

Estadística Aplicada III

Semestre 2018-II

Profesor: Dr. Jorge Francisco de la Vega Góngora
email: jorge.delavegagongora@gmail.com

Introducción

El objetivo fundamental de este curso es introducir a los estudiantes al análisis multivariado de datos. El curso se presenta en tres vertientes principales: el análisis exploratorio, el análisis multivariado de datos cuantitativos y el análisis de datos categóricos. En cada caso se revisan en segundo plano los aspectos teóricos que sustentan cada técnica y se hace un énfasis muy especial en los aspectos prácticos haciendo uso de bases de datos reales.

Objetivo General de la materia

- Introducir al alumno al manejo de métodos multivariados estadísticos
- Al final del curso el alumno deberá ser capaz de llevar a cabo por cuenta propia un proyecto de análisis de datos con su respectiva implementación en alguna plataforma computacional.

Detalle de los temas del curso

- Tema 1. Análisis exploratorio de datos (EDA)
 - 1.1 Aplicaciones de técnicas multivariadas. algunos problemas típicos
 - 1.2 La organización de los datos
 - 1.3 Uso de herramientas computacionales para el análisis y la visualización de datos
 - 1.4 Métodos gráficos para datos multivariados
- Tema 2. Vectores y matrices aleatorios y la distribución normal multivariada
 - 2.1 La distribución normal multivariada
 - 2.2 Inferencia:
 - 2.2.1 Acerca de vectores de medias
 - 2.2.2 Comparación de múltiples medias
 - 2.3 Modelos de regresión lineal multivariados
- Tema 3. Análisis de estructuras de covarianza
 - 3.1 Componentes principales
 - 3.2 Análisis de factores
 - 3.3 Análisis de correlación canónica
- Tema 4. Técnicas de agrupamiento y clasificación

- 4.1 Clasificación y análisis discriminante
- 4.2 Medidas de similitud y escalamiento multidimensional
- 4.3 Análisis de correspondencias
- Tema 5. Introducción al análisis de datos categóricos
 - 5.1 Tablas de contingencia
 - 5.2 Modelos lineales generalizados
 - 5.3 Regresión logística

Referencias generales del curso

Se espera que cada semana se lea un promedio de 40-50 páginas en lecturas asignadas y ustedes serán responsables de cubrir dichas lecturas. Más adelante especificaremos algunas revistas y artículos relevantes para el curso. Generales.

No hay un libro de texto único para este curso, principalmente:

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

Richard A. Johnson & Dean W. Wichern (2007). *Applied Multivariate Statistical Analysis*, Prentice Hall

Brian Everitt & Thorsten Hothorn (2011). *An Introduction to Applied Multivariate Analysis with R* Springer

Neil H. Timm (2002) *Applied Multivariate Analysis* Springer

Adicionalmente, hay muchas referencias que serán de utilidad. Las siguientes referencias son algunos textos que pueden ser relevantes para el curso, y no están dadas en orden de importancia. Algunas de las referencias se encuentran disponibles en la biblioteca del ITAM.

Libros más teóricos

- T. W. Anderson (2003) *An introduction to Multivariate Statistical Analysis* Third Ed. Wiley
- J. L. Schafer (1997) *Analysis of Incomplete Multivariate Data*, Chapman and Hall

Orientados a aplicaciones

- K. Gerald van den Boogart & Raimon Tolosana-Delgado (2013). *Analyzing compositional data with R*, Springer
- Tamrapamrni Dasu & Theodore Johnson (2003) *Exploratory data mining and Data Cleaning*, Wiley
- Ludwig Fahrmeir & Gerhard Tutz (1994) *Multivariate Statistical Modelling based on Generalized Linear Models*, Springer
- Wolfgang Karl Härdle & Léopold Simar (2009) *Applied Multivariate Statistical Analysis* Fourth Edition, Springer
- Edgar Santos Fernández (2012) *Multivariate Statistical Quality Control Using R* Springer Briefs in Statistics.

Algunos textos para conocer mejor R:

- Venables, W.N. y B.D. Ripley (2001). *Modern Applied Statistics with S-plus*, 4th. ed. Springer. Este libro provee un panorama general de metodología estadística a través de ejemplos en S-plus, que son casi 100% extrapolables a R.
- Becker, Richard A., John M Chambers y Allan R. Wilks (1988). *The new S Language. A programming environment for data analysis and graphics*, Wadsworth & Brooks/Cole. California.
- Spector, Phil (1994). *An introduction to S and S-Plus*, Duxbury Press, California.
- Murrell, Paul (2005). *R graphics*. Chapman and Hall.
- Templ, Matthias (2016). *Simulation for Data Science with R*, PACKT Publishing, UK.

Sobre métodos multivariados en particular:

- Ingwer Borg & Patrick J. F. Groenen (2005). *Modern Multidimensional Scaling: Theory and applications* Second Edition, Springer
- John M. Chambers & William S. Cleveland & Paul A. Tukey & Beat Kleiner (1983). *Graphical Methods for Data Analysis* Duxbury
- Trevor F. Cox & Michael A. A. Cox (2001) *Multidimensional Scaling* Chapman and Hall
- Brian S. Everitt & Sabine Landau & Morven Leese & Daniel Stahl (2011) *Cluster Analysis* Fifth Edition, Wiley
- I. T. Jolliffe (2002) *Principal Component Analysis* Second Edition, Springer
- Timothy R. C. Read & Noel A. C. Cressie (1988). *Goodness-Of-Fit Statistics for Discrete Multivariate Data* ,Springer