

Taller de programación en R: Taller 2

Facultad de Economía, Universidad de los Andes

Febrero 26/2024

Instrucciones del trabajo:

El presente taller busca que se apliquen los conceptos vistos durante las clases Magistrales de las semanas 4-7 del curso. Adicional a esto, con los ejercicios propuestos se busca desarrollar una mejor comprensión del uso de R, de los paquetes vistos en clase y fomentar un razonamiento lógico que lleve a los estudiantes a comprender e implementar de una buena manera los tópicos tratados.

El Taller 2 debe ser diligenciado en parejas (sin excepciones) y entregado el viernes 15 de Marzo a las 11:59 pm a través de la plataforma de Bloque Neón. La copia o plagio en este trabajo está totalmente prohibida e incurrir en esta práctica conlleva a una nota de cero (0) en el taller, así como a las sanciones correspondientes tenidas en cuenta en el reglamento de estudiantes de la universidad. Igualmente, está **prohibido el uso de herramientas de IA para la generación de código** como Chatgpt, Bard, entre otras.

Cada pareja debe entregar:

- Un script de R donde presente el código que realizó para solucionar el mismo.
- En un archivo .pdf, docx, entre otros, un documento donde realice las interpretaciones necesarias para cada punto.

En caso de que el **punto no requiera algún desarrollo además del código**, pueden **explicar esto en el documento**. Finalmente, los scripts deben estar organizados por puntos y subpuntos.

Puntos Por Desarrollar:

Primer Punto

Ustedes cuentan con varias bases de datos que cuenta con información del precio del Petróleo Brent, del Gas Natural, Carbón, Gasolina Motor Corriente y finalmente del IPC para Colombia desde el 2000 hasta el 2024 con una periodicidad mensual. Cada una de estas bases de datos toma el nombre del bien al que hacen referencia junto con el sufijo .csv. Por ejemplo, cuentan con "Carbón.csv" y así con los demás bienes. Cada una de estas bases cuenta con tres columnas, una primera columna con la fecha mensualizada, una segunda columna con el precio del bien para Colombia y una tercera columna con el precio del bien promedio en el mundo. En su trabajo como analistas de información se les solicita realizar un reporte sobre estas series. Para ello, su jefa les solicita realizar las siguientes tareas:

- 1.1. Establezcan un código que les permita cargar en el programa R las bases de datos de los diferentes precios de combustibles.
- 1.2. Realicen una exploración inicial de la base de datos. Para esto, para cada una de las bases de datos describan la misma. Es decir, redacten un breve párrafo comentando qué variables contiene la base de datos, el número de observaciones, el periodo de tiempo que comprende, el tipo de las variables, entre otros aspectos relevantes.

Explorando las bases de datos de los precios de forma individual identifican que no todas tienen el mismo número de observaciones. Un hecho muy común cuando se trabaja con datos en formato panel es que algunas fechas para algunos productos no estén en la base de datos, es decir, hay fechas faltantes. A continuación, ustedes quieren "explicitar" las fechas faltantes.

- 1.3. Creen una función que tenga como parámetros un dataframe, una columna de fecha y una columna de precio, posteriormente, la función tiene que agregar las fechas faltantes en la columna de fecha y añadir un missing value en su valor correspondiente en el precio para esta fecha explicitada. Finalmente, tiene que retornar el dataframe con las fechas completas.
- 1.4. Apliquen la función anterior a cada uno de los precios en la base de datos.
- 1.5. Realicen una unión de cada una de los datasets cargados que les permita juntar los precios y fechas en una misma base de datos.
- 1.6. Realice las siguientes transformaciones de los datos: para la fecha transfórmenla de formato mm/dd/aa a un formato dd/mm/aa, revisen que las variables numéricas sean identificadas de tal forma y asegúrense de que tengan un formato de dos decimales.

- 1.7. Utilizando los operadores y funciones de la librería *dplyr*, creen respectivamente columnas que correspondan al mes y año de las observaciones. Filtren la base de datos para preservar únicamente las observaciones posteriores a enero del 2005 utilizando las columnas mes y año creadas anteriormente. Finalmente, eliminen de la base de datos las columnas asociadas a los precios de los combustibles promedio en el mundo.
- 1.8. Revisen ¿Qué porcentaje de la base de datos cuenta con valores faltantes en el precio de algún bien? Sustituyan estos valores faltantes con el precio promedio de ese bien.
- 1.9. Agrupen la información de los precios a los valores promedios por mes y año.

Por otra parte, para poder realizar comparaciones correctas entre precios es importante convertir los precios nominales en precios reales (constantes de un año base), pues permite controlar por el efecto inflacionario. Para transformarlo, se normaliza por la diferencia entre los Índices de Precios al Consumidor de la forma:

$$(1) \quad P_{real,t} = \frac{P_{nominal,t} * IPC_{base}}{IPC_t}$$

Donde IPC_{base} hace referencia al IPC de un año-mes específico. IPC_t al IPC del año-mes t sobre el cual se quiere convertir el precio nominal del año-mes t ($P_{nominal,t}$).

- 1.10. Creen una función que tenga como parámetros: una columna de una serie de un precio de la base de datos, un año, un mes y la columna del IPC de Colombia. Posteriormente, la función debe crear una nueva columna cuyo nombre sea: el “*nombre del bien*” + “*año base*” + “*mes base*” + el sufijo “_transformada”. Esta columna debe ser la transformación de valores nominales de la serie a valores reales con base en el año que toma como parámetro la función. Es decir, utilizando la ecuación (1) tendrán que consolidar una función que transforme los valores nominales en precios constantes con base en cualquier mes-año.
- 1.11. Apliquen la función que desarrolló en el literal anterior para todos los bienes que definió en la base de datos, pueden usar cualquier año-mes base, lo importante es que sean explícitos.
- 1.12. Exporte la base de datos consolidada en un archivo .csv

Segundo punto

2. A partir de la base de datos consolidada en el punto anterior, este punto busca explorar y analizar los datos consolidados. Para esto, realicen lo siguiente:
 - 2.1. Presenten una tabla de estadísticas descriptivas donde evidencien el número de observaciones, promedio, min, max, desviación estándar **de los precios nominales y reales de los combustibles**. Expliquen e interpreten en el documento los valores más importantes de la tabla.
 - 2.2. Presenten una gráfica de dispersión de puntos (scatter) donde en el eje X esté el precio del carbón y en el eje Y el precio de la gasolina, ambos reales sobre un año determinado, añadan una línea de ajuste lineal, título, labels y demás elementos que permitan una gráfica autocontenida. Expliquen e interpreten en el documento los elementos más importantes de la base de datos.
 - 2.3. Presenten una gráfica de serie de tiempo con el precio real del Carbón, Gas Natural, Petróleo y Gasolina Corriente para un mismo año, asegúrense de que tenga el título, labels, colores y demás elementos que permitan una gráfica autocontenida. Expliquen e interpreten en el documento los elementos más importantes de la base de datos.
 - 2.4. Realicen una visualización de tipo “heatmap” donde puedan identificar el cambio mensual del precio del Carbón, Petróleo y Gas Natural para cada uno de los años.