

Prácticas Gestión Datos R

Daniel Orellana Pérez

#Ejercicio 1.- Supongamos que tienes un vector llamado 'edades_descubrimientos' que contiene las edades en años de diferentes descubrimientos arqueológicos importantes. ¿Cuál es la edad media de estos descubrimientos?

En este ejercicio lo que hemos hecho es decirle a R a través de un vector las edades de los descubrimientos que hemos utilizado, en este caso: 2001,2002,1995,1860 y 1999. Posteriormente hemos sacado la media con la función “mean” diciéndole a R que nos la saque de las edades que previamente le hemos dicho.

#Ejercicio2.- Si tienes un vector llamado 'cantidad_artefactos' que indica la cantidad de artefactos encontrados en cada sitio arqueológico, ¿cuántos artefactos en total se han encontrado en todos los sitios?

En este ejercicio lo que hemos dicho a R que, a través de un vector llamado “cantidad_artefactos” le digamos cuantos artefactos hay. Para ello hemos utilizado la función “sum” para que nos sume automáticamente los valores que le hemos dado (artefactos) previamente.

#Ejercicio 3.- Tienes un vector llamado 'profundidad_hallazgos' que indica la profundidad en metros de diferentes descubrimientos. ¿Cuál es la profundidad máxima de los hallazgos?

En esta ocasión, el ejercicio nos está pidiendo que, dentro de los valores del vector “profundidad_hallazgos”, le indiquemos a R que nos siga cual es el de más valor (en este caso, el que más profundidad tiene). Para ello se han generado una serie de números aleatorios gracias a la semilla que pusimos al principio de los ejercicios y le indicamos a R que ejecute la función “max”.

Respecto a los números aleatorios, se ha utilizado la función “sample” y le decimos además que me genere los números del 1 al 150 (1:150).

#Ejercicio 4.- Supongamos que tienes un vector llamado ‘materiales_encontrados’ que enumera los tipos de materiales encontrados en diferentes yacimientos. ¿Cuántos tipos de materiales distintos se han encontrado en total?

En este ejercicio hemos creado un vector denominado “materiales.encontrados” y le hemos indicado cuales son los valores de este vector (óseo, mineral, cerámica y metal). Para poder decir cuántos tipos de material se han encontrado en total, hemos utilizado la función “length”, dando como resultado en este caso 4.

#Ejercicio 5.- Si tienes un vector llamado 'años_excavaciones' que contiene los años en los que se realizaron excavaciones en diferentes sitios, ¿en cuántos años se han llevado a cabo excavaciones?

Para la realización de este ejercicio hemos creado un vector llamado “años_excavaciones” cuyos valores son los años que nosotros hemos seleccionado

(2001,2002,1995,1860 y 1999). Posteriormente hemos utilizado la función “*length*” y “*unique*”. Esta última es solo para indicar que sólo queremos que coja estos valores y no unos aleatorios. Posteriormente el resultado que nos da es 5 puesto que son 5 los valores que hay en el vector.

#Ejercicio 6.- Supongamos que tienes una matriz donde las filas representan los años y las columnas representan diferentes sitios arqueológicos, y los valores indican la cantidad de excavaciones realizadas en cada año y sitio. ¿En qué año se realizaron más excavaciones?

En este ejercicio, a diferencia de los otros se nos ha pedido generar una matriz en filas, cuyos valores son 2001,2002,1995 (años de las excavaciones) y Nertóbriga, Valencina e Itálica. Posteriormente le hemos indicado la función “*print*” para que en la consola nos salga la siguiente tabla:

```
> print(matriz)
      Nertobriga Valencina Italica
2001           1         4       7
2002           2         5       8
1995           3         6       9
```

Posteriormente usamos la función “*sumafilas*” para que nos sume los valores de las filas de la matriz, pero sin que deje de estar relacionado con su año. Una vez que le damos a ese resultado, llamamos la función “*print(sumafilas)*” para que nos diga el resultado, en este caso, sería el siguiente:

```
> print(sumafilas)
2001 2002 1995
  12   15   18
```

#Ejercicio 7.- Tienes una matriz donde las filas representan diferentes descubrimientos y las columnas representan las regiones geográficas donde se encontraron. Los valores son las edades de los descubrimientos. ¿Cuál es la región con la edad promedio más antigua de los descubrimientos?

Creamos nuestra matriz con diferentes hallazgos (hachas collar y lucerna) y sus correspondientes sitios geográficos (Nertóbriga, Valencina e Itálica), siendo los valores (años) los números generados de forma aleatoria con el mismo procedimiento que en el ejercicio 3. Posteriormente tras usar la función “*print*” y “*sumafilas*” hacemos que sumen las filas y que me diga cual es la edad promedio más antigua gracias a la función “*mean*”.

