

Análisis de regresión

Relaciones entre variables

- Considera las siguientes variables



¿Hay alguna relación entre ellas?



Objetivos del análisis de regresión

Identificar si existe una relación entre dos variables.

Crear un modelo que defina esa relación.

Determinar una medida que indique el grado de asociación entre las variables.

Utilizar el modelo para predicción, control u optimización.

Diagrama de dispersión



- Representación gráfica de dos variables cuantitativas



- Los datos se presentan en forma de puntos en un plano cartesiano

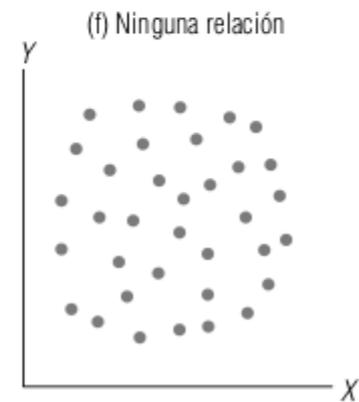
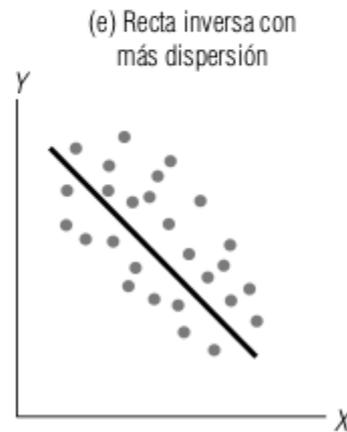
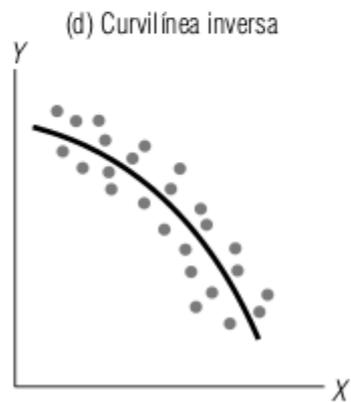
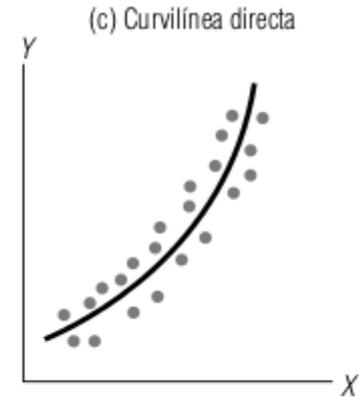
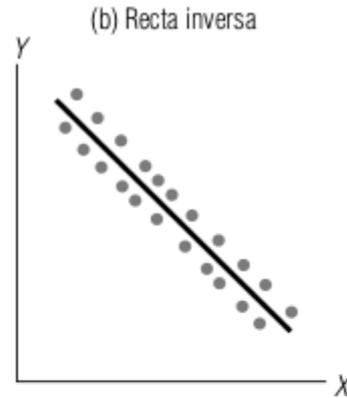
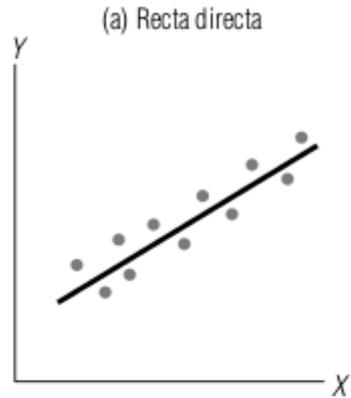


- Permite visualizar si existe una relación entre las variables y el tipo de relación que se da.



