"Quantitative Methods for the Social Sciences I + II" a.y. 2024_25 Course Syllabus Instructors: Marco Albertini (marco.albertini2@unibo.it) Francesca Zanasi (francesca.zanasi5@unibo.it)

Course objectives

The course is an introduction to empirical research methods in political and social sciences. In particular, the course will focus on quantitative data analysis and adopt a "how to do" approach, combining theoretical lectures with practical exercises. The course represents an opportunity for students to develop skills in data analysis, and interpretation and presentation of research findings, including standards for written research papers.

The course will be split in two main sections. The first part of the course will focus on the basics of quantitative methods, revising topics such as data collection, measurement, research design, statistical inference, and univariate, bivariate and multivariate data analysis. The second part will concentrate on advanced techniques of data analysis (including panel data analysis, sequence analysis, and factor analysis) and on how to present and report the results of quantitative analyses.

Regular classes will be organized in theoretical lectures followed by laboratory sessions. Approximately one third of the time will be devoted to lectures, and two third of the time will consist of laboratory sessions where students can implement data analysis using R, one of the most widely used statistical software in social sciences.

Through lectures, readings, and lab-sessions, students are expected to learn how to address empirical issues in political and social sciences, as well as how to tackle the main methodological problems researchers might encounter in quantitative empirical research. Furthermore, students are expected to develop intermediate skills in the use of R statistical software.

Week 1: Thursday November 7^h, 14-16

Introduction

Instructor: Marco Albertini

This first class will be devoted to the presentation of the main contents and methods of the course and to a presentation (and discussion) about the methodological aspects of the student's research projects/proposals.

Readings: Agresti, Ch.1; Corbetta, Ch. 3; Auspurg & Brüderl (2021)

Week 2: Thursday November 14th, 14-17

The basics of empirical research in the social and political sciences

Instructor: Francesca Zanasi

In the first part of this lecture, we will explore the primary purposes and features of surveys, providing an overview of some of the most widely used surveys in political and social sciences (e.g., the European Social Survey (ESS). We will refresh our understanding of the process of transforming abstract concepts into measurable variables, a key aspect of operationalization, with a focus on the types of variables (nominal, ordinal, discrete, continuous) and their characteristics. Additionally, we will review fundamental concepts in sampling theory, including a brief presentation of the main sampling methods. The second part of the class will be devoted to the R lab. We will examine the different data formats, and how to import datasets in R, and an overview of basic commands for data management.

<u>Readings:</u> Corbetta: Chapters "From Theory to Empirical Research", "Sampling", "The Survey" and "Scaling"; Hanneman, Kposowa, and Riddle: Chapter 1; Agresti 1: Chapter 1 (1.2)

Week 3: Thursday November 21st, 14-17 Univariate and bivariate Description

The class will focus on univariate and bivariate description. It will begin with a review of measures of central tendency, dispersion, and position for continuous variables, as well as absolute and relative frequencies for categorical variables. Following this, the class will introduce the study of relationships between two variables, characterized by their shape, strength, and direction, discussing statistical tools such as contingency tables, relative risks, and odds ratios. The R lab will concentrate on computing univariate and bivariate descriptive statistics and producing graphical visualizations of the results.

<u>Readings:</u> Agresti 1: Chapter 2; Chapter 3 (3.1); Chapter 11 (11.1-11.3); Hanneman, Kposowa, and Riddle: Chapter 2, 3, 4, 5, 6, 10, 12

Week 4: Thursday November 28th, 14-17

Statistical inference & significance testing

Instructor: Francesca Zanasi

The class will focus on statistical inference, introducing key concepts such as estimators, estimates, and population and sampling distributions. Students will learn how to calculate standard errors and confidence intervals to account for uncertainty in their analyses. The course will also cover the notion of statistical significance, explaining the main elements of significance tests, including null and alternative hypotheses, test statistics, and p-values. Each topic will be complemented by practical exercises in R to reinforce the concepts learned.

<u>Readings:</u> Agresti 1: Chapter 8, 9; Hanneman, Kposowa, and Riddle: Chapter 7, 8, 9; Bernardi et al. (2017)

Week 5: Thursday December 5th, 14-17

Regression: OLS models

Instructor: Marco Albertini

We will address the concept of directionality in the association between variables by exploring the main elements of bivariate linear regression models (OLS). By means of exercises on R, we will focus on the interpretation of beta coefficients, standard errors, confidence intervals and p-values in linear regression. We will also present measures of models' goodness of fit and discuss OLS assumptions (and their violation).

<u>Readings:</u> Agresti 1: Chapter 3 (3.2, 3.3), 12; Agresti 2: Chapter 9; Wolf & Best: Chapter 2, 4, 5; Hanneman, Kposowa, and Riddle: Chapter 13, 14

Week 6: Thursday December 12th, 14-17

Applied multivariate regressions (interpreting the outputs etc.)

Instructor: Marco Albertini

We will, first, address multivariate regression from a theoretical standpoint, by discussing the role of confounding, mediating, and moderating variables. The second part of the class will be devoted to a practical session in R in which we will perform multivariate regression models. Specifically, we will interpret beta coefficients and introduce additional estimates, including Average Marginal Effects, Average Partial Effects, and predicted values. Finally, we will address the interaction between variables both from a theoretical and practical standpoint.

<u>Readings:</u> Agresti 1: Chapter 13 (13.1, 13.5); Agresti 2: Chapter 10, 11, 13; Wolf & Best: Chapter 4-5; Hanneman, Kposowa, and Riddle: Chapter 14 (Multiple Regression)

Week 7: Thursday December 19th, 14-17

Regression models for categorical dependent variables. Logit/mlogit

Instructor: Marco Albertini

We will present the main techniques used to model binary outcomes by introducing Logistic regression and Linear Probability Models. We will see on R how to perform logistic regression models and how to interpret log-coefficients and odds ratios. The class will end with a critical examination of pros and cons of logistic vs. LPM models, and with a discussion about students' group work.

