

EVALUACIÓN MÓDULO INTRODUCCIÓN A LA PROGRAMACIÓN EN RSTUDIO.

Prof. Adrian Armando Araneda Toro.

Instrucciones

1. La Evaluación es Individual.
2. Destine un archivo (.docx) para responder a las preguntas que se pidan ser desarrolladas y/o justificación (texto). O, dicho de otra manera, escribir en un ordenador, en una hoja .docx (Word), el desarrollo textual de la prueba, para el caso de las preguntas que así lo requieran.

No se aceptará como respondida aquellas preguntas que el profesor reciba en formato manuscrito, a mano, escaneado o fotografiado.

3. Responda en un Script R (.R) para el caso de las preguntas que así lo requieran.
4. El profesor debe ejecutar exitosamente el Script. Si no se ejecuta con éxito la pregunta quedará como no respondida. Sin asignación de puntos o décimas.

Asuma que el profesor sólo hará “Ctrl+Enter” por cada línea de código. No corregirá código para que sea ejecutado exitosamente.

5. No asuma que el profesor posee los paquetes y librerías ya instalados(as) y ejecutados(as), debe enunciarlos en el script para habilitar por primera vez la función que desea utilizar.
6. Favor, colocar en la zona superior de su documento .docx y .R, su Nombre, Apellido y RUN.
7. En consecuencia, los 4 outputs o productos que se deberá entregar por alumno son:
 - I. Una hoja escrita en un ordenador (.docx),
 - II. El Script en formato (.R), que se ejecute exitosamente.
 - III. El set de datos con el que trabajó.
 - IV. Otros outputs en el formato que las preguntas lo requieran.

8. Todo lo solicitado debe ser enviado a través de un solo correo electrónico (no dos, tres, etc.). En el caso que el alumno envíe mas de un correo electrónico dentro del horario consignado para el desarrollo y entrega de la evaluación con elementos adjuntos repetidos en un correo anterior, el profesor seleccionará aquellos elementos del último correo electrónico recibido dentro del horario de entrega.
9. En dicho correo electrónico se debe adjuntar todo lo solicitado, o adjuntar un archivo comprimido con todos los elementos.
10. Enviar sólo por U-Cursos hasta las 20 hrs. No se recibirán evaluación ni elementos desde las 20.01 hrs.
11. No se aceptarán enmendaciones de las entregas enviadas fuera de la hora indicada.
12. Esta evaluación consta de Punto Base.
13. Se aplicaran las décimas ganadas y descontadas, de la tarea encomendada.

Observación: Revise que **TODOS** sus elementos se hayan adjuntado al enviar la evaluación. Adjunte y envíe 15 minutos antes de la hora de cierre indicada.

PREGUNTA 1.

Del material docente remitido y/o de lo visto en clases, indique no en más de 5 líneas, las 3 principales ventajas al utilizar un motor de procesamiento de grandes volúmenes de datos.

(0.5 pts.)

PREGUNTA 2.

De lo visto en clases explique la diferencia (excluyente) entre un paquete y una librería en RStudio.

(0.5 pts.)

PREGUNTA 3.

Para las siguientes letras, considere que x es un set de datos estructurado de tipo “dataframe”. Seleccione si las siguientes afirmaciones son Verdaderas o Falsas. Para el caso de las Falsas justifique en 1 línea.

a.) Seleccione si la siguiente sentencia es Verdadera o Falsa:

$$x[,] = x$$

Verdadera

Falsa

(0.5 pts.)

b.) De la siguiente expresión:

```
x <- cbind(x, y = rep(78,79))
```

Se puede afirmar que el dataset posee 78 rows.

Verdadero

Falso

(0.5 pts.)

c.) De la expresión anterior se puede afirmar que obtendré un vector de tipo “numeric” denominado rep.

Verdadero

Falso

(0.5 pts.)

d.) Asumiendo que x e y son dataframes, la condición **all.x = TRUE** en el uso de **merge** trae sólo la intersección de x e y.

Verdadero

Falso

(0.5 pts.)

PREGUNTA 4.

Instrucciones:

1. Responda cuidadosamente cada Letra.
2. El Script completo o cada línea de código no debe tener “x en rojo” o errores identificados por RStudio.
3. El Script debe ser auditable por el profesor. Es decir, debe ejecutarse exitosamente por completo, línea por línea. Y en el orden generado por el alumno. Sin errores. Con el fin que sea reproducible. En el caso que la línea de código no se ejecute exitosamente se considerará 0 en la asignación del puntaje correspondiente.
4. **Considere qué: Si la línea de código n+1 depende de la línea de código anterior n y esta no se ejecuta exitosamente, también se asignará 0 a la línea de código n+1 ya que no podrá ser ejecutada. Y así sucesivamente para aquellas salidas que dependan de líneas o sintaxis anteriores.**
5. Todo debe ejecutarse exitosamente. Si el profesor obtiene error para alguna ejecución, se entenderá como pregunta no respondida.
6. Escriba su nombre, RUN y DV en el encabezado de este documento y del Script.

