Intro to statistics

A brief introduction to applied statistics for quantitative social science research

Laboratorio de Investigación para el Desarrollo del Ecuador

Instructor: Daniel Sánchez, MA

Module length: 15 hours

Course level: Intermediate

Prerequisite knowledge: Basic arithmetic, algebra, cartesian graphing, Stata scripting.

GitHub repository: https://github.com/laboratoriolide/intro-to-stats

Course overview

This course introduces statistics at a basic level, focusing on its application to quantitative social science research. The course assumes no prior knowledge of statistics, but students will reach an intermediate level of undergraduate statistics by the end of the course. The material will not be proof-heavy, however, it will be rigorous enough to prepare you for the more advanced modules in the program (econometrics, causal inference, etc.)

A good grasp of statistics cannot be achieved with theory alone. I will provide practical implementations of what we cover during class mainly using R and, to a significantly lesser extent, Stata. The sister module to this course, *Intro to R*, will focus on the practical implementations and technical aspects, while this course will focus on theory and intuition.

To best understand the material, students should have a basic understanding of arithmetic, algebra and cartesian graphs (i.e. x-y plots). I will not introduce Stata in this course as it is assumed all students have already taken the introductory course on Stata from our last module, *Intro to Stata*, but I will provide guidance if needed.

Module contents

The following is a planned outline of the course. This may change depending on the pace of the class. Each lecture will have one or two assigned readings, which will all be academic articles submitted to the course's GitHub repository. See below for details on these readings.

- **Lecture 1**: Intro
    - Introduction to statistics: what is it and why do we need it in social science research?
    - Observations, variables, data types, data formats
    - Populations vs. samples
    - Descriptive vs. inferential statistics
    - **Assigned reading**: Course syllabus

- **Lecture 2**: Descriptive statistics and statistical data visualization
    - Measures of central tendency: mean, median, mode
    - Measures of dispersion: variance, standard deviation, range, IQR
    - Measures of distributional shape: skewness, kurtosis
    - Data visualization: histograms, box plots, scatter plots, cumulative distributions
    - **Assigned reading**: Lupu and Zechmeister (2020), prepare to replicate charts and apply descriptive statistics to the dataset (file provided on GitHub)

- **Lecture 3**: A brief introduction to probability
    - Why probability?
    - Basic probability: experiments, counting, set theory, conditional probability
    - Random variables and probability distributions (discrete and continuous)
    - Probability distribution functions (PDFs) and cumulative distribution functions (CDFs)
    - The normal distribution and its properties, the empirical rule
    - The standard normal distribution and z-scores
    - **Assigned reading**: Redacción Primicias (2020), will be used for a class exercise.
    - Time-permitting: bootstrapping

- **Lecture 4**: Statistical inference
    - What is statistical inference?
    - Sampling distributions
    - The central limit theorem
    - The law of large numbers
    - Expected value, variance and standard errors
    - Using simulation to understand sampling distributions

- **Lecture 5**: Point and interval estimation
    - What is estimation?

- Bias, efficiency and margin of error
- Significance levels (alpha)
- Margin of error with sampling distributions
- Confidence intervals and their interpretation
- **Assigned reading**: Barbayannis et al. (2022), with attention to the research question, charts and discussion, less on methods being used.

- **Lecture 6**: Hypothesis testing

  - What is hypothesis testing?
  - Type I and Type II errors
  - p-values and critical values
  - Single-sample hypothesis tests: one and two-tailed tests
  - One sample z-test for mean inference

- **Lecture 7**: More on hypothesis testing

  - The t-distribution, degrees of freedom and the one sample t-test
  - Two sample z and t-tests for mean inference
  - Paired t-tests (dependent samples)
  - Time permitting: critical value tables
  - Time permitting: proportion tests and chi-squared tests
  - Choosing the right test for your research question
  - **Assigned reading**: Broockman and Skovron (2018), with attention to Table 2.

- **Lecture 8**: Experimental design and ANOVA

  - What is experimental design?
  - Why experiments? Causality in social science
  - One-way and two-way ANOVA
  - **Assigned reading**: Stanley, Neck, and Neck (2023),
  - Time-permitting: factorial ANOVA

- **Lecture 9**: Ordinary least squares regression (OLS) I

  - What is regression?
  - Simple linear regression
  - Coefficient interpretation in simple regression
  - OLS features and properties: residuals, fitted values, unit transformations
  - Linearity in parameters
  - Time permitting: multiple regression
  - **Assigned reading**: Holcombe and Boudreaux (2015), with attention to regression (1) on Table 1.

- **Lecture 10**: Case study presentations

Readings and course materials

Evaluation

Assignments

There will be weekly DataCamp assignments, which will be either courses, projects or exercises. These assignments are due by 11:59pm on the Sunday of the week they are assigned, however, please check the DataCamp platform for the exact due date. The assignments will focus on R programming, statistical theory or other relevant topics. There will not be assignments in Stata, however, I will provide optional Stata exercises for those interested.

Attendance and participation

I do not have any special requirements for attendance nor participations other than the requirements set by the program. Consult the program's regulation handbook for more information. I encourage you to participate in class and ask questions, as this will help you understand the material better. Statistics typically inspires frustation, so it is important to ask questions when you are confused - as any other quantitative course, the material builds up on itself, so better to understand things sooner rather than later.

Module policies

Communication

All communications to the instructor or teaching assistant (TA) should be made through the course's Slack channel. We hope to respond to questions within 72 hours, but please be patient if we take longer.

References

Barbayannis, Georgia, Mahindra Bandari, Xiang Zheng, Humberto Baquerizo, Keith W. Pecor, and Xue Ming. 2022. "Academic Stress and Mental Well-Being in College Students: Correlations, Affected Groups, and COVID-19." *Frontiers in Psychology* 13 (May). https://doi.org/10.3389/fpsyg.2022.886344.

Broockman, David E., and Christopher Skovron. 2018. "Bias in Perceptions of Public Opinion Among Political Elites." *American Political Science Review* 112 (3): 542–63. https://doi.org/10.1017/S0003055418000011.

Holcombe, Randall G, and Christopher J Boudreaux. 2015. "Regulation and Corruption." *Public Choice* 164: 75–85.

Lupu, Noam, and Elizabeth J. Zechmeister. 2020. "Chapter 3. Social Media and Political Attitudes in the Latin American and Caribbean Region." In *The Political Culture of Democracy in Ecuador and in the Americas, 2018/19: Taking the Pulse of Democracy*, edited by Juan Carlos Donoso, Paolo Moncagatta, Arturo Moscoso, Simón Pachano, J. Daniel Montalvo, and Elizabeth J. Zechmeister, 1st ed., 1:45–72. AmericasBarometer Country Studies 2018/19. Quito: Latin American Public Opinion Project. https://www.vanderbilt.edu/lapop/ecuador/AB2018-19-Ecuador-Country-Report-Eng-V2-W-200903.pdf.

Redacción Primicias. 2020. "Mapa electoral: cuánto pesa el voto de las 24 provincias." Primicias. October 12, 2020. https://www.primicias.ec/noticias/politica/mapa-electoral-pesa-voto-provincias/.

Stanley, Matthew L, Christopher B Neck, and Christopher P Neck. 2023. "Loyal Workers Are Selectively and Ironically Targeted for Exploitation." *Journal of Experimental Social Psychology* 106: 104442.