

# Taller de R: Estadística y Programación

## Taller 3

Octubre 25 de 2021

En este taller se evalúan los módulos 4, 5 y 6 del curso. Se presentan 2 tipos de taller (A y B), pero usted solo debe desarrollar 1 de ellos. Para realizar este trabajo, podrá hacer grupo de hasta tres personas. Sea creativo en su código (no hay una respuesta única, todos los métodos que permitan obtener la misma respuesta son validos). Cuando encuentre una ayuda en línea que le permita solucionar algún problema, no olvide citar la fuente. Por último, lea atentamente las instrucciones del taller.

## Instrucciones

- No seguir las instrucciones tiene una penalización del **10%** de la nota total.
- Este taller representa el 40% de la nota del curso.
- El taller estará disponible en Bloque Neón desde las 22:00 horas del 25 de octubre de 2021.
- Al recibir el taller, uno de los integrantes del grupo debe crear un repositorio llamado **task-3** para almacenar las respuestas del taller. Este repositorio debe contener como mínimo las siguientes carpetas: **data** y **source**. El usuario que crea el repositorio debe invitar como colaborador a los demás integrantes del grupo y al usuario [eduard-martinez](#).
- El taller debe ser terminado antes de las 23:59 horas del 5 de diciembre de 2021. No se calificará ningún commit que sea hecho después de esa fecha.
- Por favor sea lo más organizado posible y comente paso a paso cada línea de código, pero recuerden **NO** usar ningún acento o carácter especial dentro del código para evitar problemas al abrir los script en los diferentes sistemas operativos.
- En las primeras líneas del script debe escribir los nombres y los códigos de las personas que trabajaron en el código.
- Recuerde que debe elegir y desarrollar solo 1 taller (A o B).

## Taller A

### 1. Datos espaciales (50%)

Toda la información empleada en esta parte del taller fue tomada del Marco Geo estadístico Nacional - MGN del DANE. Puede consultar los datos [aquí](#).

#### 1.1. Importar datos espaciales

**1.1.1** De la carpeta *data/output* importe los shapefiles de VIAS (llame al objeto *via*) y MGN\_URB\_TOPONIMIA (llame al objeto *puntos*). El primero contiene las vías del departamento de Norte de Santander y el segundo la ubicación de algunos servicios públicos (hospitales, policía,...) y comerciales (restaurantes, tiendas,...).

**1.1.2** Cree un nuevo objeto llamado *c\_medico*, que contenga las observaciones del objeto *puntos* en las que la variable CSIMBOL sea igual a “021001”, “021002” o “021003”.

**1.1.3** De la carpeta *data/output* importe los *c poblado (2017).rds* (llame al objeto *c\_poblado*), *dp deptos (2017).rds* (llame al objeto *depto*) y *\textit{victimas\_map-muse.rds}* (llame al objeto *mapmuse*). Asegúrese de dejar únicamente los centro poblados con código DANE  $\geq 54001$  &  $< 55000$ . Además, del objeto *depto*, deje únicamente el polígono del departamento de Norte de Santander.

## 1.2. Atributos de los objetos

Aplice la función *skim* de la librería *skimr* para explorar todos los objetos cargados en el punto anterior. Si considera necesario, seleccione algunas variables y pinte sobre la consola la tabla de frecuencia estas.

## 1.3. Geometrías del objeto

**1.3.1** Para todos los objetos del punto 1.1., pinte sobre la consola la caja de coordenadas (*st\_bbox*) y el CRS de cada objeto.

**1.3.2** Ahora va a re proyectar el CRS de todos los objetos. Asigne la siguiente CRS **+proj=utm +zone=19 +datum=WGS84 +units=m +no\_defs** a todos los objetos del punto 1.1..

## 1.4. Operaciones geométricas

**1.4.1** Use el objeto *depto* para hacer *clipping* y dejar los puntos de *mapmuse* que están debajo del polígono de Norte de Santander.

**1.4.2** Del objeto *c\_poblado*, seleccione cualquier municipio, use este polígono y el objeto *via*, para calcular el largo de las vías en el centro poblado que seleccionó.

