**Bioestadística** **Universidad de la República**
**Curso 2022** **Facultad de Ciencias**

## Práctico 10: Resultados

**Ejercicio 1:** (Viento en función del Ozono en airquality)

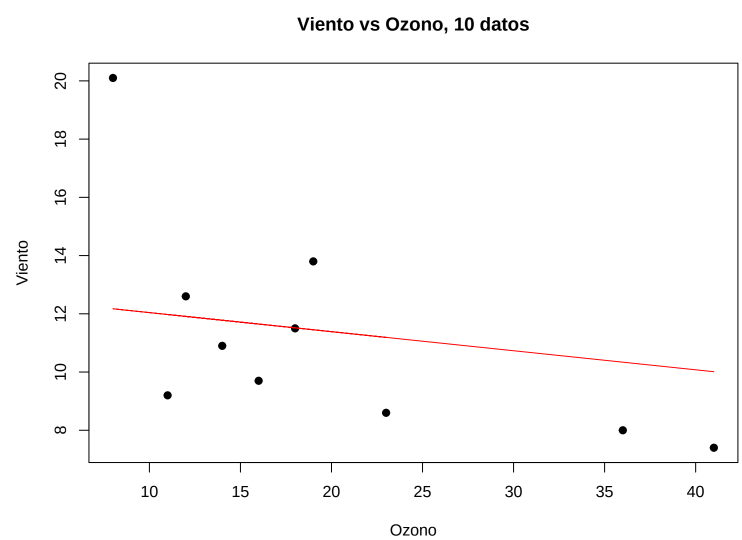
Antes de hacer el ejercicio, ver la sección **13.4** del libro del curso para ver cómo cargar los datos a R y hacer el análisis.

a) Si $X$ = **Ozono** e $Y$ = **Viento**, planteamos $Y = aX + b$, donde $a$ y $b$ serán los coeficientes que estimaremos.

b) Planteamos el test de hipótesis $\begin{cases} H_0 : a = 0, \\ H_1 : a \neq 0. \end{cases}$ Si se rechaza $H_0$, entonces hay indicios de que haya dependencia lineal entre las variables **Ozono** y **Viento**.

c) Planteamos: `WindvsOzonwe = lm(Wind~ Ozone)` y obtenemos: $\hat{a} = -0{,}066$, $\hat{b} = 12{,}70$.

d) Graficamos, por ejemplo, los primeros diez valores $(o_1, v_1), \ldots (o_{10}, v_{10})$ y obtenemos:




e) El p-valor para el test planteado en la parte b) es:

$$p - valor = \mathbb{P}(|T| > |t_{obs}|) = \mathbb{P}(|T| > |-8{,}09|) = 9{,}09 \times 10^{-13}$$

(el valor de $t_{obs}$ y el p-valor los sacamos del `summary(WindvsOzone)`). Como el p-valor da muy chiquito, rechazamos $H_0$ y podemos asumir que hay una dependencia lineal entre **Viento** y **Ozono**.

f) Planteamos $\begin{cases} H_0 : b = 0, \\ H_1 : b \neq 0. \end{cases}$ Igual que antes, el p-valor para este test lo sacamos del `summary(WindvsOzone)`:

$$p - valor = \mathbb{P}(|T| > |t_{obs}|) = \mathbb{P}(|T| > |29{,}28|) = 2 \times 10^{-16},$$

por lo tanto rechazamos $H_0$.

g) El valor del coeficiente de determinación $R^2$ lo sacamos del `summary(WindvsOzone)` como `Multiple R squared` y vale $R^2 = 0{,}3752$. Esto nos dice que la variación explicada por el modelo de regresión es casi el 40 % de la variación total, por lo tanto el modelo lineal no se ajusta muy bien a estos datos.

**Ejercicio 2:** (Teórico)

**Ejercicio 3:** (Normal condicionada)
Sea $X \sim Unif([-2,-1] \cup [1,2])$. Definimos una variable aleatoria $Y$ tal que $Y|X \sim N(3X-1, 1/4)$.

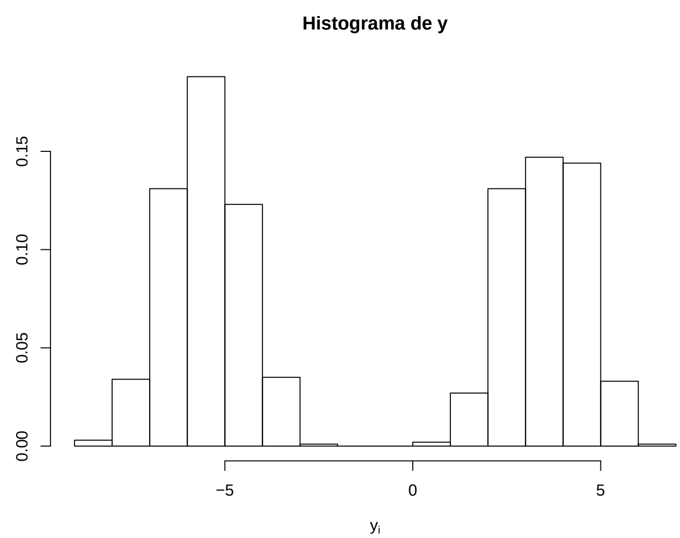
a) Observar que $X$ tiene la misma distribución que la siguiente variable:

