Nicolás Felipe Trujillo Montero 1ºB Ingeniería Informática

DNI: [redacted]

Ejercicio Extra

1) Consideremos la siguiente tabla de frecuencias absolutas, donde la variable Y representa el peso (en quintales) de la cosecha de naranjas recogidas durante el primer, segundo y tercer trimestre del año 2019.

X \ Y	1	4	8	
1	6	4	0	$N_{i1} = 10$
2	2	3	2	$N_{i2} = 7$
3	0	0	10	$N_{i3} = 10$
	$N_{1j} = 8$	$N_{2j} = 7$	$N_{3j} = 12$	$N_{ij} = 27$

a) Usar el método de los mínimos cuadrados para determinar el sistema de ecuaciones normales del modelo $Y = \sqrt{aX + bX^3}$

b) Realizar dicho ajuste y determinar la bondad del mismo.

c) ¿Es este ajuste mejor o peor que un ajuste lineal?

d) Predecir mediante el modelo $Y = \sqrt{aX + bX^3}$ el peso en quintales que se prevé que se (~~recoja~~) recoja en el primer trimestre del 2020.

e) Considera la variable $Y/_{X} \geq 2$. Calcula el coeficiente de variación CV, el percentil P_{85}, y estudia la asimetría de esta nueva variable.

a) Vamos a realizar el método de mínimos cuadrados para calcular un ajuste que se acerque a los datos dados. Para ello, usaremos la fórmula dada. Gracias al gradiente de la función (F) que lleva consigo dos variables independientes (x e y), podemos minimizar la función calculando los parámetros

$$Y = \varphi(x) = \sqrt{a\varphi_0(x) + b\varphi_1(x)} \Rightarrow \begin{cases} \varphi_0(x) = X \\ \varphi_1(x) = X \end{cases}$$

$$\Rightarrow Y^2 = aX + bX^{-3} \Rightarrow F(X,Y) = \left(Y^2 - (aX + bX^{-3})\right)^2$$

$$\nabla F(x,y) = \begin{cases} \dfrac{\partial F(x,y)}{\partial a} = \\[2ex] \dfrac{\partial F(x,y)}{\partial b} = \end{cases} \begin{pmatrix} \sum_i \sum_j Y^2 \cdot X \\[2ex] \sum_i \sum_j Y^2 \cdot X^3 \end{pmatrix} = \begin{pmatrix} \sum_i \sum_j X_i^2 & \sum_i \sum_j X_i^4 \\[2ex] \sum_i \sum_j X_i^4 & \sum_i \sum_j X_i^6 \end{pmatrix} \begin{pmatrix} a \\[2ex] b \end{pmatrix}$$

$< X, X^{-3} >$ (Base)

{Propiedad, sumatoria de una constante a, sin subíndice correspondiente, es igual a la constante por el número de sumandos}

$$\sum_i^{n=3} \sum_j^{k=3} Y_i^2 \cdot X_j = (1^2 \cdot 1 \cdot 6) + (4^2 \cdot 1 \cdot 4) + (1^2 \cdot 2 \cdot 2) + (4^2 \cdot 2 \cdot 3) + (8^2 \cdot 2 \cdot 2) + (8^2 \cdot 3 \cdot 10) = 2346$$

$$\sum_i^{n=3} \sum_j^{k=3} Y_i^2 \cdot X_j^3 = (1^2 \cdot 1^3 \cdot 6) + (4^2 \cdot 1^3 \cdot 4) + (1^2 \cdot 2^3 \cdot 2) + (4^2 \cdot 2^3 \cdot 3) + (8^2 \cdot 2^3 \cdot 2) + (8^2 \cdot 3^3 \cdot 10) = 18774$$

$$\sum_i^{n=3} \sum_j^{k=3} X_j^2 = n \sum_j^{k=3} X_j^2 = (1^2 \cdot 10) + (2^2 \cdot 7) + (3^2 \cdot 10) = 128 \cdot n = 128 \cdot 3 = 384$$

