Grupo 15

Integrantes: Santiago Wirth, Felipe Iglesias Berrondo

Tema 1

Se tiene un vector aleatorio (X,Y) donde X= Concentración de azufre en una aleación (%), Y= Dureza de la aleación. Un analista propone la siguiente función de densidad de probabilidad conjunta:

$$\boxed{f(x,y) = \frac{x^{10}}{72}\, y^4\, e^{-x^2 y}\ \mathbb{I}\{x \in [1,4], y > 0\}}$$

Consignas para todos los temas:
a) Para el vector aleatorio dado, hallar la función de regresión φ(x) = E(Y|X = x).

b) Para una semilla, simular 150 realizaciones del vector aleatorio y graficarlas en un scatter-plot.
Superponer al gráfico la función de regresión hallada en a).

c) Con las realizaciones obtenidas, aproximar la recta de regresión e incluirla en el scatter-plot. ¿Qué se observa?

d) Con las realizaciones obtenidas, aproximar la función de regresión usando el método de Nadaraya-Watson, con núcleo rectangular y dos ventanas h distintas. Incluir las curvas en el scatter-plot.¿Qué se observa?

e) Aproximar la función de densidad marginal de Y con un método que crea conveniente y graficarla. (Opcional: ¿Es fácil obtener la densidad analíticamente y superponerla al gráfico realizado? Se sugiere probar con alguna herramienta de cálculo).

Importante: Para realizar simulaciones, sólo se pueden usar los comandos runif(n), que devuelve n realizaciones de una variable U(0,1) y rnorm(n) que devuelve n realizaciones de una variable N(0,1).

Resolver el problema planteado. Se debe armar un breve informe (máximo 3 carillas) que incluya la resolución de las consignas pedidas, una explicación sobre cómo se hizo la modelización (es muy importante no incluir comentarios sobre el código en lenguaje R en esta parte), los resultados numéricos, los gráficos y cualquier otro desarrollo o conclusión que el grupo crea conveniente.
Entregar un archivo en formato pdf, que tenga como anexo un archivo en formato R con el código. El código debe arrojar los mismos resultados que figuran en el informe.

a) Cálculo analítico de la función de regresión.

$$f_{X,Y}(x, y) = f_{Y|X=x}(y). f_X(x)$$

$$f_{X,Y}(x, y) = \frac{x^{10} y^4}{24} e^{-yx^2} 1\{y > 0\} \frac{1}{3} 1\{x \in [1, 4]\}$$

$$Y|X = x \sim \Gamma(5 , x^2)$$

$\varphi(x) = E[Y|X = x] = \frac{5}{x^2}$ ⇒ REGRESIÓN VERDADERA

Ahora voy a calcular recta de regresión estimada:

$$Y(x) = \frac{Cov(x,y)}{V(x)} (x - E(x)) + E(y)$$

$$Y(x) = \frac{Cov(x,y)}{V(x)} X + E(y) - \frac{Cov(x,y)}{V(x)} E(x)$$

Siendo

$$\hat{a} = \frac{\widehat{Cov}(x,y)}{\widehat{V}(x)}$$

$$\hat{b} = \overline{Y} - \frac{\widehat{Cov}(x,y)}{\widehat{V}(x)}. \overline{X}$$

