

Statistics I
Master in Social Sciences
2021-2022

Instructors: Esther Ruiz and Guillermo Kreiman

This course will consist of a brief introduction to probability theory as well as various topics in statistics and how they can be used in Social Sciences. Subject areas will include random variables, distributions, estimation, hypothesis testing and introduction to the regression model. The course is elementary in mathematical level. The main objective is to show when, how and why statistical techniques can be used and which their limitations are so that students can understand and apply statistical techniques to answer Social Science research questions. The data sources will be taken from publicly available social data bases as, for example, the *Centro de Investigaciones Sociológicas* (CIS) or *Quality of Government* (<http://qog.pol.gu.se/data/datadownloads/qogbasicdata>). Stata will be used to perform the statistical analysis in the labs.

SYLLABUS

TOPIC 1. UNDERSTANDING RANDOM VARIABLES (3 weeks)

- 1.1 Introduction: Importance of Statistics
- 1.2 Univariate random variable. Definition. Discrete and continuous variables. Cumulative distribution, probability density and probability mass functions.
- 1.3 Summarizing information of univariate variables: mean, variance, asymmetry and kurtosis.
- 1.4 Multivariate random variable. Marginal and conditional distributions.
- 1.5 Summarizing information of multivariate variables: Covariances and correlations.
- 1.6 Some common univariate distributions: Bernoulli, Binomial, Poisson, Uniform, Normal, Student
- 1.7 The multivariate normal distribution

TOPIC 2. INFERENCE AND ESTIMATION METHODS (4 weeks)

- 2.1 Population and sample: parameters and statistics
- 2.2 Method of Moments estimator
- 2.3 Maximum Likelihood estimator
- 2.4 Properties of estimators: Small and large sample
- 2.5 Interval estimation
- 2.6 Hypothesis testing

TOPIC 3. REGRESSION MODEL (4 weeks)

- 3.1 Simple Regression model: Conditional means
- 3.2 Estimating the parameters: Least Squares estimator
- 3.3 Properties of LS estimator: Consistency, normality and efficiency
- 3.4 Residual diagnostic
- 3.5 Hypothesis testing
- 3.6 Heteroscedasticity
- 3.7 Using the regression model to predict

TEXTBOOKS

Moore, D.S., G.P. McCabe and B.A. Graig (2009), *Introduction to the Practice of Statistics*, 6th edition, New York: W.H. Freeman and Company.

Newbold, P., W.L. Carlson and B. Thorne (2013), *Statistics for Business and Economics*, 8th edition, Pearson Education.

Stevens, J.S. (2009), *Applied Multivariate Statistics for the Social Sciences*, Routledge

Leekley, R.M. (2010), *Applied Statistics for Business and Economics*, CRC Press, Boca Raton.

De Groot, M.H. (1986), *Probability and Statistics*, 2nd edition, Addison-Wesley.

Wackerly, D.D., W. Mendenhall and R.L. Scheaffer (2008), *Mathematical Statistics with Applications*, 7th edition, Cengage Learning.

EVALUATION

Final exam: 40%

Two partial exams: 20% each

Stata project: 20%

You need at least a mark of 4 in the final exam for the partials and Stata project to be considered.