

Multivariate Statistics (Practice)

Bachelor's Degrees in: Computer Science and Mathematics; Mathematics; Physics and Mathematics
(4th Academic Year - 1st semester 2023-2024)

Prof. Dr. José Luis Romero Béjar



Department of Statistics and Operations Research
Faculty of Sciences (Office 10)

Teaching period: 11/09/2023 - 22/12/2023

- 1 Relevant information of the course
 - Lecturers
 - Schedule
 - Materials
- 2 Assessment methods and tutoring timetable
 - Assesment methods
 - Tutoring timetable
- 3 Detailed syllabus
 - Syllabus
 - Tentative timeline
- 4 Bibliography
 - References
 - Recommended links

Theory

- José Miguel Angulo Ibáñez.

Practice

- José Luis Romero Béjar.

- 1 Relevant information of the course
 - Lecturers
 - **Schedule**
 - Materials
- 2 Assessment methods and tutoring timetable
 - Assesment methods
 - Tutoring timetable
- 3 Detailed syllabus
 - Syllabus
 - Tentative timeline
- 4 Bibliography
 - References
 - Recommended links

Practices

- **Group A:** Thursday 16:00-17:00 (Room M01).

Remarks

- Students are suggested to **bring their own laptop** to get full advantage of each practice session.

- 1 Relevant information of the course
 - Lecturers
 - Schedule
 - **Materials**
- 2 Assessment methods and tutoring timetable
 - Assesment methods
 - Tutoring timetable
- 3 Detailed syllabus
 - Syllabus
 - Tentative timeline
- 4 Bibliography
 - References
 - Recommended links

Resources for the students

- **Bibliographical entries** at the end of this document are appropriate as a complement to the practice sessions.
- **Digital documents:**
 - Slides with the outline of the session.
 - Newly written **R language source code** for this course.
 - Statements of the practical exercises.
 - Any other material considered by the teacher.

Remarks

- The use of paper sheets will be limited to extraordinary situations, prioritizing the digital format.
- The material prepared by the teacher and other documentation will be shared with the student through the **PRADO** platform.
- All proposed activities, mandatory and/or elective, will be delivered by students through the **PRADO** platform.
- **EXTREMELY IMPORTANT.** Along with communication in face-to-face sessions, the only other **means of communication is through the PRADO platform** (e-mail option within PRADO once authenticated).
- **Emails with queries or delivery of activities that do not come from the aforementioned platform will not be answered.**

- 1 Relevant information of the course
 - Lecturers
 - Schedule
 - Materials
- 2 Assessment methods and tutoring timetable
 - Assesment methods
 - Tutoring timetable
- 3 Detailed syllabus
 - Syllabus
 - Tentative timeline
- 4 Bibliography
 - References
 - Recommended links

Assessment methods for the practice part of the course

Continuous evaluation will be prioritized. The maximum score of the practice part represents up to 25% of the final qualification of the course, in continuous evaluation.

The distribution of scores for the practice part of the course is as follows:

- **40%**: class attendance.
- **10%**: participation and voluntary resolution of proposed exercises.
- **10%**: evaluable test-type practical quiz during class time.
- **40%**: performance of an individual or group work (maximum three people) at a proposal of the teacher.

- 1 Relevant information of the course
 - Lecturers
 - Schedule
 - Materials
- 2 Assessment methods and tutoring timetable
 - Assesment methods
 - Tutoring timetable
- 3 Detailed syllabus
 - Syllabus
 - Tentative timeline
- 4 Bibliography
 - References
 - Recommended links

Tutoring timetable

- Timetable as published in the Ugr directory following this link: [José Luis Romero Béjar](#).
- Tutoring may be individual or group.
- Initially, they will be **carried out in person**, except in exceptional cases considered by the teacher or at the express request of the student, in which case the Google Meet room will be used, whose access link is available in the PRADO space for this subject.

Remarks

- **The tutoring schedule is flexible.**
- Students must request a tutoring meeting in advance via email from the **PRADO** platform.
- The email request must **indicate a proposed day and time** for the tutoring, as well as the **object to be discussed** in it.

- 1 Relevant information of the course
 - Lecturers
 - Schedule
 - Materials
- 2 Assessment methods and tutoring timetable
 - Assesment methods
 - Tutoring timetable
- 3 Detailed syllabus
 - Syllabus
 - Tentative timeline
- 4 Bibliography
 - References
 - Recommended links

Detailed syllabus for the practice part of the course

In the different practice sessions, formal and applied aspects related to **units 3, 4, 5 and 6** of the course guide are addressed.

- **Unit 3.** Principal component analysis.
- **Unit 4.** Factor analysis.
- **Unit 5.** Discriminant analysis.
- **Unit 6.** Cluster analysis.

- 1 Relevant information of the course
 - Lecturers
 - Schedule
 - Materials
- 2 Assessment methods and tutoring timetable
 - Assesment methods
 - Tutoring timetable
- 3 Detailed syllabus
 - Syllabus
 - Tentative timeline
- 4 Bibliography
 - References
 - Recommended links