- Cada variable define una escala para representarla

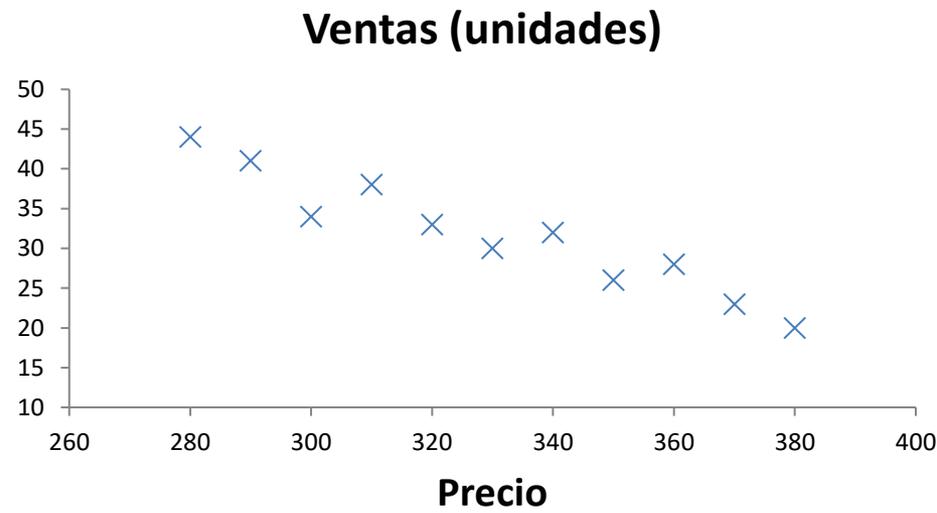
Posibles relaciones



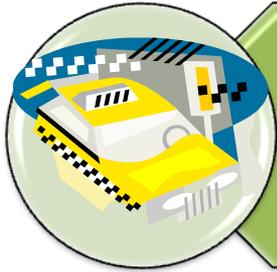
Tipos de relación

- En las relaciones **directas** si x aumenta, y también; si x disminuye, y también lo hace.
- En las relaciones **inversas** valores grandes de x corresponden a valores de y pequeños; cuando x disminuye, y aumenta.
- Por ejemplo, la ley de la demanda relaciona precios y ventas.

Precio (\$)	Ventas (unidades)
280	44
290	41
300	34
310	38
320	33
330	30
340	32
350	26
360	28
370	23
380	20



Ejemplo



Un contratista de taxis desea determinar si el peso de los autos está relacionado con el rendimiento del combustible.



Para ello selecciona una muestra de 5 autos, los cuales se pesan y después se conducen durante 100 kilómetros para determinar el kilometraje recorrido por litro.

A circular icon containing a screenshot of a data table with multiple columns and rows of text.

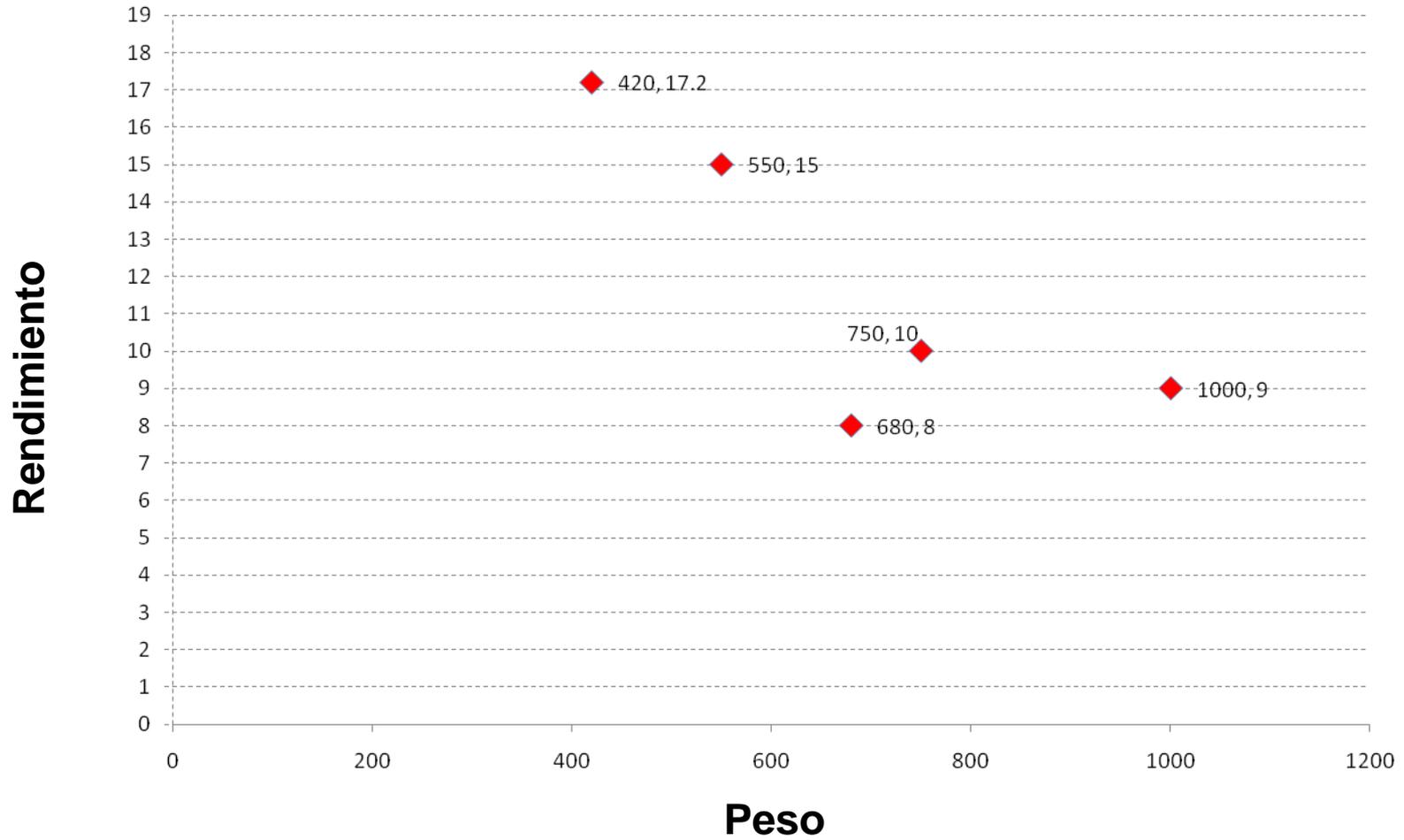
Los resultados se muestran en la siguiente tabla.

