

# Ejercicios Estadística Descriptiva

Juan Rodríguez

October 28, 2024

## 1 Problema 1

El número de cetano se emplea como indicador de la calidad de ignición del combustible utilizado en un motor diésel. En un estudio se trató de ajustar mediante un modelo lineal dicho indicador (Y) en función del índice de yodo X (en gramos). Se analizó una muestra de 14 combustibles y se obtuvieron los siguientes cálculos:

- $\sum_{i=1}^n x_i = 1307.5$
- $\sum_{i=1}^n y_i = 779.2$
- $\sum_{i=1}^n x_i^2 = 128913.93$
- $\sum_{i=1}^n y_i^2 = 43745.22$
- $\sum_{i=1}^n x_i y_i = 71347.3$

- a) Predecir utilizando un modelo lineal el número de cetano para un combustible que tiene un índice de yodo de 115 y explicar el significado del coeficiente de regresión en ese modelo. ¿Cuál es la fiabilidad de dicha predicción?
- b) Dar una medida de la bondad del ajuste e interpretarla.
- c) ¿Cuánta variabilidad en el número de cetano no se explica por el índice de yodo? Indicar también qué % representa ese valor.

Se han obtenido los siguientes datos adicionales sobre el número de cetano de la muestra de 14 combustibles:

- La mitad de ellos tiene un índice de cetano comprendido entre 87 y 114.
- La mitad de los combustibles de la muestra tienen un índice de cetano menor o igual que 100.
- Los índices de cetano mínimo y máximo obtenidos en la muestra de combustibles han resultado de 51 y 173.
- El índice de cetano más frecuente ha sido de 92.

- d) Construir un box plot para el número de cetano e interpretarlo.
- e) ¿Qué valor medio es más representativo, el del número de cetano o el del índice de yodo? Explicar por qué.

## 2 Problema 2

En una empresa se pretende analizar la relación existente entre: las ventas totales ( $X$ ) y las exportaciones ( $Y$ ), expresadas en Millones de euros. Disponemos de los siguientes datos:

- Las dos rectas de regresión se cortan en el punto  $(6, 4)$ .
- Por cada millón de euros que se incrementan las exportaciones, el incremento que experimentan las ventas totales, bajo el modelo lineal de  $Y$  sobre  $X$ , es de 1.15 M euros.
- Si se incrementan en un millón de euros las ventas totales, las exportaciones aumentarían en 810.000 euros.
- La dispersión relativa de las exportaciones ( $Y$ ), medida en términos de su coeficiente de variación, ha sido del 50 %.

Se desea:

1. Con estos supuestos, predecir las exportaciones de la empresa en un período en que las ventas totales han sido de 25 Millones de euros y dar una medida de la fiabilidad de dicha predicción.
2. Calcular el coeficiente de correlación e interpretarlo.
3. Comparar la dispersión relativa de ambas distribuciones.
4. Calcular e interpretar la varianza residual y la explicada

## 3 Problema 3

Se han tomado cinco muestras con la misma cantidad de glucógeno, se les ha aplicado una cantidad de glucogenasa (en milimoles/litro): ( $X$ ), y se ha medido la velocidad de reacción, ( $Y$ ), (en micromoles/minuto). Se han obtenido los siguientes datos:

$x_i$	0.2	0.5	1.25	2.1	3.05
$y_i$	8	10	18	35	60

¿Se puede deducir a partir de estos datos que la velocidad de reacción aumenta o disminuye linealmente con la concentración de glucogenasa?

En caso afirmativo, dar una expresión matemática del modelo de ajuste y utilizarlo para predecir la cantidad de glucogenasa aplicada en una reacción cuya velocidad ha sido de 45 micromoles/minuto.

Calcular e interpretar el coeficiente de regresión de  $y$  sobre  $x$ .

¿Cuánta variabilidad en la velocidad de reacción no queda explicada por la concentración de glucogenasa? Indicar también qué % representa ese valor.

Dar una medida de la bondad del ajuste e interpretarla.  
 Representar en un boxplot los datos correspondientes a la velocidad de reacción.  
 ¿Qué conclusiones puedes extraer de la gráfica?

## 4 Problema 4

Se trató de ajustar un modelo de regresión lineal simple para analizar la relación entre las variables:

- $X$ : producción de trigo en toneladas métricas (Tm)
- $Y$ : precio del kilogramo de harina (en euros)

Disponemos de los siguientes datos relativos a los últimos 5 años:

$$\bar{x} = 28; \quad \bar{y} = 0.414; \quad \sum_{i=1}^n x_i^2 = 3958; \quad \sum_{i=1}^n y_i^2 = 0.86; \quad \sum_{i=1}^n x_i y_i = 57.68$$

- ¿Qué puedes decir de la interdependencia entre las variables?
- Predecir la producción de trigo en un año en el que el precio de la harina fue de 0.47 euros y dar una medida de la fiabilidad de dicha predicción.
- Calcular la pendiente de la recta de regresión de  $y$  sobre  $x$  e interpretarla en el contexto del problema.
- Si nos facilitan además las producciones en Tm de los 5 años de estudio, que son: 30, 28, 32, 25, 25. ¿Entre qué dos valores estará el 50% central de la distribución de producciones?

## 5 Problema 5

Queremos analizar si existe relación lineal entre la distancia y el tiempo de viaje transcurrido entre el domicilio y el lugar de trabajo. Se han analizado las distancias (km) y tiempos (minutos) de 5 empleados y los datos son los siguientes:

$X_i$ (km)	3	7	11.1	17	12
$Y_i$ (minutos)	8	18	26	40	33

- ¿Puedes decir que existe una relación lineal entre las variables distancia y tiempo de viaje? Interpretar el valor de  $r$ .
- Predecir el tiempo que tarda en llegar al trabajo un empleado que vive a 14 km de distancia y dar una medida de la fiabilidad de dicha predicción.

- c) ¿Cuál de las dos variables es más dispersa, la distancia o el tiempo? Indica cómo efectúas la comparación.
- d) ¿Qué porcentaje de variabilidad en el tiempo de llegada al trabajo no queda explicado por la distancia?

## 6 Problema 6

Un fabricante de neumáticos quiere determinar si el diámetro interior de un neumático tiene relación con el número de pinchazos en un período de 5 años. Tenemos los datos sobre el diámetro interior ( $X$ ) y el número de pinchazos en los últimos 5 años de 5 vehículos ( $Y$ ):

$X_i$ (diámetro en cm)	57	56	59	55	61
$Y_i$ (número de pinchazos)	3	2	1	4	1

- a) ¿Qué diámetro tendrán los neumáticos de un vehículo que no ha sufrido pinchazos en los últimos 5 años? ¿Y con qué grado de fiabilidad?
- b) ¿Cuál de las dos variables presenta mayor dispersión?
- c) Dar una medida de la interdependencia entre las variables e interpretarla.
- d) Calcular la varianza explicada de  $Y$ ; explica su significado en el contexto del problema.

## 7 Problema 7

Se pretende analizar si el tiempo de duración ( $X$ ) de una serie de películas (en minutos) está relacionado con el número de semanas que están en cartel ( $Y$ ), a partir de los datos correspondientes a 5 películas:

$X$ (minutos)	115	90	83	121	100
$Y$ (semanas en cartel)	15	7	22	16	9

- a) Predecir el número de semanas en cartel de una película que dura 2 horas y dar una medida de la fiabilidad de la predicción.
- b) Calcular la varianza residual e interpretarla.
- c) Dar una medida de la interdependencia entre las variables. Interpretarlo.