# 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**: 12 hours

**Course level**: Intermediate

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

## 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 folder. See below for details on these readings.

- Lecture 1: Introduction to statistics, the logic of data and descriptive analyses
  - Introduction to statistics
  - Observations, variables, data types, data formats
  - Descriptive statistics and graphical representations
  - **Assigned reading**: Course syllabus
- Lecture 2: A brief introduction to probability
  - Why probability?
  - Basic probability: experiments, counting, set theory, conditional probability
  - Random variables and probability distributions (discrete and continous)
  - * **Assigned reading**: Barbayannis et al. (2022)
- Lecture 3: Statistical inference
  - What is statistical inference?
  - Sampling distributions, expected value and standard errors
  - Central limit theorem
  - Using simulation to understand sampling distributions
  - Confidence intervals and margin of error
- Lecture 4: Hypothesis testing
  - What is hypothesis testing?
  - Type I and Type II errors
  - Single-sample hypothesis tests: one and two-tailed tests
  - Z-tests and t-tests for mean inference (using software)
  - Time permitting: proportion tests
- Lecture 5: 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)
- Lecture 6: Ordinary least squares regression (OLS) I
  - What is regression?
  - Simple linear regression
  - Coefficient interpretation in simple regression

- Lecture 7: Ordinary least squares regression (OLS) II
  - Multiple linear regression
  - Coefficient interpretation in multiple regression
  - OLS assumptions for unbiasedness
  - Demonstrating unbiasedness with simulation
  - Functional form with OLS
- Lecture 8: Ordinary least squares regression (OLS) III
  - Hypothesis testing in regression
  - Dummy variables
  - Goodness of fit
  - Reading regression tables

## Administrative matters

### 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.

### 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.

### Materials

## 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.

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.