

El principal objetivo de esta guía es seguir programando. Para eso, en el marco de esta materia, hemos elegido ahora una excusa con datos. Es una especie de ejemplo juguete, pero poderoso (el chiquitín)!

En esta [página](#) están los datos podrás acceder a los datos que vas a utilizar para resolver esta guía. La idea es que cada uno trabaje con sus propios datos y por eso les pedimos que ingresen el número de libreta (o 5 últimos del DNI) para suministrar datos de forma personalizada. A medida que aumentan el tamaño  $n$  del conjunto de datos, incluyen nuevos casos; se agregan filas.

## 1. Calentando motores

1. Descargar de esta [página](#) un conjunto de  $n = 500$  observaciones, con todas las variables y leer el archivo en R.
2. Identificar el nombre de las columnas del `data.frame`.
3. Calcular el promedio de las alturas de los hijos para predecir la altura de un nuevo individuo.
4. Realizar un histograma de las alturas de los hijos. ¿Cuántas modas se observan? ¿A qué se puede atribuir?
5. Explor el comando `plot(density(variable))`, utilizando como variable la columna de alturas con las que se realizó el histograma. ¿Qué está pasando? ¿Cuántas modas observa? ¿A qué se puede atribuir?
6. Realizar ahora un histograma de alturas por cada género. Es decir, un histograma para las alturas correspondientes al género Masculino y otro para las alturas correspondientes al género Femenino.
7. Superponer en cada histograma del ítem anterior una función de densidad normal con los parámetros que consideres pertinentes.
8. Indicar con qué valor se puede predecir la altura de un hijo Masculino, es decir, la altura de un hijo. Indicar con qué valor se puede predecir la altura de un hijo Femenino, es decir, la altura de una hija.
9. Indicar ahora con qué valor se puede predecir la altura de un hijo (Masculino), sabiendo además que su mamá es de contextura pequeña. Comparar el valor de la predicción con la predicción obtenida en el ítem anterior. ¿Qué se puede observar?

## 2. Vamos ahora a considerar la altura de la mamá.

10. Graficar altura de mamá (en el eje x) vs. altura del hijo (eje y), utilizando un color por cada género. ¿Qué se puede observar?

En adelante, trabajaremos apenas con los datos de los hijos (género masculino).

11. Indicar si hay alguna madre de hijo varón cuya altura sea 156 cm.
12. Vamos ahora a predecir la altura de un hijo correspondiente a una mamá que mide  $x = 156$  cm haciendo *promedio móvil* centrado en 156 con ventana de tamaño  $h = 0.5$  (cm).
  - a) Indicar cuántos casos hay donde la madre registra una altura entre 155.5 y 156.5 cm., inclusive.
  - b) Calcular el promedio de la altura de los hijos cuyas madres registran una altura entre 155.5 y 156.5 cm.
13. Promedio móvil centrado en 156 con ventana de tamaño  $h = 1$  (cm).
  - a) Indicar cuántos casos hay donde la madre registre una altura entre 155 y 157 cm., inclusive.
  - b) Calcular el promedio de la altura de los hijos cuyas madres registran una altura entre 155 y 157 cm.
14. Repetir los items anteriores pero considerando ahora que la altura de la mamá es 159 cm. Es decir, nos movemos a calcular promedios en otro lado. Por eso hablamos de promedios móviles.

## 3. Implementando funciones

Hasta ahora hemos trabajado con dos posibles valores para la altura de la madre (156 y 159) y dos posibles valores de ventana ( $h = 0.5$  y  $h = 1$ ). Vamos ahora a implementar una función que permita predecir la altura de un hijo en función de la altura de la madre y el tamaño  $h$  de ventana elegida para hacer el promedio móvil.

15. Implemente una función que permita predecir la altura de un hijo en función de la altura de la madre, los datos para la altura de los individuos y de las madres y el tamaño  $h$  de ventana elegida para hacer el promedio móvil. Es decir, defina la función `predigo_altura_masculino(altura, altura_madre, altura_mama_nueva, h)`
16. Graficar la función `predigo_altura_masculino`, con  $h = 0.5$ , evaluandola a lo largo de una una grilla sobre un intervalo que cubra todas las alturas observadas en las madres.

17. Repita el item anterior usando  $h = 1$ . Repita usando  $h = 5$ . Grafique las tres funciones en un mismo gráfico utilizando un color diferente para cada valor de  $h$ .

Hasta el momento hemos trabajado utilizando siempre un mismo conjunto de datos. Vamos ahora a implementar nuevas funciones que incluyan entre sus argumentos a los datos. Es decir,

18. Implemente una función que tenga por input un conjunto de valores de  $X$ , sus correspondientes valores de  $Y$ , un nuevo valor  $x$  donde queremos predecir, y el tamaño  $h$  de la ventana que vamos a utilizar a la hora de hacer promedios móviles. `predigoVentana(X, Y, xNuevo, h)`
19. Pensar en alguna situación relacionada a su campo de trabajo donde pueda ser importante predecir cierta variable respuesta con una (o varias) variables explicativas. Contanos el problema, el marco teórico y avisanos si tenés datos que puedas compartir. Así vemos que dá!

## 4. Por si fuera poco

20. Calcular ahora predicciones para la altura del hijo conociendo la altura de la madre haciendo promedio de vecinos cercanos. Para ello, implementar una función que tenga por input un conjunto de valores de  $X$ , sus correspondientes valores de  $Y$ , un nuevo valor  $x$  donde queremos predecir, y la cantidad  $k$  de vecinos que vamos a utilizar a la hora de hacer promedios. `predigoVecinos(X, Y, xNuevo, k)`