Taller de programación en R: Taller 1

Facultad de Economía, Universidad de los Andes

Agosto 16/2024

**Instrucciones del trabajo:**

El presente taller busca que se apliquen los conceptos vistos durante las clases Magistrales de las semanas 1-4 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 1 debe ser diligenciado en parejas (sin excepciones) y entregado al finalizar la semana 4 del curso (domingo) 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:**

1. **Primer Punto:**
   1. Definan una semilla para trabajar durante el script. Respondan: ¿Por qué es importante definir una semilla?
   2. Primero, creen los siguientes 4 vectores: uno que varíe de 1 en 1 desde el 1 hasta el 50. Este corresponde al identificador del individuo. Segundo, un vector de clase *int* llamado *edad* que se distribuya de forma uniforme entre el intervalo 5 a 50. Tercero, un vector que repita el carácter “años” y un vector de 50 nombres propios aleatorios de personas. Todas las cuatro (4) listas deben tener el mismo tamaño.
   3. Creen un vector en el que cada elemento j sea la concatenación de los elementos j de los vectores del punto anterior. Ordenen y/o agreguen caracteres a cada elemento de la lista para que se consolide una oración con orden semántico que refleje la edad del individuo. Para esto, utilicen la función *pasteo()*
   4. Usando un loop realicen un código que presente (print) la concatenación lógica de la edad, nombre y años de cada uno de los individuos dentro de las listas, pero únicamente si el nombre del individuo empieza por una letra distinta de J y la edad sea distinta de un número par. Es decir, el resultado debería ser algo similar a: “Camilo tiene 29 años”
   5. Programen una función que tome como entrada un vector con valores numéricos y que su output sea el **promedio** de los valores del vector y la **desviación estándar** asociada a la misma muestra. Usando esta función respondan: ¿Cuál es la edad promedio de su lista? ¿Cuál es la desviación estándar?
   6. Programen una función que tome como entrada un vector con valores numéricos y estandarice los valores. Es decir, que los transforme a una normal estándar. Apliquen las funciones que desarrollaron en el literal 1.5) **dentro de la función** que propongan en este literal.
   7. Apliquen la función programada en el literal 1.6) para crear un vector con la edad estandarizada. Llamen este nuevo vector como
   8. Por otra parte, generen una lista llamada outcomes\_nominales. Esta lista contendrá 3 vectores de 50 observaciones cada uno con los outcomes de interés: salario, índice de salud, experiencia laboral. Para esto, cada una de estas variables tiene que seguir el siguiente proceso generador de datos.

Para todos los procesos corresponde a un error proveniente de una distribución normal con media 0 y varianza 1.

* 1. Creen una función que permita convertir un vector en una matriz para la estimación de una regresión lineal simple. Para esto, la función debe tomar como input un vector y debe tener como output una matriz **X** que concatene los datos de este vector y un vector de 1´s.
  2. A partir de la función anterior consoliden una matriz **X** con la **edad** de los individuos **estandarizada** y un vector de 1´s asociado a una constante.

1. **Segundo Punto:**
   1. Programen una función que tome como input una matriz **X** y un vector, posteriormente, el output debe corresponder a una estimación puntual del estimador () de Mínimos Cuadrados Ordinarios (MCO) para la muestra y a su error estándar asociado ().
   2. Utilizando un loop, apliquen esta función a los diferentes outcomes en las listas de outcomes\_nominales, guarden los coeficientes estimados y los errores estándar en una matriz donde la primera columna corresponde al nombre del outcome, la segunda columna al coeficiente estimado para la constante (). La tercera debe tener el coeficiente estimado asociado a correlación con la edad () y la cuarta al error estándar (). En esta matriz, cada fila representará una estimación.
   3. Interpreten los coeficientes, en particular, comparen sus resultados con los procesos generadores de datos del punto 1.8) ¿Qué patrones encuentran?