Students will apply PCA and clustering techniques to reduce data dimensions and analyze patterns. Session Material Mandatory reading: - Lecture: ISLR 12.1, 12.2, 12.4 - Lab: ISLR 12.5, 12.6, r4DS 19, 20 - Presentation readings: - Earp, B. D. (2018). The need for reporting negative results—A 90 year update. Journal of Clinical and Translational Research, 3(Suppl 2), 344. - Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. Scientometrics, 90(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7
Week 12	Lab Session
Week 13 (3 hours)	Session Title Reproducible research Session Details **Objectives** Lecture: Reproducible research Lab: Reproducible research, Iteration **Planning** - Introduction to reproducible research, emphasizing transparency and reliability in data analysis. - Lab session on reproducible research and iteration, guiding students to implement best practices using R. **Key outcomes** - Students will understand and apply reproducible research practices for transparent and reliable data analysis. - Students will implement reproducible research methods, including iteration and version control, in their projects. Session Material Mandatory reading: - r4DS 21, 26 – 30 - Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. A. (2017).

	A manifesto for reproducible science. Nature Human Behaviour, 1(1), 1–9. https://doi.org/10.1038/s41562-016-0021 • Presentation reading: • Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. Perspectives on Psychological Science, 7(6), 632–638. • Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. Perspectives on Psychological Science, 7(6), 615–631. https://doi.org/10.1177/1745691612459058

10. Complementary Activities

Activities	Detailed Content
Assessment 1 & 2	Exams (mid-term is 20% of final grade; final is 30% of final grade) The mid-term and final exams will take the form of a mulitple choice and open response questions about the content presented in class, in readings, and in lab with the students. Open-ended questions will be given to the students along the semester, from which 2 or 3 will be selected to constitute each exams (week 1-5 for mid-term; 7-12 for final exam). Each student will provide their own responses, which will be evaluated in terms of (1) quality of response, (2) reference to relevant notions and content, (3) demonstration of understanding. These evaluation criteria will be discussed in more details in class. For both exams, paper documentation is allowed. No electronic equipment (computer, phone, calculator, watch, etc.) are allowed.
Assesssment 3	Active engagement (20% of final grade) The students will be required to submit to the professor two questions per week from the assigned readings, each week: one must be from the weekly *Presentation Paper*. Questions should be constructive and appropriate for the level of the course (i.e., they should not always be purely definitional, but rather aimed at larger concepts). (10% of final grade) Students will also be assigned at least one week where they will present on the weekly *Presentation Paper* either on their own or in a small group of their peers (contingent on class size). Presentaitons should summarize the core concepts and applications and aim to make links to other concepts presented in class or other research applications. Presentations should also pose disccusstion questions and lead a brief discussion from other students. Each presentaiton will be about 20-30 minutes. (10% of final grade)
Assessment 4	Data analysis project (15% of final grade) Each student will present on a data analysis project of their choosing and submit a brief manuscript describing their work. The details of this project will be further discussed in class. This exercise has the objective of motivating students to apply knowledge on data analysis learned from course content to in a novel data analysis context. The presentation will summarize the relavant background, data sets, research questions addressed, statistical analyses deployed, and insights from analyses. It is worth 20% of the final grade. Precise instructions on the format of the presentation and its evaluation will be given in class during the first or second week.