

# Proyecto final de Modelos Lineales

## Definición

Es una tarea para realizar a lo largo del cuatrimestre, en grupos de 2 a 3 miembros.

Consiste en analizar un banco de datos, a vuestra elección, con el que penséis que se pueden plantear preguntas interesantes, esto es, objetivos a estudiar. En el análisis de esos datos debéis utilizar buena parte de las técnicas estudiadas en esta asignatura, especialmente las tratadas en los cuatro primeros temas.

## Programación

El proyecto tiene dos etapas:

En la primera etapa, debéis preparar un primer entregable, Tarea PF1, antes del jueves 6 de abril, que no debe ocupar más de una hoja. En él debéis incluir:

1. Miembros del grupo.
2. Descripción del banco de datos: dimensiones, nombres de las variables y un (muy) pequeño análisis exploratorio univariante del banco de datos. También un enlace al banco de datos.
3. Objetivos del estudio, definidos clara y concisamente, e indicación de cuál es la variable respuesta.
4. Comentarios y/o inquietudes que os hayan surgido.

Una vez os dé el visto bueno para ese análisis comenzará la segunda etapa: el análisis de los datos y extracción de conclusiones. En esa segunda etapa debéis redactar un documento, Tarea PF2, a entregar antes del 22 de mayo. Ese documento tendrá forma de artículo de investigación, de cinco o seis páginas como máximo, incluyendo gráficas y tablas. En él debéis incluir:

1. **Título y autores del trabajo.**
2. **Resumen.** Se trata de un pequeño párrafo resumiendo objetivos y resultados.
3. **Apartado 1, Introducción.** Comentando el banco de datos, los objetivos propuestos y las técnicas utilizadas para su análisis.
4. **Apartado 2, Resultados.** Recogiendo todos los resultados que consideréis interesantes, incluyendo tablas y gráficas.
5. **Apartado 3, Conclusiones.** Redacción de las conclusiones de vuestros resultados y, en su caso, comentarios sobre los puntos débiles que penséis que tiene vuestro análisis y cómo se podrían seguir estudiando esos objetivos.

## Repositorios de datos:

Podéis utilizar los datos que queráis, siempre que su análisis esté relacionado con esta asignatura. Por ello necesitaréis mi visto bueno a la Tarea PF1.

Un listado de repositorios de datos muy interesante, donde podéis encontrar bancos de datos, es:

KDD Nuggets: <https://www.kdnuggets.com/datasets/index.html>

Allí, entre muchos otros, tendréis acceso a los repositorios:

Kaggle

<https://www.kaggle.com/datasets>

Statlib Dataset Archive:

<http://lib.stat.cmu.edu/datasets/>

UCI Machine learning repository:

<http://archive.ics.uci.edu/ml/index.php>