Preguntas a Responder:

- a. Seleccione y Cargue una base de datos con la cuál usted trabaje o ejecute durante las tareas o rutinas laborales diarias en su trabajo o con su equipo de trabajo (si usted no trabajara con ninguna directamente, seleccione alguna que su equipo de trabajo sí haga uso de ella). En el caso que no desee utilizar un producto real, utilice el set de datos remitido (adjunto) por el profesor denominado USArrests.

Para lo anterior y de lo visto en clases, la base de datos con la cuál usted deberá cargar debe tener las siguientes 4 características:

- a.1. Debe poseer a lo menos 4 vectores de tipo num.
- a.2. Debe poseer a lo menos una dimensión de tipo no num.
- a.3. Debe poseer a lo menos 50 rows.
- a.4. Los 4 primeros vectores num, deben poseer 6, 5, 4 y 3 N/A por vector respectivamente (observaciones ausentes).

(0.3 pts.)

- b. En una sola línea de código, entregue un resumen de la cantidad de N/A y su porcentaje, por atributos. **(0.2 pts.)**
- c. A partir del set de datos cargado genere un nuevo set de datos con todos los N/A reemplazados por el quinto número de su RUN (leyéndolo de izquierda a derecha. Si este es cero, pase al siguiente dígito en el mismo orden). **(0.2 pts.)**
- d. Al set de datos anterior genere una variable denominada “nueva1”, la cuál debe ser el resultado de la suma de cada observación del vector 2 con el tercer número de su RUN (leyéndolo de izquierda a derecha. Si este es cero, pase al siguiente dígito en el mismo orden). **(0.2 pts.)**
- e. En el set de datos obtenido anteriormente genere una variable denominada “nueva2” la cuál debe ser el resultado de la multiplicación de cada observación del vector 3 con el cuarto número de su RUN (leyéndolo de izquierda a derecha. Si este es cero, pase al siguiente dígito en el mismo orden). **(0.2 pts.)**

- f. En el set de datos anterior, entregue en un vector nuevo, resultante de la división de la dimensión 1 con la 3. **(0.1 pts.)**
- g. Para el dataset cargado, en una línea de código, ejecute la función que sólo arroja como outputs la cantidad de rows y la cantidad de vectores **(0.2 pts.)**
- h. Exporte en dos sheet distintos, en un mismo archivo.xlsx., el dataframe original cargado y el intervenido por usted **(0.2 pts.)**
- i. Exporte en un archivo.xlsx. el archivo intervenido **(0.2 pts.)**
- j. En una sola línea de código o sintaxis muestre o imprima la observación correspondiente a la posición matricial del dataframe nuevo, “rows 17 con vector 4” **(0.2 pts.)**
- k. En una sola línea de código o sintaxis, en el set de datos nuevo, elimine la observación cuya coordenada matricial sea rows 13 y c 4. Y expórtelo como archivo o elemento rds. **(0.2 pts.)**
- l. Exporte todo a un RData. **(0.2 pts.)**
- m. Borre todo el Environment con una sola función en una línea de código. **(0.1 pts.)**
- n. Traiga el vector 2 de un registro aleatorio de la data set intervenido. **(0.1 pts.)**
- o. En una línea de código renombre todas las Variables del set de datos intervenido con nombres nuevos que usted desee. Luego, renombre las variables del set de datos cargados con los nombres que usted desee, manteniendo el vector de characters o string de ambos set de datos con el mismo nombre. Finalmente, traiga al set de datos cargado la información del set de datos nuevo **(0.4 pts.)**
- p. Para un dataset cualquiera que usted haya cargado o procesado en esta prueba, genere en él una nueva variable con información que provenga de otro vector del mismo dataset. Ocupe una **condición** para dicha generación. Todo en una sola línea de código. **(0.2 pts.)**

BONUS TRACK (se regalan 2 décimas).

Ejecute y responda al Script, siguiendo las instrucciones del mismo, que se adjunta denominado “BONUS TRACK”. Específicamente, a la pregunta:

Explique con sus palabras o a modo cualitativo, luego de seguir sus instrucciones y ejecutarlo;

“la diferencia de lo obtenido en la línea de código 30 versus lo obtenido en la línea de código 40.”