

# Ejercicio 2

Cristhian Acosta - Miguel Garzón  
Taller de Stata

20 de agosto de 2020

## Contexto

- La idea de estos ejercicios es que usted practique los temas vistos en clase.

## Recomendaciones

- Es aconsejable revisar las notas de clase para esta sección del curso y los Do-File de cada clase para revisar la sintaxis de los comandos.

## Entrega

- miércoles 26 de agosto de 2020 hasta el final del día a través de *Sicua*, sección *Ejercicios*. Únicamente debe enviar el *do-file*, este debe poder ejecutarse ininterrumpidamente.

## Instrucciones

Usted desea revisar la base de datos de medición de calidad del aire en la estación de San Juan de Sao Paulo en Brasil, cuya unidad de observación se encuentra a nivel día-mes-año en el periodo 2016-2017.

1. Defina el entorno de trabajo, estableciendo el directorio
2. Importe la base de datos de la calidad del aire de la estación San Jose de Campos - Jardim Satellite ubicada en Sao Paulo (Brasil) en el archivo “1567349900409.csv”.
3. Luego, elimine las filas de la base de datos que no contienen información útil para el análisis. Para esto elimine las primeras 3 observaciones de la base.
4. Después, guarde la información del encabezado de la tabla (“Código da estação” y “Nome da estação”) en variables. Para esto cree una nueva variable llamada *cod\_estacion* que contenga el código de la estación (“Código da estação”) y una nueva variable llamada *estacion* que contenga el nombre de la estación (“Nome da estação”). Los valores de estas variables deben ser iguales para todas las observaciones de la base.
5. Ahora, renombre las variables a partir de la información del encabezado de la tabla. En particular, renombre las variables v1, v2, v3, v4 y v5 por fecha, hora, co, pm10 y pm2\_5, respectivamente.
6. Dado que ya empleamos la información del encabezado de la tabla importada, elimine las primeras 4 observaciones de la base de datos.
7. Reemplace en la variable hora “00:00” en lugar de “24:00”.
8. Concatene la hora a la variable fecha, separada por un “/”.
9. ¿Hay una hora del día donde se concentra los valores no reportados de CO en la base?.
10. Borre la variable hora.
11. Convierta a número las variables “co”, “pm10” y “pm2”\_5 ¿qué ocurre con la variable co?
12. Cambie el formato de la variable “co” por uno donde se pueda visualizar el cero antes del decimal.
13. Ordene las variables string en primer lugar en el siguiente orden: *cod\_estacion*, *estacion* y *fecha*.
14. Estime una tabla de estadísticas descriptivas en detalle de la variable de contaminación pm10.

15. Cree una nueva variable de “pm10”, reemplazando los valores de la variable “pm10” por debajo del percentil 1 y por encima del percentil 95 con sus valores críticos.
16. Estime una nueva tabla de estadísticas descriptivas básicas (número de observaciones, mínimo, mediana, máximo y desviación estándar en este mismo orden) de las variables “pm10” y “pm10” ajustada (creada en el numeral anterior). ¿cómo cambia la distribución de la variable?