

## Relación 2 de EDIP.

Alejandro Villanueva Prados

30 de marzo de 2019

1. En una encuesta de familias sobre el número de individuos que la componen ( $X$ ) y el número de personas activas en ellas ( $Y$ ) se han obtenido los siguientes resultados:

$X \backslash Y$	1	2	3	4
1	7	0	0	0
2	10	2	0	0
3	11	5	1	0
4	10	6	6	0
5	8	6	4	2
6	1	2	3	1
7	1	0	0	1
8	0	0	1	1

- a) Calcular la recta de regresión de  $Y$  sobre  $X$ .

$$y = 0,3147x + 0,5322.$$

- b) ¿Es adecuado suponer una relación lineal para explicar el comportamiento de  $Y$  a partir de  $X$ ?

Según el coeficiente de correlación lineal

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = 0,535$$

el ajuste no es el idóneo, pero los hay peores.

2. De las estadísticas de “Tiempos de vuelo y consumos de combustible” de una compañía aérea, se han obtenido datos relativos a 24 trayectos distintos realizados por el avión DC-9. A partir de estos datos se han obtenido las siguientes medidas:

$$\begin{aligned}\sum y_i &= 219,719 & \sum y_i^2 &= 2396,504 & \sum x_i y_i &= 349,486 \\ \sum x_i &= 31,470 & \sum x_i^2 &= 51,075 & \sum x_i^2 y_i &= 633,993 \\ \sum x_i^4 &= 182,977 & \sum x_i^3 &= 93,6\end{aligned}$$

La variable  $Y$  expresa el consumo total de combustible, en miles de libras, correspondiente a un vuelo de duración  $X$  (el tiempo se expresa en horas, y se utilizan como unidades de orden inferior fracciones decimales de la hora).

- a) Ajustar un modelo del tipo  $Y = aX + b$ . ¿Qué consumo total se estimaría para un programa de vuelos compuesto de 100 vuelos de media hora, 200 de una hora y 100 de dos horas? ¿Es fiable esta estimación?

*Solución:* Comenzamos analizando nuestra población y los datos que tenemos: observamos que la población es de tamaño  $n = 24$ , además nos indican que los 24 vuelos son distintos así que  $n_i = 1; \quad \forall i \in \{1 \dots 24\}$

Una vez tomadas estas consideraciones, nos centramos en la pregunta: encontrar un ajuste lineal mediante un polinomio de grado 1. La expresión de la función será:

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x}),$$

así que pasamos a calcular los datos necesarios:

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{31,470}{24} \approx 1,311 \text{ Horas de vuelo,}$$

$$\bar{y} = \frac{1}{n} \sum y_i = \frac{219,719}{24} \approx 9,155 \text{ Miles de libras de combustible,}$$

$$\sigma_{xy} = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) = m_{11} - m_{10}m_{01} = \frac{1}{n} \sum x_i y_i - \frac{1}{n^2} \sum x_i \sum y_i \approx 2,557$$

$$\sigma_x^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2, \approx 0,40854$$

finalmente, la recta de regresión queda:

$$y = 6,26x + 0,946.$$

Pasamos ahora a la segunda cuestión del apartado, para ello aplicamos la función obtenida:

$$6,26 \cdot 0,5 + 0,946 = 4,076 \text{ (Consumo estimado para un vuelo de media hora)}$$

$$6,26 \cdot 1 + 0,946 = 7,206 \text{ (Consumo estimado para un vuelo de una hora)}$$

$$6,26 \cdot 2 + 0,946 = 13,466 \text{ (Consumo estimado para un vuelo de dos horas)}$$

Ahora escalamos estos resultados multiplicando por el número de vuelos y sumamos todo, obteniendo un consumo total de 3195,4 miles de libras de combustible.

Para cuantificar la fiabilidad de la predicción, vamos a emplear el coeficiente de correlación lineal, dado por

$$r = \pm \sqrt{r^2} = \frac{\sigma_{xy}}{\sigma_x \sigma_y},$$

calculamos  $\sigma_x$  y  $\sigma_y$ :

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{0,401} \approx 0,640$$

$$\sigma_y = \sqrt{\frac{1}{n} \left( \sum y_i^2 - \frac{1}{n} \sum y_i \right)} \approx 4,005,$$

con estos datos, el coeficiente de correlación lineal queda 0,99758, lo que nos indica un ajuste casi perfecto y una muy alta fiabilidad.

- b) Ajustar un modelo del tipo  $Y = a + bX + cX^2$ . ¿Qué consumo total se estimaría para el mismo programa de vuelos del apartado a)?

*Solución:* Los coeficientes de nuestra función de ajuste vienen dados por las soluciones del siguiente sistema:

$$\begin{cases} m_{01} &= a_0 + a_1m_{10} + a_2m_{20}, \\ m_{11} &= a_0m_{10} + a_1m_{20} + a_2m_{30}, \\ m_{21} &= a_0m_{20} + a_1m_{30} + a_2m_{40}. \end{cases}$$

Para ello calculamos los distintos momentos valiéndonos de los datos iniciales y resolvemos el sistema, obtenemos:

$$a_0 = 0,800 \quad a_1 = 6,558 \quad a_2 = -0,112.$$

Con estos resultados nuestra función de ajuste nos queda:  $Y = -0,112X^2 + 6,558X + 0,8$ . Volviendo a calcular las predicciones con esta función nos queda un consumo estimado de 3198,275 miles de libras de combustible.

- c) ¿Cuál de los dos modelos se ajusta mejor? Razonar la respuesta.

*Solución:* Para comparar los dos modelos vamos a usar el coeficiente de correlación, definido como sigue:

$$\eta^2_{Y/X} = \frac{\sigma_{ey}^2}{\sigma_y^2},$$

donde  $\sigma_{ey}^2 = \frac{1}{n} \sum (\hat{y}_j - \bar{y})^2 =$

**NOTA:** no está terminado el apartado c), no sé cómo hacerlo :’(