TECNOLÓGICA DEL ORIENTE
INSTITUCIÓN DE EDUCACIÓN SUPERIOR

ESTADÍSTICA DESCRIPTIVA-APLICADA

CONCEPTO DE REGRESIÓN LINEAL

CONCEPTO DE REGRESIÓN LINEAL

La regresión lineal se basa en la suposición de que existe una relación lineal entre dos variables, lo que significa que un cambio en una variable, está asociado con un cambio proporcional en la otra variable. Esta relación puede ser positiva (cuando una variable aumenta, la otra también lo hace) o negativa (cuando una variable aumenta, la otra disminuye).

Imagine que está estudiando la relación entre las horas de estudio (variable independiente, x) y las calificaciones obtenidas en un examen (variable dependiente, y). Intuitivamente, esperaría que más horas de estudio se asocien con calificaciones más altas. Si representa cada estudiante como un punto en un gráfico, con sus horas de estudio en el eje x y su calificación en el eje y, es probable que observe una tendencia: los puntos se alinean aproximadamente a lo largo de una línea recta, con pendiente positiva. Esta es la idea central de la regresión lineal.

En un contexto empresarial, la regresión lineal es una herramienta valiosa para entender cómo ciertos factores influyen en variables de interés. Por ejemplo, un gerente de ventas podría usar regresión lineal para modelar cómo el gasto en publicidad (variable independiente) afecta los ingresos por ventas (variable dependiente). Encontrar la línea de mejor ajuste permitiría predecir las ventas esperadas para diferentes niveles de inversión publicitaria, información clave para la toma de decisiones y la planificación presupuestaria.

Modelo de regresión lineal simple

El modelo de regresión lineal simple, se expresa mediante la ecuación de una línea recta:

$$y = \beta 0 + \beta 1 * x + \varepsilon$$

Donde:

y es la variable dependiente.

x es la variable independiente.

$\beta 0$ es la ordenada al origen (el valor de y cuando $x = 0$).

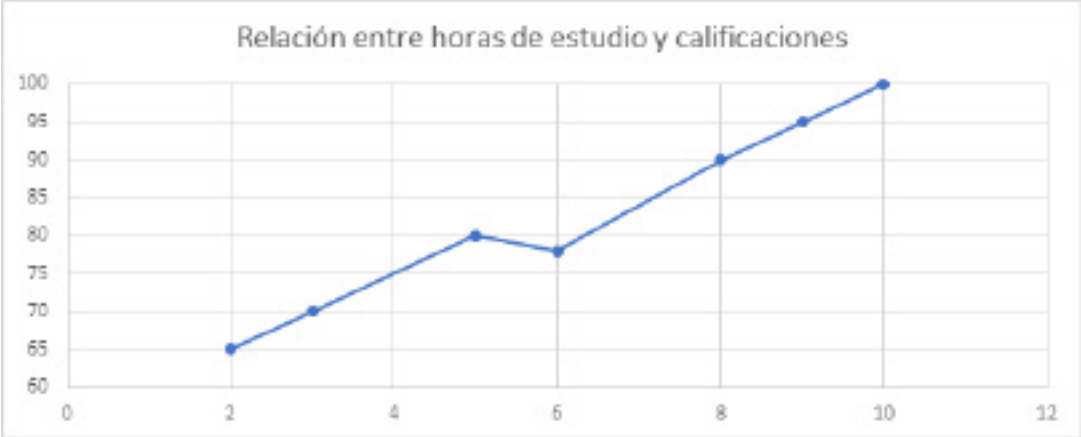
$\beta 1$ es la pendiente de la recta (cuánto cambia y por cada unidad que aumenta x).

ε es el término de error (la desviación de cada observación respecto a la línea).

Para entender mejor cada componente, consideremos el ejemplo de las horas de estudio y las calificaciones:

Horas de estudio (X)	Calificaciones (Y)
2	65
3	70
5	80
6	78
8	90
9	95
10	100

Figura 1. Relación entre horas de estudio y calificación



Ahora, imaginemos que encontramos la línea de mejor ajuste para estos datos y tiene la ecuación:

Calificación = 50 + 5*horas_estudio

En este caso:

- La variable dependiente (y), son las calificaciones.
- La variable independiente (x), son las horas de estudio.
- La ordenada al origen (β0) es 50, lo que sugiere que un estudiante que no estudia en absoluto, esperaría una calificación de 50.
- La pendiente (β1) es 5, lo que indica que, por cada hora adicional de estudio, la calificación esperada aumenta en 5 puntos.
- El término de error (ε) representaría cuánto se desvía la calificación real de cada estudiante de la predicha por el modelo.

Es importante tener en cuenta que este es un modelo simplificado. En realidad, muchos otros factores además de las horas de estudio, podrían influir en las calificaciones (como la aptitud del estudiante, la calidad de la enseñanza, etc.). Estos factores se incorporan colectivamente en el término de error.

En resumen, el modelo de regresión lineal simple, busca la línea recta que mejor capture la relación entre dos variables. La ordenada al origen ($\beta 0$) y la pendiente ($\beta 1$), definen esta línea, y el término de error (ε), representa la variabilidad no explicada por el modelo. A continuación, veremos cómo se estiman los coeficientes $\beta 0$ y $\beta 1$, a partir de los datos, usando el método de mínimos cuadrados.