Ejemplo

Auto	Peso (kg)	Rendimiento (km/l)
1	750	10.0
2	420	17.2
3	550	15.0
4	1000	9.0
5	680	8.0

- Asignamos como x a la *variable independiente* y como y a la *variable dependiente*.
- En este caso, el rendimiento (y) depende del peso (x).
- Posteriormente, las parejas (x,y) se grafican como puntos en un plano cartesiano.

Ejemplo



Ejemplo

¿Qué tipo de relación existe entre el peso y el rendimiento?



Coeficiente de correlación

Indica la fuerza de la relación lineal entre dos variables cuantitativas.

Es un valor numérico entre -1 y 1.

-1 indica una fuerte relación lineal negativa.

0 indica que no existe relación lineal.

1 indica una fuerte relación lineal positiva.

Para una muestra, se denota como "r".

Coeficiente de correlación

- Para calcular el coeficiente de correlación r usamos

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

donde n = número de datos.

Ejemplo

Vamos agregar a la tabla de datos tres columnas que correspondan a xy , x^2 , y^2 .
Se obtienen los valores correspondientes y se suman las columnas.

auto	Peso	Rendimiento			
	x	y	xy	x^2	y^2
1	750	10	7500	562500	100
2	420	17.2	7224	176400	295.84
3	550	15	8250	302500	225
4	1000	9	9000	1000000	81
5	680	8	5440	462400	64
Sumas	3400	59.2	37414	2503800	765.84

Ejemplo

Se obtiene el valor de “r” sustituyendo los valores de las sumas indicadas:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

$$r = \frac{n(37414) - (3400)(59.2)}{\sqrt{5(2503800) - (3400)^2} \sqrt{5(765.84) - (59.2)^2}} = -0.81$$

Esto indica que en la muestra existe una relación lineal negativa clara entre el peso y el rendimiento de los automóviles.

Recta de regresión

Dado que el análisis de regresión se hace con el propósito de usar el valor de la variable independiente para pronosticar el valor de la variable dependiente, el siguiente paso, después de obtener el diagrama de dispersión es obtener una curva de aproximación que permita realizar ese pronóstico.

De todas las curvas de aproximación a una serie de datos puntuales, la que mejor ajusta es la recta de regresión muestral.

