Regresión Lineal

1) Cinco niños de 2, 3, 5, 7 y 8 años de edad pesan, respectivamente, 14, 20, 32, 42 y 44 kilos.
   a. Hallar la ecuación de la recta de regresión que estime el peso en función de la edad.
   b. ¿Cuál sería el peso aproximado de un niño de seis años?

2) Un centro comercial sabe en función de la distancia, en kilómetros, a la que se sitúe de un núcleo de población, acuden los clientes, en cientos, que figura en la tabla siguiente:

Número de clientes (X)	Distancia (Y)
8	15
7	19
6	25
4	23
2	34
1	40

   a. Calcular el coeficiente de correlación lineal e interpretarlo.
   b. Si el centro comercial se sitúa a 2km, ¿cuántos clientes puede esperar?
   c. Si desea recibir a 5 clientes, ¿a qué distancia del núcleo de población debe situarse?

3) Las notas obtenidas por cinco alumnos en Matemáticas y Química son:

Matemáticas	Química
6	6.5
4	4.5
8	7
5	5
3.5	4

   a. Determinar la rectas de regresión que pone la nota de Química en función de la de Matemáticas.
   b. Calcular la nota esperada en Química para un alumno que tiene 7.5 en Matemáticas.

4) La siguiente tabla muestra el salario anual para el primer trabajo y la calificación media obtenida durante la licenciatura de 8 personas recién licenciadas escogidas al azar en una Facultad de Ciencias Económicas y Empresariales.

Calificación	5.2	5.8	6.3	7.5	7.8	8.2	8.6	8.8
Salario	1.5	1.6	1.8	2.6	2.7	3	3.1	3.8

a. Ajustar el modelo de regresión lineal que explique el salario en función de la calificación.
b. Calcular el coeficiente de determinación e interpretar el resultado.
c. Obtener un intervalo de confianza al 95% para la pendiente de la recta ajustada, e interpretar el resultado concluyendo si el coeficiente es significativo o no.

5) Sea el modelo lineal simple $Y = \beta_0 + \beta_1 X$ donde Y representa el consumo y X representa la renta familiar disponible. Una m.a.s. de 25 familias de una determinada ciudad arrojó los resultados siguientes:

$\sum X_i = 5740, \quad \sum X_i^2 = 1450400, \quad \sum Y_i = 4804, \quad \sum Y_i^2 = 1009589, \quad \sum Y_i X_i = 1209691$

a. Estime los parámetros del modelo.
b. Obtenga una medida de la bondad de ajuste del modelo.

6) La siguiente tabla muestra el consumo por cada 100km de un vehículo en distintas rutas y su velocidad media.

Consumo	11	15	19	16
Velocidad media	50	63	90	75

a. Ajuste un modelo lineal simple que explique el consumo por cada 100km en función de la velocidad media.
b. ¿Es cierto que por cada km/hora que aumenta la velocidad, el consumo se incrementa en 0.3 litros a los 100km? Usa un contraste de hipótesis al nivel de significación 5%.

Resumen de teoría y fórmulas

Ejemplos de aplicación de los modelos de regresión lineal:

- Estudiar cómo influyen las características de una vivienda en su precio.
- Predecir la tasa de paro según la edad de la persona.
- Aproximar la calificación obtenida en una materia según el número de horas de estudio semanal.
- Prever el tiempo de computación de un programa en función de la velocidad del procesador.

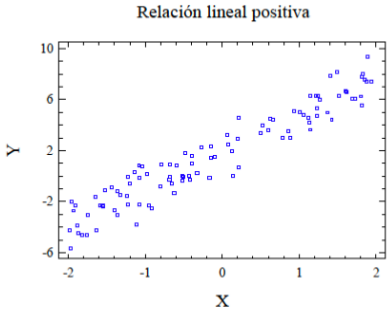
Modelo de regresión lineal simple:

$$Y = \beta_0 + \beta_1 X$$

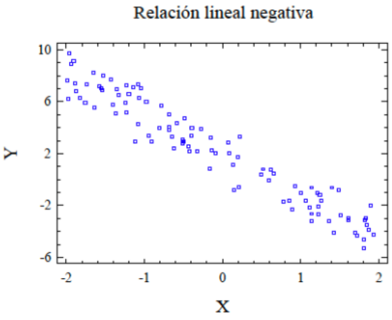
Y: Variable dependiente o explicada
X: Variable independiente o explicativa
β_0: Constante o intercepto
β_1: Pendiente

Tipo de relación lineal según el signo de la pendiente:

$\beta_1 > 0$: Relación lineal positiva



$\beta_1 < 0$: Relación lineal negativa



<u>Estadísticos que se necesitan para estimar el modelo de regresión:</u>
Media muestral de la variable independiente: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

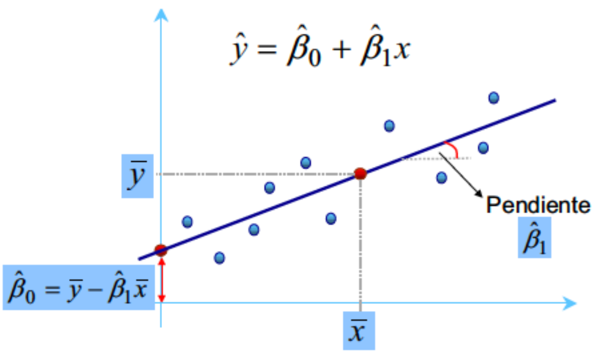
Media muestral de la variable dependiente: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
Varianza muestral de la variable independiente: $s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$
Varianza muestral de la variable dependiente: $s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2$
Covarianza entre las dos variables: $cov(x,y) = s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$
Correlación: $r_{xy} = \frac{s_{xy}}{s_x s_y}$

<u>Estimadores de mínimos cuadrados de los parámetros del modelo:</u>

$$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \, \bar{x}$$

Alternativa para calcular $\hat{\beta}_1 = r_{xy}\frac{s_y}{s_x}$



Coeficiente de determinación $R^2 = r_{xy}^2$ (coeficiente de correlación de Pearson al cuadrado): mide la variabilidad explicada por el modelo (el % de variabilidad de Y explicada por la variable X – mientras mayor sea ese % mejor).

Inferencia en el Modelo de Regresión:

Contraste de hipótesis:
$H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

Estadístico de prueba para testear la significación estadística del coeficiente $\hat{\beta}_1$:

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{s_R^2}{(n-1)s_x^2}}} \sim t_{n-2}$$

Donde $s_R^2 = \sum_{i=1}^{n} \frac{e_i^2}{n-2}$ se denomina varianza residual. Los residuos son: $e_i = y_i - \hat{y}_i$

Intervalo de confianza para β_1:

$$\hat{\beta}_1 \pm t_{n-2,\alpha/2} \sqrt{\frac{s_R^2}{(n-1)s_x^2}}$$

La longitud del intervalo disminuirá si:
- Aumenta el tamaño de la muestra
- Aumenta la varianza de las x_i
- Disminuye la varianza residual

Interpretar el resultado: Si el cero está fuera del intervalo de confianza, rechazamos la hipótesis nula.

Interpretación de los coeficientes del modelo:

$$Y = \beta_0 + \beta_1 X$$

La variable Y cambia en β_1 unidades, cuando la variable X aumenta en una unidad.
- El cambio depende del signo de β_1.
- Tener en cuenta la unidad de medida de cada variable en el problema que se tenga.