$$\sum_i^{n=3} \sum_j^{k=3} X_j^4 = n \sum_j^{k=3} X_j^4 = (1^4 \cdot 10) + (2^4 \cdot 7) + (3^4 \cdot 10) = 932 \cdot n = 2796$$

$$\sum_i^{n=3} \sum_j^{k=3} X_j^6 = n \sum_j^{k=3} X_j^6 = (1^6 \cdot 10) + (2^6 \cdot 7) + (3^6 \cdot 10) = 7748 \cdot n = 23244$$

$$\begin{pmatrix} 2346 \\ 18774 \end{pmatrix} = \begin{pmatrix} 384 & 2796 \\ 2796 & 23244 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\boxed{a = \frac{149}{81} \qquad b = \frac{95}{162}}$$

$$Y = \sqrt{\frac{149}{81} X + \frac{95}{162} X^{-3}}$$

6) Y_i: 1 1 1 1 1 ¦ 4 4 4 4 ¦ 1 1 ¦ 4 4 4 ¦ 8 8 ¦ 8 8 8 8 8 8 8 8 8 8

X_i: 1 1 1 1 1 1 ¦ 1 1 1 1 ¦ 2 2 ¦ 2 2 2 ¦ 2 2 ¦ 3 3 3 3 3 3 3 3 3 3

Para realizar la bondad del ajuste, vamos a recurrir al coeficiente de determinación:

$$\boxed{R^2 = 1 - \frac{V_r}{V(y)}}$$

Para ello, necesitamos calcular la varianza residual y la varianza de y.

$$V_r = \sum_i f_i (e_i - \bar{e})^2$$

$$V_y = \sum_i f_i (y_i - \bar{y})^2$$

$$V_y = \sum_i f_i (y_i - \bar{y})^2 = \frac{1}{N}\sum_i n_i (y_i - \bar{y})^2 = \frac{1}{27}\cdot\left[\left(8\cdot\left(1-\frac{132}{27}\right)^2\right) + \left(7\cdot\left(4-\frac{132}{27}\right)^2\right) + \left(12\cdot\left(8-\frac{132}{27}\right)^2\right)\right] = \frac{728}{81}$$

$$\boxed{\bar{y} = \frac{\sum_i^{n=3} y_i n_i}{N}} \Rightarrow \frac{1\cdot 8 + 4\cdot 7 + 8\cdot 12}{27} = \frac{132}{27}$$

$$V_r = \sum_i f_i (e_i - \bar{e})^2 = \frac{1}{N}\sum_i n_i (e_i - \bar{e})^2 = \frac{1}{27}\cdot\left\{\left[8\left(\left(1-\sqrt{\frac{149}{81}+\frac{95}{162}}\right)-2'7377\right)^2\right] + \left[\left(7\cdot\left(\left(4-\sqrt{\frac{149}{81}\cdot 1+\frac{95}{162}}\right)-2'73777\right)^2\right)\right] + \left[12\cdot\left(\left(8-\sqrt{\frac{149}{81}\cdot 2^{-}+\frac{95}{162}\cdot 2^3}\right)-2'73777\right)^2\right]\right\} = 5'7346$$

$$\bar{e} = \frac{\sum_i e_i n_i}{N} = \{e_i = y_i - y_{est}\} = \frac{1}{27}\cdot\left[8\left(1-\sqrt{\frac{149}{81}\cdot 1+\frac{95}{162}\cdot 1^3}\right) + 7\left(4-\sqrt{\frac{149}{81}\cdot 1+\frac{95}{162}\cdot 1^3}\right) + 12\left(8-\sqrt{\frac{149}{81}\cdot 2+\frac{95}{162}\cdot 2^3}\right)\right] = 2'7377$$

$$\boxed{y_{est} = \sqrt{\frac{149}{81}x^{-}+\frac{95}{162}x^{-3}}}$$

Vamos a realizar el coeficiente de determinación

$$\boxed{R^2 = 1 - \frac{5'7346}{\frac{728}{81}} = 0'36395 \quad (36'395\%)}$$

* No sé si la realización de dicho ejercicio es correcta debido a que al calcular la y_{est} existen varias x para una misma y, por eso he escogido las mayores x.