Date	Content developed throughout the session
09/14/2023	<p>Introducing the course.</p> <ul style="list-style-type: none"> - R-language introduction: R-console and RStudio interface. - Talking to data. ¡A challenge for the students!
09/21/2023	<p>Unit 3. Introduction to principal component analysis (PCA).</p> <ul style="list-style-type: none"> - Formal aspects: assumptions and analytical calculation of the principal components.
09/28/2023	<p>Practice 1.1: PCA with R-language.</p> <ul style="list-style-type: none"> - In this practice: verifying desirable assumptions, obtaining the principal components, choosing the optimal number of principal components and introducing some bi-plot graphical outputs.
10/05/2023	<p>Practice 1.2: PCA with R-language.</p> <ul style="list-style-type: none"> - In this practice the previous one will be strengthened and different methods for choosing the optimal number of principal components will be introduced, as well as more informative bi-plot graphs.
10/19/2023	<p>Unit 4. Introduction to factorial analysis (FA).</p> <ul style="list-style-type: none"> - Formal aspects: latent variables, assumptions, approach and fundamental equation.
10/26/2023	<ul style="list-style-type: none"> - Methods for obtaining latent factors: method of principal factors and method of maximum likelihood. - Simplifying the representation of reality: rotation of factors. <p>Practice 2: FA with R-language.</p> <ul style="list-style-type: none"> - In this practice: verifying of previous assumptions, obtaining latent factors, choosing an appropriate number of factors and simplifying the representation by means of varimax rotation. New visual tools will be introduced for correlation analysis, etc.
11/02/2023	<p>Practice 2: FA with R-language (continuation).</p> <ul style="list-style-type: none"> - Self-assessment test-quiz in PRADO platform (10-15 minutes). <p>Final activity (assessable): proposal and statement.</p>

Date	Content developed throughout the session
11/09/2023	<p>Unit 6. Introduction to cluster analysis (CA).</p> <ul style="list-style-type: none"> - Unsupervised learning: problem statement. - Drawbacks and Applicability.
11/16/2023	<ul style="list-style-type: none"> - Hierarchical methods. Numerical example with Ward's method. - Non-hierarchical methods: k-means algorithm.
11/23/2023	<p>Practice 3.</p> <ul style="list-style-type: none"> - In this practice, an example of the application of a hierarchical algorithm will be illustrated with the interpretation of the output dendrogram. The same example will be faced through the non-hierarchical approach with the k-means algorithm. <p>New visual tools (heatmaps, etc.) that help to identify the appropriate number of clusters are also introduced.</p>
11/30/2023	<p>Unit 5. Introduction to discriminant analysis (DA).</p> <ul style="list-style-type: none"> - Supervised learning: problem statement. - Linear (LDA) and quadratic (QDA) discriminant analysis formal aspects: Bayes and Fisher approaches.
12/14/2023	<p>Practice 4: DA with R-language.</p> <ul style="list-style-type: none"> - In this practice, a classification problem will be faced by fitting both linear and quadratic discriminant models. To do this, the assumptions will be verified, the models will be fitted, a primary cross-validation will be carried out, validity measures will be obtained (sensitivity, specificity, PPV, NPV, ROC curve) and inference will be made on the models obtained.
12/21/2023	<p>Workshop: other multivariate techniques applied to supervised learning such as logistic regression, decision trees, support vector machine (SVM), etc.</p> <ul style="list-style-type: none"> - Final activity (assessable): final recommendations for delivery.

Final remarks

- The practice part of the course represents 1.5 ECTS credits that correspond to 13 one-hour sessions.
- Taking into account the holidays, there are 13 class sessions scheduled.
- The practice sessions **end on December 21, 2023** because December 7, 2023 has been considered **non-working day**. If it were not, they would end on December 14, 2023.

- 1 Relevant information of the course
 - Lecturers
 - Schedule
 - Materials
- 2 Assessment methods and tutoring timetable
 - Assesment methods
 - Tutoring timetable
- 3 Detailed syllabus
 - Syllabus
 - Tentative timeline
- 4 Bibliography
 - **References**
 - Recommended links

- [1] Anderson, T.W. (2003, 3^a ed.). An Introduction to Multivariate Statistical Analysis. John Wiley & Sons.
- [2] Gutiérrez, R. y González, A. (1991). Estadística Multivariable. Introducción al Análisis Multivariante. Servicio de Reprografía de la Facultad de Ciencias. Universidad de Granada.
- [3] Härdle, W.K. y Simar, L. (2015, 4^a ed.). Applied Multivariate Statistical Analysis. Springer.
- [4] Johnson, R.A. y Wichern, D.W. (1988). Applied Multivariate Analysis. Prentice Hall International, Inc.
- [5] Rencher, A.C. y Christensen, W.F. (2012, 3^a ed.). Methods of Multivariate Analysis. John Wiley & Sons.
- [6] Timm, N.H. (2002). Applied Multivariate Analysis. Springer.
- [7] Baillo, A. y Grané, A. (2008). 100 Problemas Resueltos de Estadística Multivariante. Delta Publicaciones.
- [8] Fujikoshi, Y., Ulyanov, V.V. y Shimizu, R. (2010). Multivariate Statistics. High-Dimensional and Large-Sample Approximations. John Wiley & Sons.
- [9] Hair, J., Anderson, R., Tatham, R. y Black, W. (2000, 5^a ed.). Análisis Multivariante. Prentice Hall.
- [10] Härdle, W.K. y Hlávka, Z. (2015, 2^a ed.). Multivariate Statistics. Exercises and Solutions. Springer.
- [11] Tusell, F. (2016). Análisis Multivariante.
- [12] Zelterman, D. (2015). Applied Multivariate Statistics with R. Springer.

- 1 Relevant information of the course
 - Lecturers
 - Schedule
 - Materials
- 2 Assessment methods and tutoring timetable
 - Assesment methods
 - Tutoring timetable
- 3 Detailed syllabus
 - Syllabus
 - Tentative timeline
- 4 Bibliography
 - References
 - Recommended links

Recommended links

- 1 [R project for Statistical Computing.](#)
- 2 [RStudio.](#)
- 3 [Book \[11\] in pdf format.](#)
- 4 [Ugr PRADO platform.](#)
- 5 [Digital bookstore of Springer publisher.](#)
- 6 [Digital bookstore of ScienceDirect.](#)