$$X = \begin{cases} U, & \text{con probabilidad } \frac{1}{2}, \\ -U, & \text{con probabilidad } \frac{1}{2} \end{cases} ,$$

siendo $U \sim U(1,2)$. Es decir que podemos tirar una moneda, si sale CARA sortear $U$ y si sale NÚMERO, sortear $-U$ (ver el script de R).

b) Para cada dato $x_i$ obtenido antes, definimos $Y_i \sim N(3x_i - 1, \sigma = 0{,}5)$ y simulamos una realización de $Y_i$ (ver el script)

c) Obtuvimos el siguiente histograma para $(y_1, \ldots, y_{1000})$:

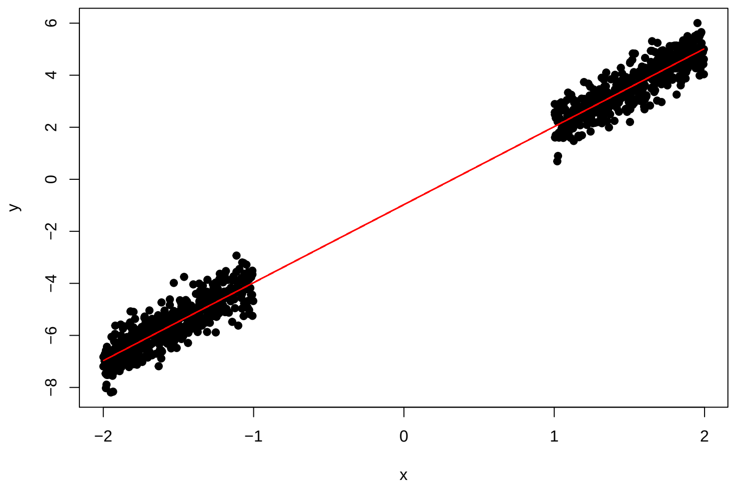



No parece corresponder a una distribución normal.

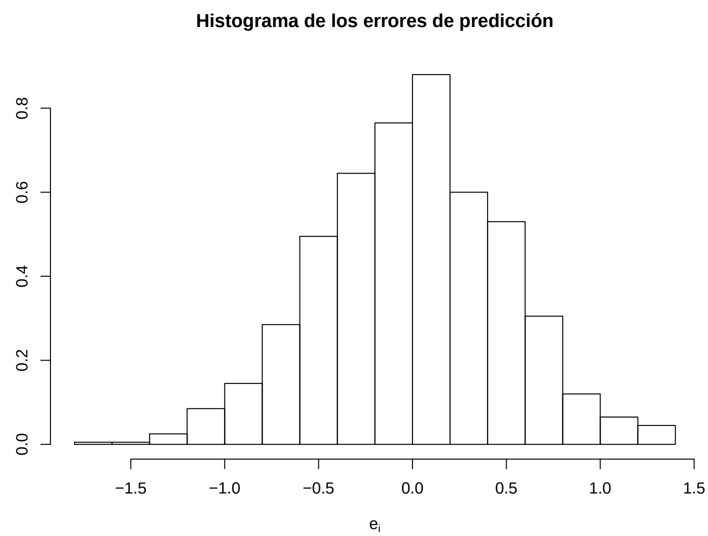
d) Realizamos una regresión lineal para ajustar los valores $(x_1, y_1), \ldots (x_n, y_n)$ obtenidos a un modelo $Y = \hat{\beta}_0 + \hat{\beta}_1 X$ y obtenemos:

$$\hat{\beta}_0 = 0{,}97 (\text{ es el } b \text{ de las notas}) \quad \hat{\beta}_1 = 3{,}00 \text{ (es el } a \text{ de las notas)}$$

Obtenemos:




e) Sean $e_i = \hat{y}_i - y_i$ los errores de predición del modelo, siendo $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. El siguiente gráfico contiene el histograma de los errores $e_i$:




Finalmente, plenteamos un test $\begin{cases} H_0 : \text{Los } e_i \text{ tienen distribución normal} , \\ H_1 : \text{ NO } H_0. \end{cases}$

Por ejemplo, realizamos un test de Lilliefors para comparar los datos con un $N(\mu, \sigma^2)$ con $\mu$ y $\sigma$ desconocidos y obtenemos que el p-valor para este test es casi 1, por lo que no podemos rechazar $H_0$.