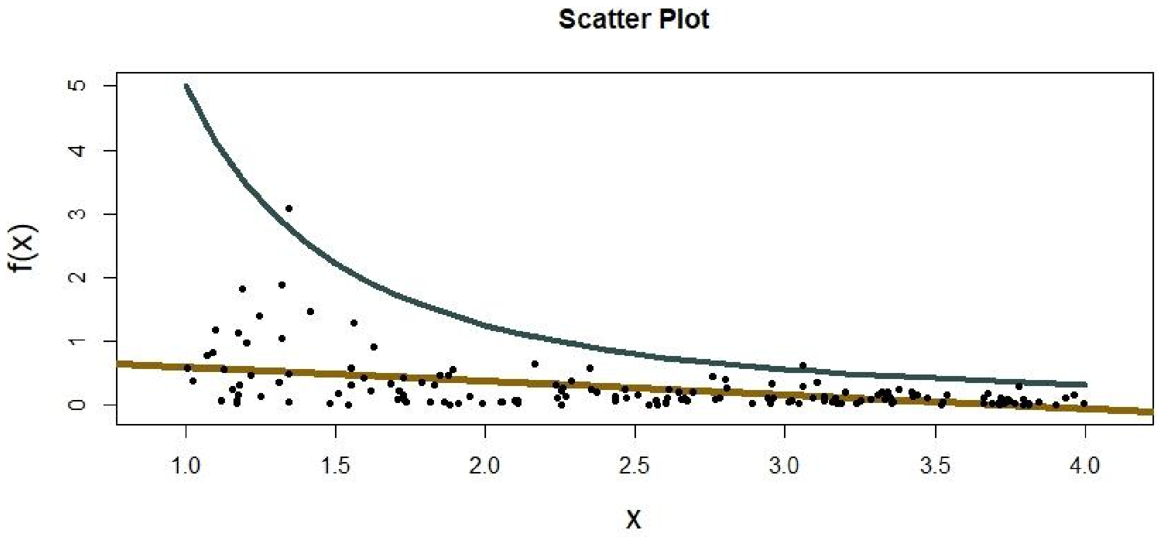
$$\Rightarrow \widehat{Y}(x) = \hat{a}\, x + \hat{b}$$

$\Rightarrow \widehat{Y}(x) =- 0,2182\, x + 0,8102$ ⇒ RECTA DE REGRESIÓN ESTIMADA

MODELIZACIÓN EN R

b) Lo primero que hacemos es simular 150 realizaciones de X, y luego para esas x, 150 realizaciones de Y dado X.
Después juntamos estos valores en una matriz para obtener los pares de puntos (x,y) para luego graficarlos en el scatter plot.
Ahora añadimos al gráfico la función de regresión obtenida analiticamente.

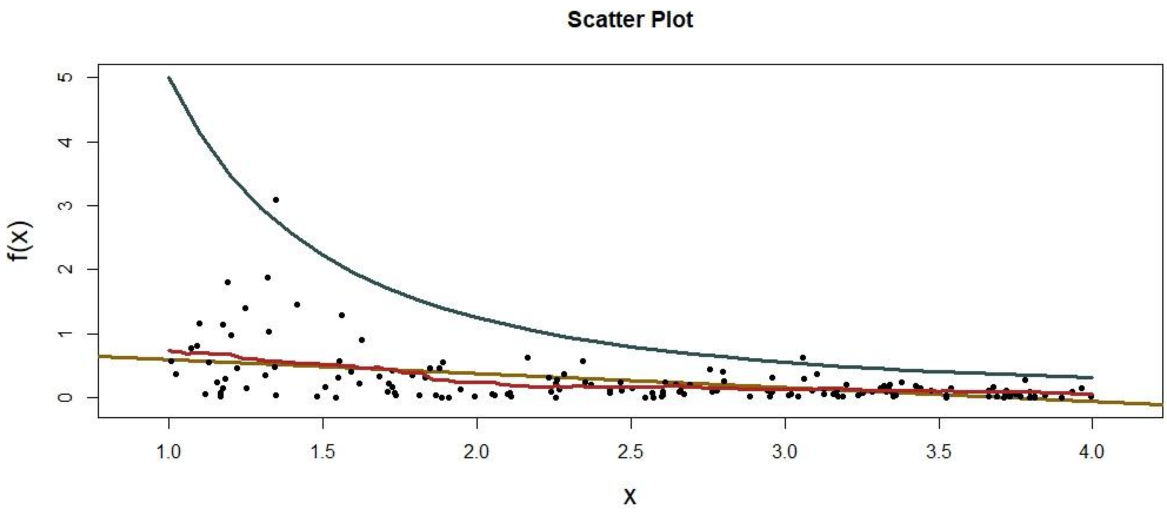
c) En este punto añadimos la recta de regresión aproximada, en el gráfico, lo que se puede observar es que la recta aproxima muy bien a la función ya que esta minimiza el error cuadrático medio lo que lo hace el mejor predictor entre x e y.



