

**Laboratorio 1**  
**Estadística Computacional**  
Universidad Técnica Federico Santa María  
Departamento de Informática

Héctor Allende Olivares  
<hallende@inf.utfsm.cl>

Erick López Ovando  
<elopez.ovando@gmail.com>

Camilo Valenzuela Carrasco  
<camilo.valenzuela@alumnos.usm.cl>

21 de Octubre del 2015

## 1. Objetivo

Realizar un análisis exploratorio de datos, utilizando todas las herramientas aprendidas en la experiencia.

Se busca que el alumno utilice las herramientas correctas para responder las preguntas, y haga un análisis completo de los resultados, no sólo mostrarlos.

## 2. Motivación

Nuestro querido ayudante de cátedra Rodrigo Naranjo, luego de años de ahorrar su sueldo de ayudante, ha decidido comprar un Auto.

Pero no puede ser cualquier Auto, si va a gastar sus ahorros de toda una vida, mejor comprar algo con estilo, por eso está en búsqueda de un Auto antiguo, una verdadera joyita para lucirse mientras maneja.

El problema es que todavía no decide qué Auto comprar, tiene que ser un auto rápido, pero también hay que pensar en el medio ambiente por lo que no tiene que gastar mucha bencina.

Por este motivo nos ha pedido al curso de Estadística Computacional realizar un análisis estadístico de los datos existentes, y así tomar una decisión más informada.

### 3. Datos a utilizar

Para el análisis se utilizará un conjunto de datos del UCI Repository llamado Auto MPG<sup>1</sup>, éste contiene 398 instancias, cada una de ellas tiene los datos sobre un modelo de Auto. Cada instancia tiene 9 variables

- 1.- mpg (Miles per Galon): Consumo de combustible.
- 2.- cylinders: Cantidad de cilindros del motor.
- 3.- displacement: Cilindrada del motor<sup>2</sup>.
- 4.- horsepower: Caballos de fuerza (Potencia).
- 5.- weight: Peso del modelo.
- 6.- acceleration: Aceleración del modelo.
- 7.- model year: Año del modelo.
- 8.- origin: Lugar de construcción.
- 9.- car name: Nombre del modelo.

Como supuesto diremos que el conjunto de datos contiene todos los modelos que salieron a la venta en los distintos años.

Los archivos necesarios para la realización del laboratorio se encontrarán en Moodle.

### 4. Preguntas

- 1.- Antes de comenzar, estudie la dispersión de las variables, solo en los casos que sea posible. ¿Cuál variable es más Homogénea?. (Hint: Utilice el coeficiente de variación).
- 2.- En relación a la cantidad de modelos que se generan por año, como es el comportamiento del mercado automotriz.
- 3.- ¿Está relacionado la cantidad de cilindros y el cilindraje de un Auto? ¿Pasa lo mismo si vemos la relación entre la cantidad de cilindros y la aceleración del auto?. Justifique.
- 4.- Entre qué años se encuentran los modelos con mayor potencia. Justifique.
- 5.- ¿Qué origen tienen los autos que consumen menor cantidad de combustible?
- 6.- Sobre el conjunto de autos de origen con menor consumo de combustible, ¿Existe alguna relación entre la aceleración del modelo y los caballos de fuerza que este tiene? Argumente.
- 7.- Tomando en cuenta los análisis realizados en los puntos 3) y 4), tome la cantidad de cilindros que obtienen la mayor aceleración, y un rango de años donde la potencia es la mayor. De ese subconjunto de datos elija el año donde el consumo de gasolina sea el menor, y muestre que tan menor es el valor en comparación a los otros años. (Sea lo más preciso que pueda).
- 8.- Luego de realizado el análisis en los distintos puntos, concluya sobre cuál es la mejor opción para que nuestro Ayudante se luzca en las calles de Valparaiso.

---

<sup>1</sup><https://archive.ics.uci.edu/ml/datasets/Auto+MPG>

<sup>2</sup><https://es.wikipedia.org/wiki/Cilindrada>

## 5. Sobre la Entrega

- 1.- El informe debe ser realizado en  $\text{\LaTeX}$ .
- 2.- Estructura del informe: El informe tiene que contener lo siguiente:
  - 2.1.- Introducción.
  - 2.2.- Desarrollo.
  - 2.3.- Conclusiones generales.
  - 2.4.- Referencias, si es que utilizaron. (Tienen que estar citadas en el informe).
- 3.- Los calculos y gráficos deben realizarse utilizando R, este código no tiene que estar incluido dentro del informe.
- 4.- El informe y el código utilizado tiene que ser subido a moodle en formato nombreCompleto1-rol1-nombreCompleto2-rol2.tar.gz.
- 5.- Fecha de entrega: **Miércoles 28 de Octubre**. Se descontarán 20 puntos por cada día de atraso.