## 1.5. Pintar mapas

**1.5.1** Use la función *leaflet* para visualizar en un mismo mapa: los polígonos de los centros poblados, el polígono del departamento de Norte de Santander y los hospitales y puestos de salud del objeto *c\_medicos*.

**1.5.2** Use las librerías *ggplot*, *ggsp* y las demás que considere necesarias para visualizar en un mismo mapa: los polígonos de los centros poblados, el polígono del departamento de Norte de Santander y los hospitales y puestos de salud del objeto *c\_medicos*. Asegúrese de poner la barra de escalas, la estrella del norte y las etiquetas que permitan diferencias cada uno de los objetos. Exporte el mapa en formato .pdf a la carpeta *views*.

# 2. Regresiones (30%)

La Oficina del Alto Comisionado para la Paz (OACP) tiene el registro oficial de víctimas por minas antipersona (MAP) y municiones sin explotar (MUSE). Estos registros pueden obtenerse en la página oficial de la [OACP](#). Para este ejercicio usted cuenta con una base de datos que contiene los registros de las víctimas de MAP-MUSE en el departamento de Norte de Santander. La variable *fallecido* toma el valor de 1 si la persona fallece en el accidente y 0 si resulta herida.

**2.1.** Importe el archivo *data/output/df\_mapmuse.rds* y estime un modelo de probabilidad lineal en el que *fallecido* es la variable dependiente. Y use las demás variables como variables explicativas. Almacene los resultados de la estimación en un objeto llamado *ols*.

**2.2.** Exporte a la carpeta *views* los gráficos con los coeficientes (*coef-plot*) de las estimaciones.

**2.3.** Ahora estime la ecuación del punto 2.1. usando un modelo logit y un modelo probit, almacene los resultados de las estimaciones en dos objetos llamados *logit* y *probit* respectivamente.

**2.4.** Exporte los resultados de los tres modelos en una misma tabla usando la función *outreg*.

**2.5.** De los objetos *logit* y *probit* exporte a la carpeta *views* dos gráficos con el efecto marginal de la distancia a un centro medico sobre la probabilidad de fallecer.

### 3. Web-scraping (10%)

**3.1.** Desde la consola de Rstudio lea la siguiente url [https://es.wikipedia.org/wiki/Departamentos\\_de\\_Colombia](https://es.wikipedia.org/wiki/Departamentos_de_Colombia) y cree un objeto que contenga el HTML de la página como un objeto *xml\_document*.

**3.2.** Use el *xpath* para extraer el título de la página (Departamentos de Colombia).

**3.3.** Extraiga la tabla que contiene los departamentos de Colombia.

## Taller B

### 1. Datos espaciales (40%)

Toda la información empleada en esta parte del taller fue tomada del Marco Geo estadístico Nacional - MGN del DANE. Puede consultar los datos [aquí](#).

#### 1.1. Importar datos espaciales

**1.1.1** De la carpeta *data/outpu* importe los shapefiles de VIAS (llame al objeto *via*) y MGN\_URB\_TOPONIMIA (llame al objeto *puntos*). El primero contiene las vías del departamento de Norte de Santander y el segundo la ubicación de algunos servicios públicos (hospitales, policía,...) y comerciales (restaurantes, tiendas,...).

**1.1.2** Cree un nuevo objeto llamado *c\_medico*, que contenga las observaciones del objeto *puntos* en las que la variable CSIMBOL sea igual a "021001", "021002" o "021003".

**1.1.3** De la carpeta *data/outpu* importe los *c poblado (2017).rds* (llame al objeto *c\_poblado*), *dp deptos (2017).rds* (llame al objeto *depto*) y *\textit{victimas\_map-muse.rds}* (llame al objeto *mapmuse*). Asegúrese de dejar únicamente los centros poblados con código DANE  $\geq 54001$  &  $< 55000$ . Además, del objeto *depto*, deje únicamente el polígono del departamento de Norte de Santander.

