

Ejercicios conceptos básicos:

En este curso de estadística aplicada vamos a utilizar R (Rstudio) como herramienta de trabajo. Si no tienes este programa instalado en tu ordenador de trabajo, puedes conseguirlo siguiendo estos pasos:

1. Descarga e instala la última versión de R: <https://cran.rstudio.com/>
2. Descarga e instala la última versión de RStudio:
<https://www.rstudio.com/products/rstudio/download/#download>

Un pequeño *recordatorio* de Rstudio: Rstudio presenta cuatro pantallas, el *script*, la *consola*, el *environment*, y el panel de *ayuda/gráficos*. Únicamente aquello que sea escrito en el *script* se puede guardar como documento de texto directamente. En la *consola*, veremos aquellas líneas de código que hayamos corrido (run o Ctrl+Enter), junto con su resultado.

¡A tope!

Estos ejercicios nos van a ayudar a afianzar los conceptos tratados en la parte de teoría de hoy. En primer lugar, para mantener limpio nuestra sesión de R, vamos a borrar cualquier variable que hayamos utilizado previamente pero que siga en el *environment*. ¿Cómo? Ay dios, yo tampoco me acuerdo... Vamos a buscar en Google *cómo eliminar variables de environment en R*.

```
rm(list=ls())
```

¡Genial!

Para llevar a cabo los siguientes ejercicios vamos a necesitar algunos datos, así que utilizaremos algunas de las bases de datos (bd) que R tiene incorporadas. Para ello necesitamos instalar y cargar el paquete **datasets**. Una vez que este paquete esté instalado y cargado, puedes incluir en tu environment una de las bd que incluye. Aquí te dejo el nombre de algunos de estos datos que me parecen interesantes:

- ChickWeight: BD del peso de pollos de distintas edades y con distintas dietas.
- airquality: BD de medidas de la calidad del aire de Nueva York.
- CO2: BD de CO2 fijado por especies de plantas.
- trees: BD de dimensiones de cerezos.
- ToothGrowth: BD del tamaño de los dientes de cobayas con distintos suplementos de vitamina C.
- Iris: BD de dimensiones de las flores de distintas especies de Iris.

¡Hay muchas más bd en este paquete! Si estas no cumplen tus expectativas o no te resultan curiosas, puedes elegir cualquier otra

Ejercicios:

1. Carga la BD que hayas elegido.

```
data(NombreDeBaseDeDatosElegida)
```

Al utilizar la función `data()`, la bd elegida se carga en el environment, siendo un nuevo objeto que se llama exactamente igual que la bd. ¿Le cambiamos el nombre? Prueba a crear un nuevo objeto cuyo nombre te resulte más cómodo para el resto de ejercicios.

2. Échale un vistazo a la base de datos para comprender la estructura que tiene, las variables que comprende y cómo se llaman. Prueba las funciones `>View()`, `>head()`, o `>str()`. ¿Qué te está mostrando cada una?
3. Escoge una variable de tu bd que te parezca interesante y calcula la media aritmética. ¿Habrá alguna función que la calcule directamente?

Recuerda que para seleccionar una sola variable de una base de datos, utilizamos `$`. Por ejemplo, si tenemos una base de datos llamada *casa* con 3 variables: *cuarto*, *usuario* y *dimensiones*, y queremos mostrar los valores de la variable *usuario*, utilizaremos: `casa$usuario`

4. Con la misma variable que has elegido en el paso anterior, calcula tres medidas de dispersión.
5. Grafica un histograma de la variable elegida y comprueba que los valores obtenidos en los ejercicios 3. y 4. tienen sentido. Extra: utiliza la función `>summary()`, la cual te proporciona un resumen de la variable elegida, y comprueba el resto de valores que muestra el histograma. Extra 2: Añade líneas verticales en el histograma en los valores del cuartil 1 y 3 de la variable elegida.
6. Volviendo a la base de datos, vuelve a pensar en todas las variables que incluye. ¿Qué pregunta podrías contestar con ellas? Escribe las hipótesis nula y alternativa para esa pregunta.

Recuerda que para añadir comentarios en el script de R, puedes utilizar la almohadilla/hastag (`#`); si no, R lo interpretará como código.

7. Extra: Cuando acabamos de recoger nuestros datos, es muy interesante (y satisfactorio) verlos graficados de alguna manera, intentando testar visualmente nuestras hipótesis. Crea un gráfico que creas que puede resolver tu hipótesis planteada en el ejercicio 6. (Pista: si las variables son continuas, utiliza la función `>plot()`; si alguna de las variables es categórica, utiliza la función `>boxplot()`).

Recuerda que para representar una ecuación o la relación entre variables dependientes e independientes es necesario el símbolo `~`. Otra alternativa es seguir los argumentos de la función a utilizar: generalmente, primero va el valor de la variable independiente (x), seguida de la var. dependiente (y).