<u>Readings:</u> Agresti 1: Chapter 13 (13.6); Agresti 2: Chapter 15; Wolf & Best, Ch. 8; Mood (2010); Hanneman, Kposowa, and Riddle: Chapter 15

Week 8: Thursday January 9th, 14-16

Assignment presentation

Instructor: Francesca Zanasi

The session will be devoted to the presentation of students' group works, and to the clarification of issues raised by students about the first part of the course.

Week 9: Thursday January 16th, 14-17

Visualizing and presenting results

Instructor: Francesca Zanasi

We will explore some of the main techniques for visualising data in R software. We will use the ggplot package and explore its main features (types of graphs, aesthetics, settings). We will also focus on the visualisation of spatial data by plotting maps.

Readings: Wickham (2016)

Week 10: Thursday January 23rd, 14-17 Summarizing the information: Factor analysis and PCA

Instructor: Francesca Zanasi

The session will focus on dimensionality reduction techniques such as Factor Analyses (FA) and Principal Component Analysis (PCA). We will present the main characteristics of FA and PCA, while also discussing about their similarities and differences. In addition, we will revise the main differences between Exploratory and Confirmatory Factor Analysis (EFA or CFA). In the practical session on R, we will address the main elements of Factor Analysis, including factor loadings, commonality, eigen-values, and factor scores.

Readings: Fabrigar & Wegener (2012): Chapter 1, 2, 3, 5; Schulz et al. (2018)

Week 11: Thursday January 30th, 14-17

Sequence analysis

Instructor: Francesca Zanasi

We will, first, present what is sequence analysis and how it can be utilized in social and political sciences. The second part of the class will be devoted to a practical exercise with R introducing basic data management commands to prepare the data set for sequence analysis and basic commands to describe sequences and their main characteristics.

Readings: Aisenbrey & Fasang (2010); Brzinsky-Fay, Kohler, & Luniak, (2006)

Week 12: Thursday February 6th, 14-17

Panel data I: Theory, data management and fixed-effects

Instructor: Marco Albertini

We will, first, present the characteristics of panel surveys and panel data with a brief discussion about their strengths and weaknesses. Next, we will explore the basics of fixed-effect regression models and the differences with pooled OLS regression models both on a methodological and practical standpoint. The R lab will concentrate on panel data management and on how to perform fixed-effect regression models.

Readings: Wolf & Best (2014): Chapter 15; Alison (2009): Chapter 1, 2, 3; Bell et al. (2019)

Week 12: Thursday February 13th, 14-17

Panel data II: Random effects and hybrid models

Instructor: Marco Albertini

Following on the previous class, we will address random effects models in panel data analysis. Furthermore, we will see how to separately account for the variation within and between individuals by introducing and discussing hybrid models.

Readings: Wolf & Best (2014): Chapter 15; Alison (2009): Chapter 1, 2, 3; Bell et al. (2019)

Week 14: Wednesday February 20th, 14-17 Final assignment presentation

Instructor: Marco Albertini

The last session will be devoted to the presentation of students' final work.

References:

- 1. Agresti, A., Franklin, C. A, Klingenberg, H. (2018). *Statistics: the art and science of learning from data*. England: Pearson. Available as ebook on SBA Almastart (In Syllabus: called **Agresti 1**)
- 2. Agresti, Alan (2018). *Statistical Methods for the Social Sciences*, Fifth edition. Pearson. Available as ebook on SBA Almastart (In Syllabus: called Agresti 2)
- 3. Aisenbrey, S. & Fasang, A. E. (2010) New lifes for old ideas: The "second wave" of sequence analysis. Bringing the "course" back into the life course. *Sociological Methods and Research*, 38(3):420-462.
- 4. Alison, D. P. (2009) Fixed effects regression models. London: Sage.
- 5. Auspurg K. & Brüderl J. (2021) Has the credibility of the social sciences been credibly destroyed? Reanalyzing the "Many analysts, one data set" project. Socius, 7: 1-14.
- 6. Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality & quantity*, *53*(2), 1051-1074.
- 7. Bernardi, F., Chakhaia, L., & Leopold, L. (2017). 'Sing me a song with social significance': the (mis) use of statistical significance testing in European sociological research. *European Sociological Review*, 33(1), 1-15.
- 8. Brzinsky-Fay, C., Kohler, U. & Luniak, M. (2006) Sequence analysis with Stata. *TheStata Journal*, 6(4): 435-460.
- 9. Corbetta, P. (2003). *Social research: Theory, methods and techniques*. SAGE Publications, Ltd. Available as ebook on SBA Almastart
- 10. Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.
- 11. Hanneman, R. A, Kposowa, A. J, Riddle, M. D. (2013) *Basic Statistics for Social Research*. San Francisco, CA: Jossey-Bass. Available as ebook on SBA Almastart
- 12. March, C. L. & Cormier R. D. (2002) Spline regression models. Sage.
- 13. Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European sociological review*, 26(1), 67-82.
- 14. Schulz, A., Müller, P., Schemer, C., Wirz, D. S., Wettstein, M., & Wirth, W. (2018). Measuring populist attitudes on three dimensions. *International Journal of Public Opinion Research*, 30(2), 316-326
- 15. Wolf, C. & Best, H. (Eds.) (2014) The SAGE handbook of regression analysis and causal inference. London: Sage.