#### 1.2. Atributos de los objetos

Aplique la función *skim* de la librería *skimr* para explorar todos los objetos cargados en el punto anterior. Si considera necesario, seleccione algunas variables y pinte sobre la consola la tabla de frecuencia estas.

#### 1.3. Geometrías del objeto

**1.3.1** Para todos los objetos del punto 1.1., pinte sobre la consola la caja de coordenadas (*st\_bbox*) y el CRS de cada objeto.

**1.3.2** Ahora va a re proyectar el CRS de todos los objetos. Asigne la siguiente CRS **+proj=utm +zone=19 +datum=WGS84 +units=m +no\_defs** a todos los objetos del punto 1.1..

#### 1.4. Operaciones geométricas

**1.4.1** Use el objeto *depto* para hacer *clipping* y dejar los puntos de *mapmuse* que están debajo del polígono de Norte de Santander.

**1.4.2** Del objeto *c\_poblado*, seleccione cualquier municipio, use este polígono y el objeto *via*, para calcular el largo de las vías en el centro poblado que seleccionó.

#### 1.5. Pintar mapas

**1.5.1** Use la función *leaflet* para visualizar en un mismo mapa: los polígonos de los centros poblados, el polígono del departamento de Norte de Santander y los hospitales y puestos de salud del objeto *c\_medicos*.

**1.5.2** Use las librerías *ggplot*, *ggsp* y las demás que considere necesarias para visualizar en un mismo mapa: los polígonos de los centros poblados, el polígono del departamento de Norte de Santander y los hospitales y puestos de salud del objeto *c\_medicos*. Asegúrese de poner la barra de escalas, la estrella del norte y las

etiquetas que permitan diferencias cada uno de los objetos. Exporte el mapa en formato .pdf a la carpeta *views*.

## 2. Regresiones (45%)

La Oficina del Alto Comisionado para la Paz (OACP) tiene el registro oficial de víctimas por minas antipersona (MAP) y municiones sin explotar (MUSE). Estos registros pueden obtenerse en la página oficial de la [OACP](#). Para este ejercicio usted cuenta con una base de datos que contiene los registros de las víctimas de MAP-MUSE en el departamento de Norte de Santander.

**2.0.** En este punto usted debe emplear el objeto *mapmuse* y calcular la distancia más cercana de cada punto a una *via*, un centro medico (*c\_medico*) y a un centro poblado (*c\_poblado*). Asegúrese de nombrar las variables de distancia *dist\_vias*, *dist\_cmedico* y *dist\_cpoblado*, respectivamente. Construya la variable *fallecido* que tome el valor de 1 si el estado de la víctima es muerta y 0 si el estado es herida. Finalmente convierta el objeto en un *data.frame* solamente y elimine las variables *estado*, *month* y *geometry*.

**2.1.** Estime un modelo de probabilidad lineal en el que *fallecido* es la variable dependiente. Y use las demás variables como variables explicativas. Almacene los resultados de la estimación en un objeto llamado *ols*.

**2.2.** Exporte a la carpeta *views* los gráficos con los coeficientes (*coef-plot*) de las estimaciones.

**2.3.** Ahora estime la ecuación del punto 2.1. usando un modelo *logit* y un modelo *probit*, almacene los resultados de las estimaciones en dos objetos llamados *logit* y *probit* respectivamente.

**2.4.** Exporte los resultados de los tres modelos en una misma tabla usando la función *outreg*.

**2.5.** De los objetos *logit* y *probit* exporte a la carpeta *views* dos gráficos con el efecto marginal de la distancia a un centro medico sobre la probabilidad de fallecer.

## 3. Web-scraping (5%)

**3.1.** Desde la consola de Rstudio lea la siguiente url [https://es.wikipedia.org/wiki/Departamentos\\_de\\_Colombia](https://es.wikipedia.org/wiki/Departamentos_de_Colombia) y cree un objeto que contenga el HTML de la página como un objeto *xml\_document*.

**3.2.** Use el *xpath* para extraer el título de la página (Departamentos de Colombia).

**3.3.** Extraiga la tabla que contiene los departamentos de Colombia.