

■ Guía Explicativa: S25 - Actividad 3

Correlación y Regresión Lineal Simple

Asignatura: Estadística | **Docente:** Alexandra Cataño López

Institución: Institución Universitaria Digital de Antioquia

Equipo de Trabajo: Jorge Alejandro Rubio Giraldo, Luisa Fernanda Parra Arboleda, Maribel Román Ospina

■ Objetivo de la Actividad

Analizar la relación bivariada entre variables cuantitativas mediante modelos de **correlación de Pearson** y **regresión lineal simple** ($y = mX + b$), interpretando el porcentaje de variabilidad explicada (R^2) y realizando predicciones aplicadas a la toma de decisiones.

■ Conceptos Clave Masticaditos

1. Correlación vs. Regresión

- **Correlación (r):** Mide si existe asociación lineal entre dos variables cuantitativas y qué tan fuerte es dicha asociación. No implica relación causa-efecto directa.
- **Regresión Lineal ($y = mX + b$):** Modelo matemático que establece una recta de ajuste para **predecir** el valor de la variable dependiente (Y) en función de la variable independiente (X).

2. Coeficiente de Correlación de Pearson (r)

El valor de r se ubica siempre en el intervalo **[-1.0, +1.0]**:

- **Positivo ($r > 0$):** Relación directa. Al aumentar X, también aumenta Y.
- **Negativo ($r < 0$):** Relación inversa. Al aumentar X, disminuye Y.
- **Intensidad Fuerte:** $|r| \geq 0.70$ | **Moderada:** $0.40 \leq |r| < 0.70$ | **Débil/Nula:** $|r| < 0.40$

3. Coeficiente de Determinación (R^2) y Explicación de Variaciones

El R^2 se obtiene al elevar al cuadrado el coeficiente de correlación ($R^2 = r^2$) y expresarlo en porcentaje ($\% = R^2 \times 100$).

¿Qué representa el R^2 ?

Representa la proporción de la **variabilidad total** de la variable dependiente (Y) que es explicada por la variable independiente (X) a través del modelo de regresión lineal. El porcentaje restante ($100\% - R^2$) corresponde a factores no contemplados en el modelo o variaciones aleatorias de la vida real.

4. Componentes de la Recta de Regresión ($y = mX + b$)

- **Pendiente (m):** Indica la tasa de cambio promedio estimada en Y por cada incremento de 1 unidad adicional en X.
- **Intercepto (b):** Representa el valor base estimado de Y cuando la variable independiente es cero ($X = 0$).

5. Evaluación de Predicciones y Confiabilidad

- **Predicción:** Sustituir el valor de interés X_{eval} en la fórmula ajustada $y = m \cdot X_{eval} + b$.

- **Criterios de Confiabilidad:** Modelos con $R^2 > 60\%$ ofrecen predicciones de alta precisión. Las interpolaciones dentro del rango observado son altamente seguras.

■ Formulario Matemático y Estadístico Oficial

Formulaciones matemáticas aplicadas para el cálculo de mínimos cuadrados y coeficientes bivariados:

1. Coeficiente de Pearson (r) $r = [n \sum XY - (\sum X)(\sum Y)] / \sqrt{ [n \sum X^2 - (\sum X)^2] [n \sum Y^2 - (\sum Y)^2] }$
2. Coeficiente de Determinación (R^2) $R^2 = (r)^2 \implies R^2(\%) = R^2 \times 100\%$
3. Pendiente de la Recta (m) $m = [n \sum XY - (\sum X)(\sum Y)] / [n \sum X^2 - (\sum X)^2]$
4. Intercepto de la Recta (b) $b = Y_{\text{■}} - m X_{\text{■}} = [\sum Y - m \sum X] / n$
5. Ecuación de la Recta Estimada $Y_{\text{■}} = mX + b$
6. Modelo de Predicción en Tiempo Real $Y_{\text{■_pred}} = m \cdot X_{\text{eval}} + b$

■ Resumen de Variables Cuantitativas Seleccionadas por Base de Datos

En cada base de datos se seleccionaron **3 variables cuantitativas** para analizar **2 modelos de correlación y regresión**:

Base de Datos	3 Variables Cuantitativas	Modelos de Regresión Analizados
DB 1: Intoxicaciones (250 Casos)	1. Dosis Estimada (mL) 2. Días Estancia 3. Tiempo Atención (hrs)	Mod 1: Dosis (X1) vs Días Estancia (Y1) $r = 0.1566, R^2 = 2.45\%$ Mod 2: Tiempo Atención (X2) vs Días Estancia (Y2) $r = -0.0008, R^2 = 0.00\%$
DB 2: Estudiantes (60 Alumnos)	1. Nota 1 2. Promedio Final 3. Nota 4	Mod 1: Nota 1 (X1) vs Promedio Final (Y1) $r = 0.8412, R^2 = 70.76\%$ Mod 2: Nota 4 (X2) vs Promedio Final (Y2) $r = 0.7950, R^2 = 63.20\%$

DB 3: Clientes (500 Registros)	1. Inversión Ad (\$ USD) 2. Ventas Totales (\$ USD) 3. Visitas Web	Mod 1: Inversión Ad (X1) vs Ventas (Y1) $r = 0.9450, R^2 = 89.30\%$ Mod 2: Visitas Web (X2) vs Ventas (Y2) $r = 0.7895, R^2 = 62.33\%$
--	--	---