

## Escuela de Ingeniería Informática

### Métodos Estadísticos, Curso 2019-20

#### Ejercicios Prácticos Lab4

**Ejercicio 1.** Obtener del Instituto Canario de Estadística ([ISTAC](#)) la distribución por edades de la población entre los años 2000 a 2017. Representar las pirámides de población correspondientes con la librería *pyramid*, y analizar la evolución anual. Razonar las conclusiones y realizar una pequeña animación (en formato GIF o similar) de la evolución de las pirámides de población en esos años.

Para obtener más fácilmente los datos se puede ir directamente a la dirección: ([link](#)) y consultar los datos de “población según sexos y grupos de edad grandes y quinquenales”.

**Ejercicio 2 (Opcional).** Con datos del [INE](#) que indican la proyección 2018-2068 de la evolución de la población residente en España a 1 de enero, por sexo, edad y año:

<http://www.ine.es/jaxi/Tabla.htm?path=/t20/p278/p01/2018-2068/pob/I0/&file=01001.px&L=0>

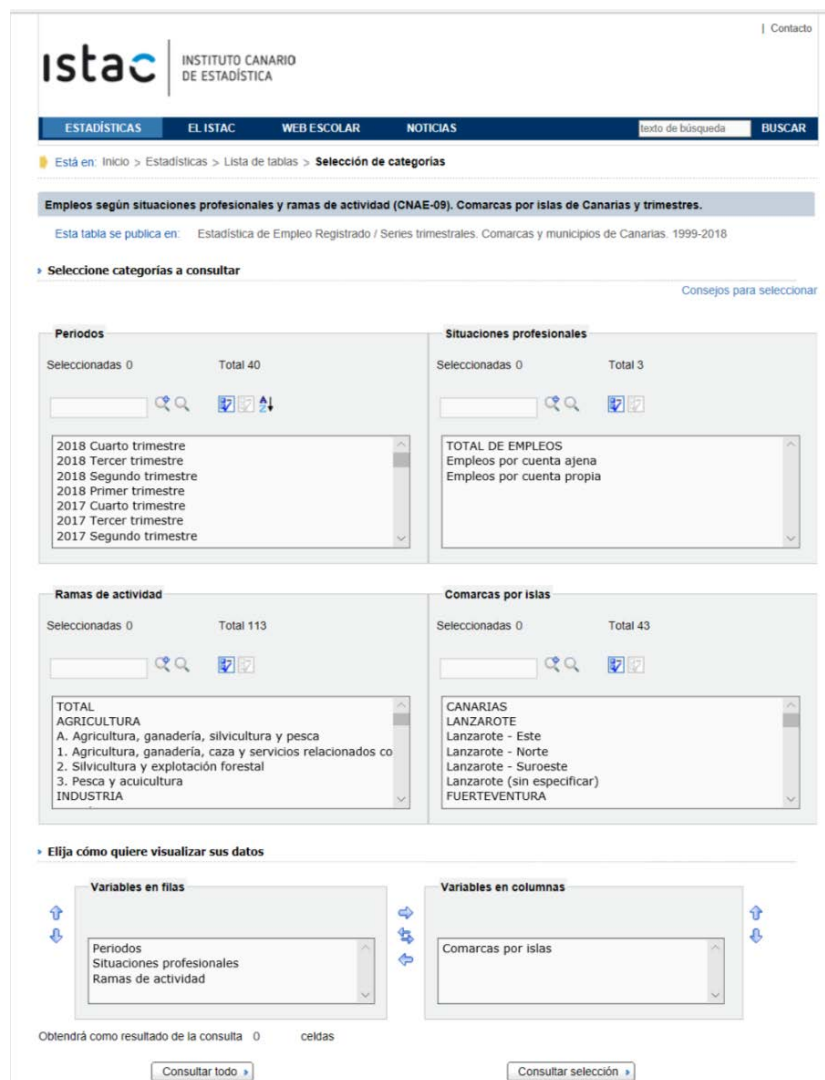
Desarrollar una animación con los datos poblaciones del periodo 2018-68 que permita visualizar las proyecciones y los cambios. Analizar y establecer conclusiones al respecto con su posible proyección a Canarias.

**Ejercicio 3.** El fichero “*germinacion.csv*” contiene los datos de germinación de semillas de dos genotipos de la planta parásita Orobanché y dos extractos de plantas huésped (judía y pepino) que se utilizaron para estimular la germinación. La variable “*count*” representa el número de semillas que germinaron de un lote de tamaño “*sample*”. Con estos datos se pide:

- Crear un data frame con los datos de la variable “*count*” y una columna adicional que incluya en número de semillas que no germinó.
- Calcular los parámetros de centralización y dispersión del conjunto de muestras para cada genotipo y tipo de planta huésped y analizar gráficamente el efecto del genotipo en la germinación. Explicar las conclusiones.
- Utilizar la función *lm()* para ver la tendencia e influencia de los genotipos en la germinación. ¿Son estadísticamente independientes las variables de genotipo (“*Orobanché*”) y de tipo de planta huésped (“*extract*”)? Razonar y justificar las respuestas.

