

RE4015: Tarea #1, Parte 2

Roberto Ponce
rpl@tec.mx

Escuela de Gobierno y Transformación Pública
Tecnológico de Monterrey
Entrega: 12 de febrero del 2021

Generales

Esta tarea es sobre SQL/Tidyverse y Regresión Lineal.



Info: Entregar la solución de la tarea en un archivo en formato de Word o PDF a través de Canvas. Corre el siguiente código en R para descargar los resultados del Censo 2020 de INEGI. En esta tarea trabajaremos con estos datos.

```
censo20.R

rm(list=ls())
# Fija tu directorio de trabajo
setwd("C:/clase/tareas/tarea1_2")

# Descarga el archivo del censo 2020 por AGEB para
# Nuevo Leon
download.file(url=
'https://www.inegi.org.mx/contenidos/programas/ccpv/
2020/microdatos/
ageb_manzana/RESAGEBURB_19_2020_csv.zip',
destfile='censo20_NL.zip', method='curl')

# Descomprime el archivo .zip
zipF<- "censo20_NL.zip"
outDir<- "./unzipfolder20"
unzip(zipF, exdir=outDir)

# Lee el archivo que descomprimes.
d20 <- read.csv(
paste(outDir, "/", "RESAGEBURB_19CSV20.csv", sep=""),
encoding="UTF-8", na.strings="*",
stringsAsFactors = FALSE)

head(d20)
```

El resultado del código anterior debe de ser un objeto en memoria llamado "d20" con la base de datos de manzanas y Áreas Geostatísticas Básicas (AGEBs) del Censo del 2020 de INEGI. Consulta el pdf adjunto en el Canvas de esta tarea que contiene el diccionario de datos de la base que acabas de descargar. Esta base de datos contiene los resultados por manzana, AGEB, localidad, municipio y estado para Nuevo León. Explora el contenido de la base y sus variables.

1 Tidyverse y SQL (25 puntos)

Utiliza Tidyverse o SQL en R para resolver los siguientes ejercicios. El objetivo es que generes una base de datos con los resultados de las manzanas y generes una serie de estadísticos.

- (a) Filtra tu base de datos de tal forma que solamente despliegue los resultados de las manzanas, elimina los totales por AGEb, localidad, municipio y estado. Quédate solamente con las manzanas.
- (b) Filtra tu base de datos del punto anterior, de tal forma que contenga solamente las manzanas de los municipios de la ZMM (once o dieciocho municipios, según la definición que utilices) y guárdalo como un nuevo objeto. En este objeto selecciona solamente las columnas siguientes: grado promedio de escolaridad, población total, viviendas particulares habitadas, viviendas con internet, promedio de ocupantes por cuarto, población sin afiliación a servicios de salud.
- (c) En la base de datos del inciso anterior, elimina las manzanas que contengan valores nulos en alguna de las variables que seleccionaste.
- (d) Con esta base de datos final, genera una tabla que despliegue el mínimo, máximo y promedio de grado promedio escolaridad por municipio a partir de las manzanas. Es decir, agrega las manzanas por municipio y genera el mínimo, máximo y promedio. Presenta tu tabla en los resultados, así como el código de R.

2 Regresión Lineal (25 puntos)

- (a) Con la base de datos del ejercicio anterior, calcula las siguientes dos nuevas variables: a) promedio de viviendas particulares habitadas con internet; b) porcentaje de la población sin afiliación a servicios de salud.
- (b) Estima una regresión lineal de grado promedio de escolaridad por manzana sobre porcentaje de viviendas con internet. Presenta el resultado e interpreta el coeficiente.
- (c) Dibuja un plot donde grafiques grado promedio de escolaridad contra porcentaje de viviendas, agregando al plot de puntos una línea en color rojo que represente tu línea de regresión. ¿Qué te dice este plot?
- (d) Estima una regresión lineal de grado promedio de escolaridad por manzana sobre porcentaje de viviendas con internet, promedio de ocupantes por cuarto y porcentaje de población sin afiliación a servicios de salud. Presenta los resultados e interpreta. ¿Cómo cambiaron los estimadores respecto al ejercicio anterior? ¿Por qué?
- (e) ¿Tu regresión tiene el sesgo de la variable omitida o no? ¿Por qué? ¿Son confiables tus estimadores? Explique.



Notice: La entrega de la tarea es individual. Si trabajas con alguien en la resolución de los ejercicios, especificar el nombre de las personas que colaboraron en la resolución de los ejercicios. Pega el código de R en tu documento.