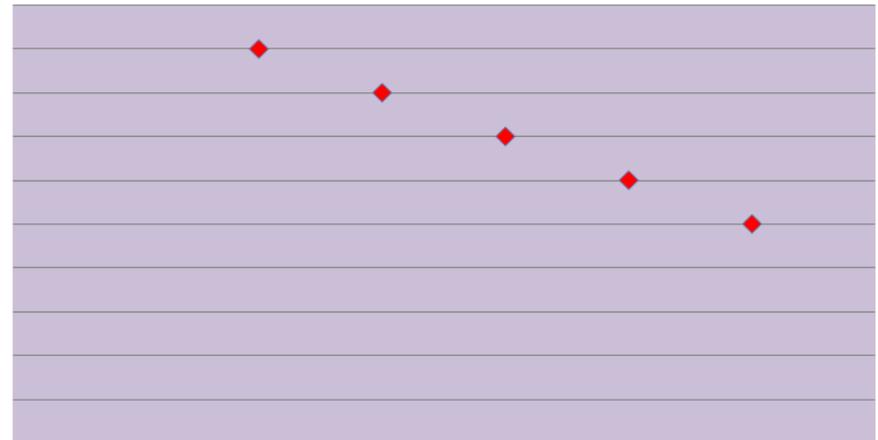
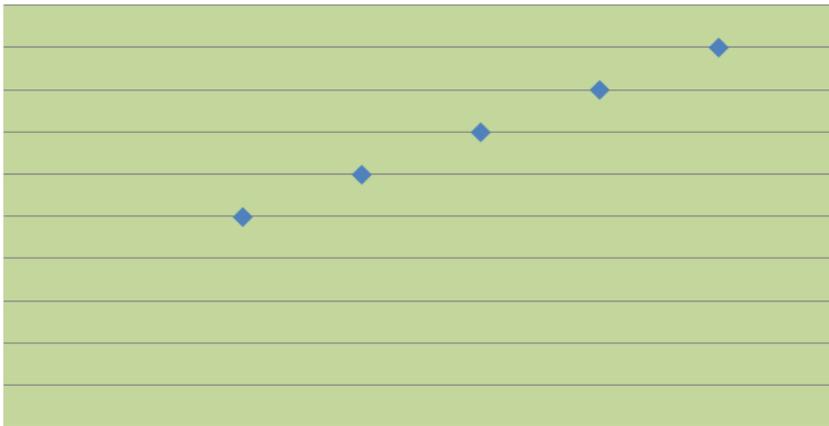
Relaciones lineales

- La relación lineal se expresa como

$$y = \alpha + \beta x$$

donde α la **ordenada al origen** y β es la **pendiente**.

- La pendiente es positiva para relaciones directas y negativa para relaciones inversas.



Recta de regresión

- Las ecuaciones para obtener los parámetros α y β son

$$\beta = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$\alpha = \bar{y} - \beta \bar{x}$$

donde \bar{x}, \bar{y} es el promedio de las x e y respectivamente.

A partir de los resultados obtenidos en la tabla para la detección del coeficiente de correlación, se determinan los parámetros de la recta de regresión.

auto	Peso	Rendimiento			
	x	y	xy	x^2	y^2
1	750	10	7500	562500	100
2	420	17.2	7224	176400	295.84
3	550	15	8250	302500	225
4	1000	9	9000	1000000	81
5	680	8	5440	462400	64
Sumas	3400	59.2	37414	2503800	765.84

