

Programa: Modelos Estadísticos Lineales

Profesor: Alejandra Tabares Pozos

Información General

Este curso de nivel maestría ofrece una visión integral de los modelos estadísticos lineales y su extensión a los modelos lineales generalizados (GLMs), haciendo uso del lenguaje de programación R. A lo largo del semestre, los estudiantes explorarán los fundamentos teóricos de la modelización estadística, la derivación matemática de los estimadores, la importancia de las suposiciones y el diagnóstico de los modelos. Además, se abordarán diferentes familias de distribución y tipos de datos, comprendiendo cómo adaptar las herramientas estadísticas a contextos y problemas variados en ciencia de datos.

Forma de evaluación del curso

- *Tareas Prácticas (aprox. 5): 25 %* Las tareas consistirán en ejercicios aplicados con R, análisis de datos reales, interpretación de resultados y discusión crítica.
- *Examen Parcial (Semana 8): 20 %* Prueba escrita (teórico-práctica) que evaluará la comprensión de los fundamentos teóricos, el ajuste e interpretación de modelos lineales y la introducción a los GLMs.
- *Proyecto Final (presentación en Semana 15): 30 %* Los estudiantes analizarán un conjunto de datos real o simulado, seleccionarán y ajustarán modelos estadísticos lineales y/o GLMs apropiados, y presentarán los resultados, conclusiones y recomendaciones. Se valorará la solidez metodológica, la corrección técnica, la claridad y el rigor al comunicar hallazgos.
- *Examen Final (Semana 16): 25 %* Prueba integradora (teórico-práctica) que abarcará todo el contenido del curso, desde los modelos lineales clásicos hasta las familias avanzadas de GLMs y sus diagnósticos.

Conocimientos Preliminares:

Matemática Básica: Familiaridad con álgebra lineal (operaciones con matrices, vectores, determinantes, inversas y descomposiciones básicas). Comprensión de funciones, derivadas, máximos y mínimos, así como conceptos de análisis básico de funciones reales.

Estadística y Probabilidad: Conocimiento de estadística descriptiva, distribuciones de probabilidad (especialmente las más comunes: normal, binomial, Poisson), intervalos de confianza, pruebas de hipótesis y teoría de la estimación. Entendimiento de conceptos como valor esperado, varianza, covarianza, independencia y función de densidad/probabilidad.

Inferencia Estadística: Familiaridad con el concepto de estimadores insesgados, consistentes y de mínima varianza. Conocimiento del método de estimación por mínimos cuadrados para la regresión lineal simple.

Recomendamos estos recursos en línea que pueden ayudar a refrescar su memoria:

- Algebra for educational scientists
- Estadística básica con R

Objetivos de Aprendizaje:

1. Dominar la formulación matemática, las suposiciones estadístico-probabilísticas, y el razonamiento que subyace a los estimadores por mínimos cuadrados y por máxima verosimilitud.
2. Aprender a implementar las herramientas prácticas para ajustar modelos lineales y GLMs, diagnosticar su validez, y comunicar resultados con rigor y claridad.
3. Entender cómo seleccionar la familia de distribución, el enlace y las metodologías más adecuadas según la naturaleza de los datos (continuos, binarios, conteos, etc.).
4. Evaluar supuestos, comparar modelos alternativos, revisar suposiciones, detectar problemas de ajuste y proponer soluciones, sustentando las decisiones con criterios estadísticos sólidos.

Competencias a Desarrollar en los Estudiantes:

Con la finalización del curso, se espera que los estudiantes sean capaces de enfrentar desafíos reales de análisis de datos con las herramientas y el criterio estadístico sólido que caracteriza al ámbito profesional y de investigación avanzada. De forma desagregada, las competencias mas relevantes pueden ser enunciadas como sigue:

1. Capacidad de entender la estructura matemática de los modelos lineales y lineales generalizados, así como las implicaciones teóricas de sus estimadores e inferencias.

2. Habilidad para utilizar eficazmente R en el ajuste, diagnóstico y comunicación de resultados de modelos estadísticos, aprovechando librerías relevantes y produciendo código limpio y reproducible.
3. Aptitud para identificar el tipo de modelo estadístico más adecuado según la problemática, evaluar la validez de las suposiciones, detectar anomalías en los datos y modelos, y proponer acciones correctivas o alternativas metodológicas.
4. Capacidad de traducir hallazgos estadísticos en información clara y útil para audiencias técnicas y no técnicas, argumentando las decisiones metodológicas y los resultados obtenidos con base en evidencia empírica y teórica.

