

Ejercicio 3

Cristhian Acosta - Miguel Garzón
Taller de Stata

25 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

- viernes 28 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 seguir revisando 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 hora-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 “*read_sanjose.dta*”.
3. Modifique el nombre de la estación siguiendo los siguientes pasos:
 - a) Reemplace “S.” por “San”.
 - b) Reemplace “Jd.Satelite” por “jardin satelital”.
 - c) Elimine las tildes.
 - d) Deje la cadena de texto en mayúscula.
4. Cree las variables *año*, *mes* y *día* a partir de la variable *fecha*.
5. Convierta la variable *año* a número. ¿De qué años es la información de la estación? Conserve únicamente los años completos.
6. Cree una variable que identifique las observaciones con un 1 si son “missing” en todas las variables de contaminación.
7. ¿Cuántas observaciones cumplen esta condición? Elimine estas observaciones.
8. Cree una variable con el promedio y otra variable con la desviación estándar de la variable *co* a nivel de hora.
9. Estandarice la variable *co* con su media y desviación estándar a nivel de hora.
10. ¿Qué pasó con la organización de la base de datos? ¿Cuántos valores diferentes toman las nuevas variables de media y desviación en la base de datos?
11. Borre las variables media, desviación estándar y la creada en el numeral 6.
12. Etiquete las variables:
 - a) *co*: “Monóxido de Carbono (ppm)”
 - b) *pm10*: “MP10 (Partículas inhaladas)-microgramo por metro cúbico($\mu\text{g}/\text{m}^3$)”
 - c) *pm2_5*: “MP2.5 (Partículas inhaladas finas)-microgramo por metro cúbico($\mu\text{g}/\text{m}^3$)”
 - d) *pm10_w*: “MP10 winsorizada”
 - e) *z_co*: “Monóxido de Carbono (en desviación estándar)”

13. Cree una variable que contenga, para cada hora, un ranking de los días con mayores valores de la variable *co* estandarizada del numeral 9.
14. Identifique para cada hora, el día-mes de mayor nivel de la variable *co* estandarizada. Para esto cree una variable que contenga el día y el mes únicamente para la observación u observaciones (en caso de haber más de una observación en el primer lugar del ranking) de mayor valor de *co* estandarizada.
15. Por último, comprima la base de datos y guárdela en un archivo *.dta* llamado “*clean_sanjose.dta*”