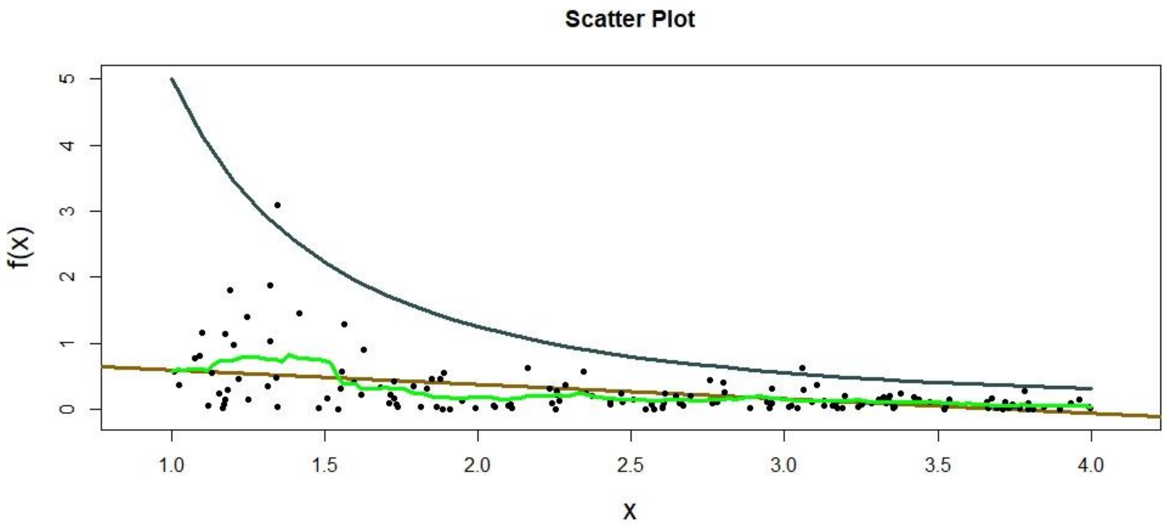
d) Para la simulación de la función de regresión mediante el método Nadaraya-Watson, nos basamos en el método de resolución analítica, que dice que para cada punto x, se puede plantear la siguiente ecuación:

$$\hat{\varphi}(x) = \frac{\sum_{i=1}^{n} Y_i \, \mathbf{1}\{|X_i - x| \leq \mathrm{h}\}}{\sum_{i=1}^{n} \mathbf{1}\{\,|X_i - x| \leq \mathrm{h}\}}$$

Por lo tanto, procedimos a armar una función la cual tomara todos nuestros datos de x y realizara la evaluación necesaria para la ecuación, es decir $\left|x_i - x\right| \leq h$, para dos valores diferentes de h. Una vez obtenidos dichos valores, registramos cuales son y cuántos son, y por último, tomamos cada valor de y correspondiente a dichos x, y tomamos su valor promedio. Para el gráfico, tomamos cada valor promedio obtenido y lo graficamos en forma de línea, uniendo puntos los cuales fueran $(x_i, y(x_i))$. A continuación, agregamos los gráficos correspondientes a las rectas de regresión obtenidas con $h_1 = 0.5, h_2 = 0.2$. Nótese que las curvas fueron realizadas por sobre el gráfico previo. La curva roja corresponde a h_1, y la verde a h_2



Scatter Plot
f(x)
x
0
1
2
3
4
5
1.0
1.5
2.0
2.5
3.0
3.5
4.0



Scatter Plot
f(x)
x
0
1
2
3
4
5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

e) Para aproximar la función de densidad marginal de Y, realizamos un histograma de los valores de $Y|X = x$, y por sobre él un gráfico de la densidad de $Y|X = x$.