**Ejercicio 4.** Obtener del Instituto Canario de Estadística ([ISTAC](http://istac.gob.es)) el fichero con los datos de empleo en actividades relacionadas con la Ingeniería Informática (códigos de rama de actividad CNAE-09: 62,63 y 95) en Canarias por trimestres en el periodo 2009 a 2018. (ojo con el formato de los datos y los separadores de decimales). Se pide

- Analizar gráficamente la variación de cada tipología de empleo en las Islas Canarias (por islas y totales) en el periodo considerado e intentar explicar sus valores singulares.
- Utilizando la librería **mgcv()**, encontrar un modelo de seguimiento del empleo representar gráficamente su evolución y predicciones (efectuar una posible a diciembre de 2018).
- Encontrar la isla donde hay más empleo y en qué etapa.
- Analizar comparativamente la evolución durante dos años del empleo en dos islas diferentes y explicar sus variaciones y sus aspectos comunes.



The screenshot shows the ISTAC (Instituto Canario de Estadística) website interface for selecting data. The main navigation bar includes 'ESTADÍSTICAS', 'EL ISTAC', 'WEB ESCOLAR', and 'NOTICIAS'. A search bar is present with the text 'texto de búsqueda' and a 'BUSCAR' button.

The breadcrumb trail indicates the current path: 'Está en: Inicio > Estadísticas > Lista de tablas > Selección de categorías'.

The main heading is 'Empleos según situaciones profesionales y ramas de actividad (CNAE-09). Comarcas por islas de Canarias y trimestres.' Below this, it states 'Esta tabla se publica en: Estadística de Empleo Registrado / Series trimestrales. Comarcas y municipios de Canarias. 1999-2018'.

The 'Selección de categorías a consultar' section is divided into four panels:

- Periodos:** Total 40. Selected: 0. The list includes: 2018 Cuarto trimestre, 2018 Tercer trimestre, 2018 Segundo trimestre, 2018 Primer trimestre, 2017 Cuarto trimestre, 2017 Tercer trimestre, 2017 Segundo trimestre.
- Situaciones profesionales:** Total 3. Selected: 0. The list includes: TOTAL DE EMPLEOS, Empleos por cuenta ajena, Empleos por cuenta propia.
- Ramas de actividad:** Total 113. Selected: 0. The list includes: TOTAL, AGRICULTURA, A. Agricultura, ganadería, silvicultura y pesca, 1. Agricultura, ganadería, caza y servicios relacionados co, 2. Silvicultura y explotación forestal, 3. Pesca y acuicultura, INDUSTRIA.
- Comarcas por islas:** Total 43. Selected: 0. The list includes: CANARIAS, LANZAROTE, Lanzarote - Este, Lanzarote - Norte, Lanzarote - Suroeste, Lanzarote (sin especificar), FUERTEVENTURA.

Below these panels, there is a section 'Elija cómo quiere visualizar sus datos' with two sub-sections:

- Variables en filas:** Includes Periodos, Situaciones profesionales, and Ramas de actividad.
- Variables en columnas:** Includes Comarcas por islas.

At the bottom, it states 'Obtendrá como resultado de la consulta 0 celdas' and provides two buttons: 'Consultar todo' and 'Consultar selección'.

**Ejercicio 5. (opcional)** Considérese el data set “*mtcars*” de **R**. Este data set contiene la información sobre varios modelos de vehículos:

```
# DATOS ORIGINALES mtcars
# =====
#[, 1]   mpg   Miles/(US) gallon
#[, 2]   cyl   Number of cylinders
#[, 3]   disp  Displacement (cu.in.)
#[, 4]   hp    Gross horsepower
#[, 5]   drat  Rear axle ratio
#[, 6]   wt    Weight (1000 lbs)
#[, 7]   qsec  1/4 mile time
#[, 8]   vs    Engine (0 = V-shaped, 1 = straight)
#[, 9]   am    Transmission (0 = automatic, 1 = manual)
#[,10]   gear  Number of forward gears
#[,11]   carb  Number of carburetors
```

Con ellos realizar las siguientes operaciones:

- a) Obtener un data set transformado cuyas columnas expresen magnitudes en unidades del sistema métrico decimal. Para la variable **qsec** (tiempo en recorrer ¼ de milla) considérese el equivalente a 100 metros. Análogamente para la variable **mpg** (millas (US) por galón) tomar los kilómetros realizados por litro de combustible. En general, pueden utilizarse los siguientes factores de conversión:

```
# FACTORES CONVERSION
# =====
# 1 mpg = 0.5 km/l
# 1 cubic inch = 16.3871 centímetros cúbicos
# 1 hp = 1.01386 cv
# 1000 lbs = 453.592 kg
# 0,25 mile = 402.236 metros
# sec_100= qsec *(100/402.236)=qsec*0.2486 en seg a 100 metros
```

- b) Añadir al data set de a) una nueva columna/variable que contenga una medida de la aceleración (de 0 a 100 metros), medida en **m/s<sup>2</sup>**, de los modelos de vehículos incluidos en el data set original.
- c) Diseñar una función que permita obtener una predicción del consumo, expresado en litros por cada 100 Km, desde las variables representativas del Número de Cilindros, Cilindrada, Aceleración y Tipo de Transmisión.
- d) Diseñar una aplicación interactiva con **Shiny** que permita visualizar los resultados de la predicción de consumo (en L/100Km) en función de tres datos de entrada (Cilindrada, Aceleración y Tipo de Transmisión).