#Ejercicio 8.- Si tienes una matriz donde las filas representan diferentes periodos de tiempo y las columnas representan diferentes tipos de artefactos, y los valores son la cantidad de cada tipo de artefacto encontrada en cada período, ¿cuál es el período con la mayor cantidad total de artefactos?

En esta ocasión debemos hacer una matriz teniendo en cuenta que, cada fila es un periodo de tiempo diferente y las columnas que tipo de resto es, siendo los valores el nº de cada artefacto en cada época. Sumamos por columnas y, con los resultados, utilizamos la función “*max(sumafilas)*” para que nos diga del resultado de la operación anterior, cual

es la edad que más artefactos, siendo en este caso 18 que corresponde con la Época Moderna.

#Ejercicio 9: Supongamos que tienes una matriz donde las filas representan diferentes sitios arqueológicos y las columnas representan diferentes excavaciones en esos sitios. Los valores indican la profundidad de los hallazgos en cada excavación. ¿Cuál es el sitio con la menor profundidad promedio de los hallazgos?

A diferencia del caso anterior, los pasos serían exactamente iguales, pero al final, como nos está pidiendo que se indique dónde está la menor profundidad, tras sumar las filas se realiza utilizando la función “*min(sumafilas)*” para que nos indique el resultado de menor valor.

#Ejercicio 10.- Tienes una matriz donde las filas representan diferentes períodos de tiempo y las columnas representan diferentes tipos de materiales. Los valores indican la cantidad de cada tipo de material encontrada en cada período. ¿Cuál es el tipo de material más común a lo largo de todos los períodos?

En este ejercicio lo que hay que tener en cuenta es que, tras crear la matriz, hay que usar la función “*sumacolumnas*” (y no “*sumafilas*”) para que nos diga cual es el material más común en cada una de las edades.

#Ejercicio 11- Crea un data frame llamado 'registro_artefactos' que contenga información sobre artefactos encontrados, incluyendo el sitio arqueológico, el tipo de artefacto, la fecha de descubrimiento y una descripción.

El objetivo de este ejercicio es conseguir poner de forma ordenada un yacimiento, tipo de artefacto, fecha de descubrimiento y descripción asignado a cada valor. Los valores de yacimiento, tipo de artefacto y descripción lo hemos seleccionado nosotros, pero para la fecha de descubrimiento lo hemos generado aleatoriamente con números que oscilan desde el 1 hasta el 150.

#Ejercicio 12.- Crea un data frame llamado 'excavaciones_equipo' que contenga información sobre las excavaciones realizadas por diferentes equipos arqueológicos, incluyendo el equipo, el sitio arqueológico, la fecha de inicio y la fecha de finalización.

Hay que crear un data frame generado fechas de inicio y de fin de excavación de forma aleatorias. Pero hay que tener en cuenta no pueden empezar desde el mismo punto puesto que, si no saldrían fecha de fin superiores a la de inicio. Por ello ponemos que se genere el número a partir de la finalización del primero, es decir si ponemos 1998:2010, la finalización debe ser 2010:2023.

#Ejercicio 13.- Crea un data frame llamado 'datos_esqueletos' que contenga información sobre esqueletos descubiertos en diferentes yacimientos, incluyendo el sitio arqueológico, la edad estimada, el sexo y cualquier característica especial.

En este ejercicio, una vez creado el data frame he puesto los valores que queremos, como son los datos sobre su sitio arqueológico, su edad, sexo y características especiales. En este caso, se le ha dado libertad a R para que le asigne estos valores hasta llegar a diez, puesto que en al principio del ejercicio, cuando creamos la semilla, le dijimos que creara 10 registros. La edad ha sido acotada de 1 año a 40 (1:40).

#Ejercicio 14.- Crea un data frame llamado 'ubicaciones_geograficas' que contenga información sobre la ubicación geográfica de diferentes yacimientos, incluyendo el sitio arqueológico, la latitud, la longitud y la altitud.

Para finalizar, en este ejercicio hemos creado el data.frame donde hemos puesto que los valores seas en sitio arqueológico, la longitud, latitud y altitud. Para el sitio arqueológico si que hemos señalado nosotros los tres yacimientos, pero para los otros valores, los ha elegido R de forma aleatoria, pero de forma acotada: Longitud (0:180), Latitud (0:90) y Altitud (1:3000).