Bibliografía

Bibliografía Principal

Texto Base:

- Dunn, P.K. & Smyth, G.K. (2018). *Generalized Linear Models With Examples in R*. Springer.

Bibliografía Secundaria

Modelos Lineales y Regresión:

- Kutner, M.H., Nachtsheim, C.J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill.
- Montgomery, D.C., Peck, E.A., & Vining, G.G. (2012). *Introduction to Linear Regression Analysis*. Wiley.
- Faraway, J.J. (2004). *Linear Models with R*. Chapman & Hall/CRC.

Generalized Linear Models (GLMs) y Categorías Especiales:

- McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*. Chapman & Hall/CRC.
- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley.
- Agresti, A. (2013). *Categorical Data Analysis*. Wiley.
- Dobson, A.J. & Barnett, A.G. (2008). *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC.

Perspectivas Ampliadas:

- Casella, G. & Berger, R.L. (2002). *Statistical Inference*. Duxbury.

- Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R*. Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

Semana 1: Contexto y Rol de los Modelos Estadísticos (Cap. 1)

Contenidos: Filosofía de la modelización estadística, variabilidad sistemática vs. aleatoria, representaciones gráficas, codificación de factores, introducción a R.

Bibliografía secundaria:

- James, G., Witten, D., Hastie, T., Tibshirani, R. *An Introduction to Statistical Learning*, Cap. 2.
- Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Caps. 1–2.

Recursos web/videos:

- RStudio Education for Beginners
- R for Data Science Online

Paquetes R recomendados: tidyverse, skimr, summarytools

Semana 2: Estimación por Mínimos Cuadrados y Propiedades Teóricas (Cap. 2.1–2.3)

Contenidos: Derivación OLS, propiedades del estimador, inferencia sobre coeficientes.

Bibliografía secundaria:

- Montgomery, D.C., Peck, E.A., Vining, G.G. *Introduction to Linear Regression Analysis*, Caps. 1–2.
- Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W. *Applied Linear Statistical Models*, Cap. 2.

Recursos web/videos:

- Simple Linear Regression: The Least Squares Regression Line

- Penn State STAT 501 Materials

Paquetes R recomendados: broom, car

Semana 3: Extensión a Regresión Múltiple y Formulación Matricial (Cap. 2.4–2.5)

Contenidos: Regresión múltiple, ANOVA para modelos anidados, notación matricial.

Bibliografía secundaria:

- Draper, N.R., Smith, H. *Applied Regression Analysis*.
- Faraway, J.J. *Linear Models with R*, Caps. 2–3.

Recursos web/videos:

- mumfordbrainstats: Multiple linear regression with matrices (YouTube)
- Matrix Approach to Multiple Linear Regression (YouTube)
- R-Bloggers: artículos sobre regresión múltiple

Paquetes R recomendados: GGally, car

Semana 4: Diagnóstico del Modelo Lineal y Construcción de Modelos (Cap. 3)

Contenidos: Supuestos, residuos, outliers, transformaciones (Box–Cox), variables categóricas.

Bibliografía secundaria:

- Fox, J., Weisberg, S. *An R Companion to Applied Regression*.
- Cook, R.D., Weisberg, S. *Residuals and Influence in Regression*.

Recursos web/videos:

- Jarad Niemi:Regression Diagnostics
- Mikko Rönkkö: Regression diagnostics and analysis workflow (versión extendida del primero)

Paquetes R recomendados: car, MASS (Box–Cox)

Semana 5: Más Allá de la Regresión Lineal: Máxima Verosimilitud (Cap. 4)

Contenidos: Limitaciones del modelo normal, MLE, información de Fisher.

Bibliografía secundaria:

- Casella, G., Berger, R.L. *Statistical Inference*.
- Rice, J.A. *Mathematical Statistics and Data Analysis*.

Recursos web/videos:

- StatQuest: Maximum Likelihood
- Linear Regression: MLE for Methods of Estimation

Paquetes R recomendados: `bbmle`, `optim()` (base R)

Semana 6: Extensión a Modelos No Normales e Idea General de GLMs (Cap. 4)

Contenidos: Motivación hacia GLMs, familia exponencial.

Bibliografía secundaria:

- Agresti, A. *Foundations of Linear and Generalized Linear Models*, Caps. 1–2.
- McCullagh, P., Nelder, J.A. *Generalized Linear Models*, Cap. 1.