c) Obviamente, realizando un ajuste lineal, no nos saldría una bondad buena. Esto es debido a que como he descrito antes, existen varias x para una misma y (caso que no ocurre en las rectas, a no ser que la recta sea horizontal, que tampoco es el caso)

d) Para ello, vamos a implementar una $X=5$, tal que sea un "quinto trimestre" del 2019.

$$Y = \sqrt{\frac{149}{81} \cdot 5 + \frac{95}{362} \cdot (5)^3} = 9'083 \simeq 9 \text{ quintales}$$

e) Variable $\frac{Y}{X} \geq 2 \Rightarrow$ {Casos posibles $(X;Y) = (1,4), (2,4), (2,8), (3,8)$}

→ Para calcular el coeficiente de variación de la nueva variable, tenemos que:

$$\boxed{CV = \frac{\sigma}{|\bar{x}|}} = \frac{\sigma}{|\overline{\frac{y}{x}}|} = \frac{0'72928}{|\frac{170}{57}|} = 0'24452 = 24'452\%$$

$$\sigma = +\sqrt{V} = \sqrt{\frac{1}{N}\sum_{i=1}^{K}\left(\left(\frac{y}{x}\right)_i - \left(\overline{\frac{y}{x}}\right)\right)^2 \cdot n_i} = \sqrt{\frac{1}{19}\left[\left(4\cdot\left(4-\frac{170}{57}\right)^2\right) + \left(3\cdot\left(2-\frac{170}{57}\right)^2\right) + \left(2\cdot\left(4-\frac{170}{57}\right)^2\right)\right.}$$

$$\left.+\left(10\cdot\left(\frac{8}{3}-\frac{170}{57}\right)^2\right)\right] = \sqrt{\frac{1}{19}\left(\frac{13456}{3249}+\frac{3136}{1083}+\frac{6728}{3249}+\frac{360}{361}\right)} = 0'72928$$

$$\left(\overline{\frac{y}{x}}\right) = \frac{\sum\left(\frac{y}{x}\right)_i n_i}{N} = \frac{4\cdot4+2\cdot3+4\cdot2+\frac{8}{3}\cdot10}{19} = 2'9825 = \frac{170}{57}$$

→ Calcular el percentil P_{85}

$P_{85} = \frac{Y}{X}\left(\frac{85}{100}\right)$ = {Como existen tres valores $\left(\frac{8}{3}, 2, 4\right)$ para los que $\frac{y}{x} \geq 2$ en los datos dados tenemos que calcular el 85%}

$\frac{y}{x}$	n_i	f_i		F_i
$\frac{4}{1}$	4	$\frac{4}{19}$	$\approx 0'21053$	$\frac{4}{19} \approx 0'21053$
$\frac{4}{2}$	3	$\frac{3}{19}$	$\approx 0'15789$	$\frac{7}{19} \approx 0'36842$
$\frac{8}{2}$	2	$\frac{2}{19}$	$\approx 0'10526$	$\frac{9}{19} \approx 0'47368$
$\frac{8}{3}$	10	$\frac{10}{19}$	$\approx 0'52632$	$\frac{19}{19} = 1$

$N = 19$

→ Para encontrar el percentil P_{85}, tenemos que buscar en la tabla (→ Columna de Frecuencias Acumuladas) cual es el valor mayor o igual a $0'85$, es decir $\frac{8}{3} \approx P\left(\frac{85}{100}\right) = \frac{8}{3}$

→ Para realizar el estudio de la simetría de esta nueva variable $\frac{Y}{X}$, voy a utilizar el método de Pearson

$$A_p = \frac{\overline{\frac{y}{x}} - Mod}{\sigma} = \frac{\frac{170}{57} - \frac{8}{3}}{0'72928} = 0'43302 < 1 \quad \text{Asimetría a la izquierda}$$