Tareas de Clase 4

Ejercicio Datos simulados de altura.

Instrucciones: **En esta dirección página podrán acceder a los datos simulados que va a utilizar para resolver esta guía. La idea es que cada grupo trabaje con sus propios datos y por eso les pedimos que ingresen el número de libreta (o 5 últimos del DNI) de uno de los integrantes del equipo para suministrar datos de forma personalizada. Cabe destacar que 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. Los datos corresponden en la primera columna a la altura de une hije, en la segunda a su género, en la tercera la contextura de su madre y en la cuarta la altura de su madre. Nos concentraremos en el análisis de la altura de les hijes.

2. Identificar el nombre de las columnas del data.frame.

3. Realizar un histograma de las alturas de les hijes. ¿Cuántas modas se observan? ¿A qué se puede atribuir?

4. Calcular la ventana de convalidación cruzada para hacer una estimación no paramétrica de la densidad basada en el núcleo normal utilizando como variable la columna de alturas con las que se realizó el histograma. Explorar el comando `plot(density(variable))` usando la ventana de convalidación cruzada calculada.

   ¿Qué está pasando? ¿Cuántas modas observa? ¿A qué se puede atribuir?

5. 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. Repetir el item anterior para cada género.

6. Superponer en cada histograma del ítem anterior un estimador de la función de densidad normal usando la ventana de convalidación cruzada obtenida para cada caso.

7. Indicar con qué valor se puede predecir la altura de un hijo (Masculino). Indicar con qué valor se puede predecir la altura de una hija (Femenino).

8. Indicar ahora con qué medida se puede predecir la altura de un hijo, sabiendo que su mamá es de contextura pequeña, y obtener su valor para su conjunto de datos. Comparar el valor de la predicción con la predicción obtenida en el ítem anterior para el mismo género. ¿Qué se puede observar?

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

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

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

10. Indicar si hay alguna madre de hijo varón cuya altura sea 156 cm.

11. Vamos ahora a predicir la altura de un hijo correspondiente a una mamá que mide $x = 156$ cm haciendo *promedio local* centrado en 156 con ventana de tamaño h =1 (cm).

    a) Indicar cúantos casos hay donde la madre registra 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.

    c) repetir con $h = 2$.

12. Por otra parte, realizar la predicción para la altura de un hijo de una mamá que mide x = 156 cm calculando el promedio de los $k = 7$ vecinos más cercanos.

3. Implementando funciones

Hasta ahora hemos trabajado con un valor posible de altura de la madre (156) y dos valores posibles de ventana ($h = 1$ y $h = 2$). 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 elegidos para hacer el promedio local.

13. Implementar 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_masc,altura_madre_masc,altura_mama_nueva,h)`

14. Graficar la función `predigo_altura_masculino` usando h =1 y evaluándola en 100 valores de `altura_mama_nueva` que varían sobre una grilla de puntos equidistantes entre 155 y 160.

15. Repetir el ítem anterior usando $h = 2$. Repetir usando $h = 5$. Graficar las tres funciones en un mismo gráfico utilizando un color diferente para cada valor de h.

16. Hallar mediante el criterio de Convalidación Cruzada $CV(h)$ la ventana óptima, h_{opt}. Realizar la búsqueda en una grilla para valores de h entre 2.2 y 4 con paso 0.05. Graficar un plot de h vs. $CV(h)$. ¿Cuánto vale $CV(h_{\text{opt}})$?