

## **MÁSTER EN ANÁLISIS POLÍTICO Y ELECTORAL, 2023-2024**

### **Análisis de Encuestas**

Profesora: Marga Torre

E-mail: [margarita.torre@uc3m.es](mailto:margarita.torre@uc3m.es)

Despacho: 18.2.D08 (Campus de Getafe)

### **DESCRIPCIÓN DEL CURSO**

El análisis de datos es un elemento clave en la carrera de profesionales que, cada vez más a menudo, se encuentran con la necesidad de obtener, gestionar y analizar información de diferente naturaleza. Este curso se centra en el aprendizaje de técnicas de análisis de datos estructurados y no estructurados, desde una perspectiva aplicada. Todos los temas serán abordados de manera teórico-práctica, haciendo uso del lenguaje de programación R.

El curso se estructura en 4 grandes bloques, cada uno de ellos compuesto por diferentes temas:

1. Análisis exploratorio y limpieza de datos
  - a. Limpieza y preparación de los datos
  - b. Análisis univariado
  - c. Introducción al tratamiento de casos perdidos
2. Análisis bivariado
  - a. Variables continuas
  - b. Variables categóricas
3. Técnicas estadísticas multivariantes
  - a. Regresión lineal multivariante
  - b. Algoritmos de clasificación
  - c. Algoritmos de caja negra
4. Minería de opiniones
  - a. Análisis de textos: minería de opiniones en las Redes Sociales
  - b. Análisis de sentimiento

### **CRITERIOS DE EVALUACIÓN**

Componentes de la nota final:

- Participación en clase: 10%
- Prácticas evaluadas realizadas en clase\*: 90%

\*Es necesario tener una calificación de 4 o superior en todas las prácticas para hacer media entre los diferentes ejercicios

## CRONOGRAMA DE LA ASIGNATURA \*\*

### **SEMANA 1: Limpieza y preparación de los datos**

- Limpieza y preparación de variables
- Los fundamentos de R Markdown

### **SEMANA 2: La prueba estadística**

- Pruebas de decisión estadística
  - o Recordatorio: prueba t
  - o Recordatorio: intervalos de confianza
- Comparación de dos muestras
  - o Comparación de medias en muestras independientes
  - o Comparación de medias en muestras independientes
  - o Comparación de varianzas

### **SEMANA 3: Análisis bivariado I**

- Relación entre variables numéricas
  - o Covarianza
  - o Correlación de Pearson
- El Análisis de Correlación en R
- Correlación vs causalidad

### **SEMANA 4: Análisis biariado II**

- Relación entre variables categóricas (atributos)
  - o Tablas de contingencia
  - o Análisis de los residuos
  - o Chi<sup>2</sup>
  - o Gazapos estadísticos

### **SEMANA 5**

- Ejercicio práctico evaluado (20%)

### **SEMANA 6: Regresión lineal (I)**

- Regresión lineal simple
  - o Objetivos
  - o Mínimos Cuadrados Ordinarios (MCO)
  - o Bondad de ajuste del modelo
  - o Las variables en la regresión lineal
  - o Los resultados de la regresión
  - o Evaluación de los supuestos de la regresión lineal
- Ejemplo práctico
  - o Preparación de variables y análisis preliminar
  - o Estimación del modelo de regresión simple
  - o Interpretación de los resultados
  - o Bondad de ajuste del modelo

- Diagnóstico de la regresión
- Exportación y presentación de los resultados
  - Librería *stargazer*
  - Librería *jtools*

### **SEMANA 7: Regresión lineal (II)**

- El modelo de regresión lineal múltiple
- Ejemplo práctico:
  - Preparación de variables y análisis preliminar
  - Estimación del modelo de regresión múltiple
  - Interpretación de los resultados
  - Visualización de los coeficientes
  - ¿Qué variable es la más importante?
  - Bondad de ajuste del modelo
  - Diagnóstico de la regresión
  - Valores predichos
- Exportación de los resultados
  - Librería *stargazer*
  - Librería *jtools*
- Comparación de modelos: AIC, BIC

### **SEMANA 8: Validación del modelo de regresión lineal**

- Explicación vs predicción
- El equilibrio entre el sesgo y la varianza
- Validación del modelo de regresión lineal
  - El conjunto de validación
  - Método *Leave-one-out* (LOOCV)
  - Método K-fold

### **SEMANA 10: Regresión logística (I)**

- El análisis de clasificación
- Regresión logística vs regresión lineal
  - Supuestos básicos de la regresión logística
  - El modelo de regresión logística
- Estimación del modelo
  - Significatividad de las variables
  - Interpretación de los coeficientes
    - Primera transformación de los coeficientes: *Odds ratio*
    - Segunda transformación de los coeficientes: Probabilidades predichas
- Diagnóstico de la regresión

### **SEMANA 11: Regresión logística (II)**

- Los efectos Marginales
- Exportación y presentación de resultados
- Ejercicio práctico de regresión logística

## **SEMANA 12: Validación de los modelos de clasificación**

- Validación del modelo de clasificación
  - o Matriz de confusión
  - o Curva ROC
- Mejora de modelos
  - o Efectos cuadráticos
  - o Interacciones

## **SEMANA 13**

- Ejercicio práctico evaluado (70%)

## **SEMANA 14: Introducción al text mining**

- Introducción al *text mining*
  - o ¿Qué es la minería de texto?
  - o Librerías
  - o Datos
  - o Corpus de palabras
  - o Nube de palabras
  - o Extracción de los términos más frecuentes
- El análisis de sentimiento
  - o Preparación del conjunto de entrenamiento y validación
  - o Análisis
    - Logit
    - Naive Bayes
    - LDA - Linear Discriminant Analysis
    - Random Forest
  - o Comparación de modelos
  - o Clasificación
- Conclusiones y propuestas de mejora

\*\* El cronograma es orientativo y puede sufrir variaciones en función del nivel y ritmo de la clase

## **BIBLIOGRAFÍA BÁSICA**

- Wickham, Hadley & Garret Golemund. 2020. [R for Data Science](#). O'Reilly

## **BIBLIOGRAFÍA COMPLEMENTARIA**

- García Ferrando, Manuel y Modesto Escobar. 2017. *Socio-estadística*. Alianza Editorial
- Guillén, Mauro. 2014. *Análisis de regresión múltiple*. Cuadernos Metodológicos 4. CIS, Madrid
- James, Gareth, Daniel Witten, Trevor Hastie, & Robert Tibshirani. 2013. *An introduction to Statistical Learning with applications in R*. Springer

- Long, Scott J. & Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables*
- Peña, Daniel & Juan romo. *Introducción a la Estadística para las Ciencias Sociales*. McGraw Hill
- Silge, Julia and David Robinson. 2018. *Text mining with R. A tidy approach*. O'Reilly