

‘Prueba de Evaluación R’

Autor: _____

Actualizado septiembre 2020

Pregunta 1 (20 puntos): Un boleto del sorteo de la ONCE consta de dos partes, la primera es un número de 4 dígitos y la segunda es un número de tres dígitos que forman la serie del boleto.

Aquí consideramos sólo el número, por ejemplo,

0	2	0	9
---	---	---	---

Se pide:

- Genera todos los números que entran en el sorteo de la ONCE y mostrarlos con los cuatro dígitos.
- ¿Cuál es la suma de los números de un boleto que más se repite?

Pregunta 2 (20 puntos): En la carpeta covid_19 hay una serie de archivos sobre el covid-19 en España

(Fuente:

<https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/situacionActual.htm>). Se pide:

- Leer los archivos “datos_provincias.csv”, “CodProv.txt*” y “CodCCAA.dat “. Añade el código de la comunidad autónoma al fichero “datos_provincias.csv” (no manualmente).
- Selecciona los datos de la comunidad autónoma que te corresponda. Para saber cuál es tu comunidad autónoma realiza la siguiente operación

DNI o Pasaporte mod 17 por ejemplo (12345678 %% 17 = 6 → Castilla y León)
hay que seleccionar las provincias de Castilla y León

Gráficos

- Realizar un gráfico que muestre adecuadamente la evolución de los casos nuevos. Justifica el gráfico elegido.
- Presenta en único gráfico la evolución de las distintas variables (columnas) por medio de un gráfico de líneas múltiples. Utiliza diferentes colores y añade una leyenda muestre el origen de cada línea.

Pregunta 3 (20 puntos): Consideramos un fichero de datos en formato SAS de nombre “punt.sas7bdat” que contiene datos sobre alumnos matriculados en diversos cursos. Las variables son:

- SEGSOC (nº de la seguridad social)
- ENROLLED (fecha de matrícula)
- COURSE (curso)
- TEACHER (profesor)
- TEST1 puntuaciones de los estudiantes
- TEST2 puntuaciones de los estudiantes
- TEST3 puntuaciones de los estudiantes
- TEST4 puntuaciones de los estudiantes

##	SEGSOC	ENROLLED	COURSE	TEACHER	TEST1	TEST2	TEST3	TEST4
##	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	113-31-5	12-01	EC0560	GOMEZ	81	69	81	81

##	2	113-74-5	07-20	ECO560	GOMEZ	92	92	61	81
##	3	445-33-6	11-29	MATE500	GONZALEZ	78	87	92	91
##	4	526-86-6	02-06	ECO560	GOMEZ	81	69	75	95
##	5	546-14-8	07-25	MATE500	GONZALEZ	87	91	100	95
##	6	628-35-8	02-26	HIST200	LOPEZ	87	92	91	95

- Importa el fichero de datos y guárdalo en un objeto de nombre `punt`. Comprueba la estructura del objeto `punt`. Si es necesario conviértelo en un data frame.
- Obtener una nueva variable `overall` que de la puntuación media de los cuatro test para cada estudiante suponiendo que el último test se pondera el doble.
- Formar una nueva variable denominada `start` compuesta por el mes y día de `ENROLLED` y por el año corriente y presenta en pantalla las variables `SEGSOC`, `COURSE` y `star`.
- Formar un nuevo data frame de nombre `level500` que contenga los estudiantes cuyo curso acaba en 500. Crear dos nuevas variables carácter, una de nombre `subject` con el código de curso (parte literal) y otra de nombre `level` con el número del curso (parte numérica).
- Escribe la información de `level500` en fichero ASCII de nombre “level500.dat”.

Pregunta 4 (40 puntos): La siguiente tabla representa puntuaciones de sensación de ardor para 16 sujetos en un estudio para probar un nuevo hidrogel. La primera columna da el número del sujeto. Las siguientes columnas dan la puntuación de sensación de ardor (en una escala de 1 a 4) para semanas 1 (S1) a 7 (S7). (La matriz de datos se encuentra en “matriz.R”)

Nr	S1	S2	S3	S4	S5	S6	S7
1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	2
3	1	1	1	1	1	2	3
4	1	1	1	1	1	3	4
5	1	1	1	1	2	3	3
6	1	1	1	1	1	1	1
7	1	1	1	3	4	2	2
8	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1
10	1	1	1	1	1	1	4
11	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1
13	1	2	1	3	2	3	4
14	1	1	1	2	2	4	4
15	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1

Propondremos una forma interesante de mostrar estos datos.

1.- Para la semana S_7 , calcule el vector $(f_1, 1 - f_1, f_2, 1 - f_2, f_3, 1 - f_3, f_4, 1 - f_4)$ donde f_i es la frecuencia de la modalidad $i \in \{1, 2, 3, 4\}$ observada en la semana S_7 sobre los 16 sujetos. (Sugerencia: use las funciones `tabulate()`, `cbind()`, `t()` y `as.vector()`)

2.- Ahora, use la función `apply()` para hacer el mismo cálculo para todas las demás semanas. Almacene el resultado en una matriz.

3.- Utilice la función `barplot()` y el argumento `col = c ("black", "white")` en esta matriz. El gráfico que se obtiene ofrece una descripción general de la evolución de la Sensación de ardor con el tiempo.

4.- Cambie el gráfico anterior para que las barras que representan las frecuencias estén en rojo. Los números de las semanas deben estar en azul y en la parte superior del gráfico en lugar del fondo. Los números de modalidad deben estar a la izquierda, en azul. Agrega un título al gráfico

Instrucciones generales:

1. Es obligatorio enviar el código R que utilices en un fichero ASCII (texto plano).
2. Si el ejercicio lo requiere, se adjuntará un documento en pdf que explique lo que se ha hecho y lo que se ha obtenido.
3. Se puede sustituir los documentos (1) y (2) por un único documento en markdown.
4. Se recomienda utilizar la codificación UTF-8.
5. Si se utiliza alguna librería debe decirse cual es y las ventajas respecto al código base.
6. Siempre que se pueda utilizar las funciones de R Base.
7. Si la función que necesita no existe se recomienda crear funciones para responder a las preguntas (si es posible y necesario).
8. Las funciones deben ser los más generales posibles.