Multivariate Statistics
Course Outline

Pedro Galeano
Departamento de Estadística
Universidad Carlos III de Madrid
pedro.galeano@uc3m.es

Course 2017/2018

Master in Mathematical Engineering

Instructor

- Instructor: Pedro Galeano.

- Email: pedro.galeano@uc3m.es

- Webpage: http://www.est.uc3m.es/pgaleano

- Office: 10.1.26 (Getafe Campus).

- Phone: 91 624 8901

- Office hours: By appointment.

Course information

- Classes: From September 18th to October 30th.

- Timetable:

    - Mondays, 10:00 to 14:00 at classroom 7.3.J08.

- The classes will contain some computer work so it is advisable to bring a laptop with you or to share one with someone else.

- The software taught in the classes is **R**.

Course outline

- Chapter 1: Multivariate data.

- Chapter 2: Multivariate distributions and inference.

- Chapter 3: Principal components analysis.

- Chapter 4: Factor analysis.

- Chapter 5: Multidimensional scaling.

- Chapter 6: Cluster analysis.

- Chapter 7: Classification.

- Chapter 8: Canonical correlation analysis.

Course evaluation

- 50% final exam:

  - The exam will be held next November, 6th at 10:00 at classroom 7.3.J08.

  - The exam will include theoretical questions such as:

    - Explain the k-means algorithm.

    - Explain the non-identifiability of the factor model with respect to rotations.

    - Justify the Fisher linear discriminant rule and explain when and why this rule is optimal.

Course evaluation

- 50% project:

    - ▶ The project should be done individually.

    - ▶ The projects will be focused on analyzing a certain data set with the techniques seen during the course.

    - ▶ The idea is to use the codes in **R** that will be given during the course.

    - ▶ The advances of the projects have to be presented in the classes every week, starting on the third week.

    - ▶ A written project with the complete results should be given at the end of the course.

    - ▶ The first thing to do during the first week is to look for a suitable data set (more information about it will be given during the classes).

Course evaluation

- Data repositories:

    - http://www.inside-r.org/howto/finding-data-internet

    - https://archive.ics.uci.edu/ml/datasets.html

    - http://www.stat.ufl.edu/∼winner/datasets.html

    - http://www.census.gov/2010census/data/

    - https://www.ehdp.com/vitalnet/datasets.htm

    - https://data.europa.eu/euodp/en/data

    - http://www.datasciencecentral.com/profiles/blogs/big-data-sets-available-for-free

Course contents

- All the material (slides, **R** scripts, data sets) of this course will be posted in Aula Global.

- The material will be also posted in the following webpage:

    http://www.est.uc3m.es/pgaleano/esp/2017-2018/MultivariateStatistics.html

- Some basic background on linear algebra will help a lot (some explanation will be given if needed).

Textbooks

- Peña, D. (2002). Análisis de Datos Multivariantes. McGraw Hill.

- Härdle, W. K. and Simar, L. (2007). Applied Multivariate Statistical Analysis. Springer.

- Johnson, R. A. y Wichern, D.W. (2007). Applied Multivariate Statistical Analysis (6a Ed.). Prentice Hall.

- Koch, I. (2013). Analysis of Multivariate and High-Dimensional Data. Cambridge University Press.

- Manly, B. F. M. (2005). Multivariate Statistical Methods: A Primer (3a Ed.). Chapman-Hall.

- Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). Multivariate Analysis. Academic Press.

Where to start?

- The first step is to introduce multivariate data analysis.

- This is done in:

Chapter 1: Multivariate data