Recursos web/videos:

- Stanford online: Statistical Learning: 4.8 Generalized Linear Models (YouTube)

Paquetes R recomendados: `stats` (`glm`), `family()`

Semana 7: Estructura de los GLMs (Cap. 5)

Contenidos: Definición GLM, función enlace, EDM, devianza.

Bibliografía secundaria:

- Agresti, A. *Categorical Data Analysis*, secciones introductorias a GLMs.
- Dobson, A.J., Barnett, A.G. *An Introduction to Generalized Linear Models*, Caps. 1–3.

Recursos web/videos:

- GLM Exponential Family, Example with Normal distribution finding functions a, b, and c

Paquetes R recomendados: `faraway`, `glm2`

Semana 8: Estimación en GLMs (Cap. 6) y Examen Parcial

Contenidos: IWLS, ecuaciones de puntuación, examen parcial.

Bibliografía secundaria:

- McCullagh, P., Nelder, J.A. *Generalized Linear Models*, Cap. 2.
- Fahrmeir, L., Tutz, G. *Multivariate Statistical Modelling Based on GLMs*.

Recursos web/videos:

- StatQuest videos sobre IWLS (YouTube)
- R-Bloggers: Búsqueda "IWLS"

Paquetes R recomendados: statmod, MASS

Receso Académico (Semana 9)

Semana 10: Inferencia en GLMs (Cap. 7)

Contenidos: Pruebas Wald, LRT, análisis devianza, intervalos.

Bibliografía secundaria:

- Agresti, A. *An Introduction to Categorical Data Analysis*.
- Collett, D. *Modelling Binary Data*.

Recursos web/videos:

- Pablo Inchausti Estadística: GLM teoría (YouTube)

Paquetes R recomendados: AICcmodavg, profile() (base R)

Semana 11: Diagnóstico de GLMs (Cap. 8)

Contenidos: Residuos en GLMs, leverage, outliers, remedios.

Bibliografía secundaria:

- Fox, J., Weisberg, S. *An R Companion to Applied Regression*.
- Atkinson, A.C., Riani, M. *Robust Diagnostic Regression Analysis*.

Recursos web/videos:

- Notes for Predictive Modeling

Paquetes R recomendados: DHARMA, car

Semana 12: Modelos Binomiales (Cap. 9)

Contenidos: Logístico, probit, cloglog, sobre-dispersión binomial.

Bibliografía secundaria:

- Agresti, A. *Categorical Data Analysis*.
- Gelman, A., Hill, J. *Data Analysis Using Regression and Multilevel Models*.

Recursos web/videos:

- StatQuest: Logistic Regression

Paquetes R recomendados: `epicalc`, `effects`, `ResourceSelection`

Semana 13: Modelos Poisson y Binomial Negativa (Cap. 10)

Contenidos: GLMs para conteo, Poisson, binomial negativa, sobredispersión.

Bibliografía secundaria:

- Cameron, A.C., Trivedi, P.K. *Regression Analysis of Count Data*.
- Hilbe, J.M. *Negative Binomial Regression*.

Recursos web/videos:

- Penn State STAT 504 Analysis of Discrete Data

Paquetes R recomendados: `pscl`, `MASS` (`glm.nb`)

Semana 14: Modelos Gamma e Inversa Gaussiana (Cap. 11)

Contenidos: GLMs para datos continuos positivos, Gamma, Inverse Gaussian.

Bibliografía secundaria:

- Dunn, P.K., Smyth, G.K. Artículos sobre EDM y Tweedie.
- Jørgensen, B. *Exponential Dispersion Models*.

Recursos web/videos:

- Gamma Regression in R (YouTube).

Paquetes R recomendados: `flexplot`, `gam`

Semana 15: Tweedie GLMs (Cap. 12) y Proyecto Final

Contenidos: Familia Tweedie, combinando conteo y continuo, índice de potencia, presentaciones finales.

Bibliografía secundaria:

- Dunn, P.K., Smyth, G.K. (2005), trabajos sobre Tweedie.
- Jørgensen, B. *The Theory of Linear Models*.

Recursos web/videos:

- Intro to Tweedie Distributions

Paquetes R recomendados: `tweedie`, `cplm`, `mgcv`

Semana 16: Examen Final y Cierre

Contenidos: Examen final integrador, discusión, perspectivas futuras.