$$\beta = \frac{5(37414) - (3400)(59.2)}{5(2503800) - (3400)^2} = -0.015$$

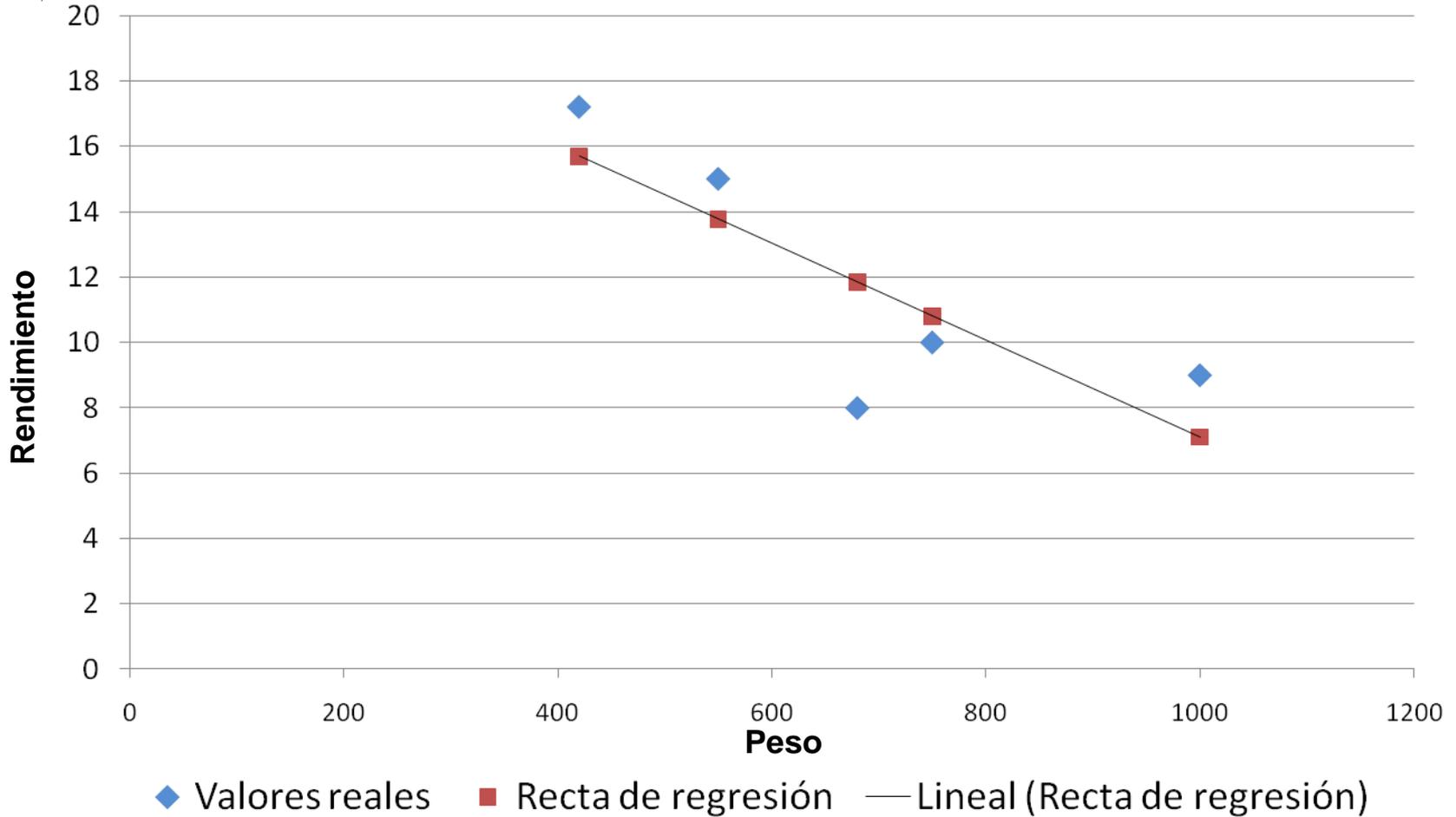
$$\alpha = (11.84) - (-0.015)(680) = 22.04$$

Por tanto, la ecuación de la recta de regresión es:

$$y = 22.04 - 0.015x$$

El valor negativo de la pendiente indica que la relación es inversa; es decir, que para cada unidad de aumento en el peso del automóvil, el rendimiento se reduce en promedio 0.015 km/l.

Recta de regresión



Aplicación del análisis de regresión a los pronósticos

El modelo de regresión se puede utilizar para hacer predicciones o previsiones sobre la variable dependiente, partiendo de un valor futuro supuesto de la variable independiente.

Sin embargo es importante hacer notar que éste modelo sólo es útil para predecir valores que se encuentren dentro del rango de los datos utilizados para hacer la predicción ya que no se cuenta con información sobre la naturaleza de la relación entre las variables fuera de este rango y se pueden cometer errores graves.

¿Cuál será el rendimiento de un automóvil al que se le adiciona un remolque de tal forma que su peso pasa de 750 kg a 980 kg?



- Utilizando la recta de regresión, se sustituye el valor nuevo de:

$$x = 980$$

$$\hat{y} = 22.04 - 0.015x$$

$$\hat{y} = 22.04 - 0.015(980)$$

$$\hat{y} = 7.34$$

Esto quiere decir, que el auto incrementó su rendimiento de 10 km/l a 7.34 km/l.

Tarea!!!