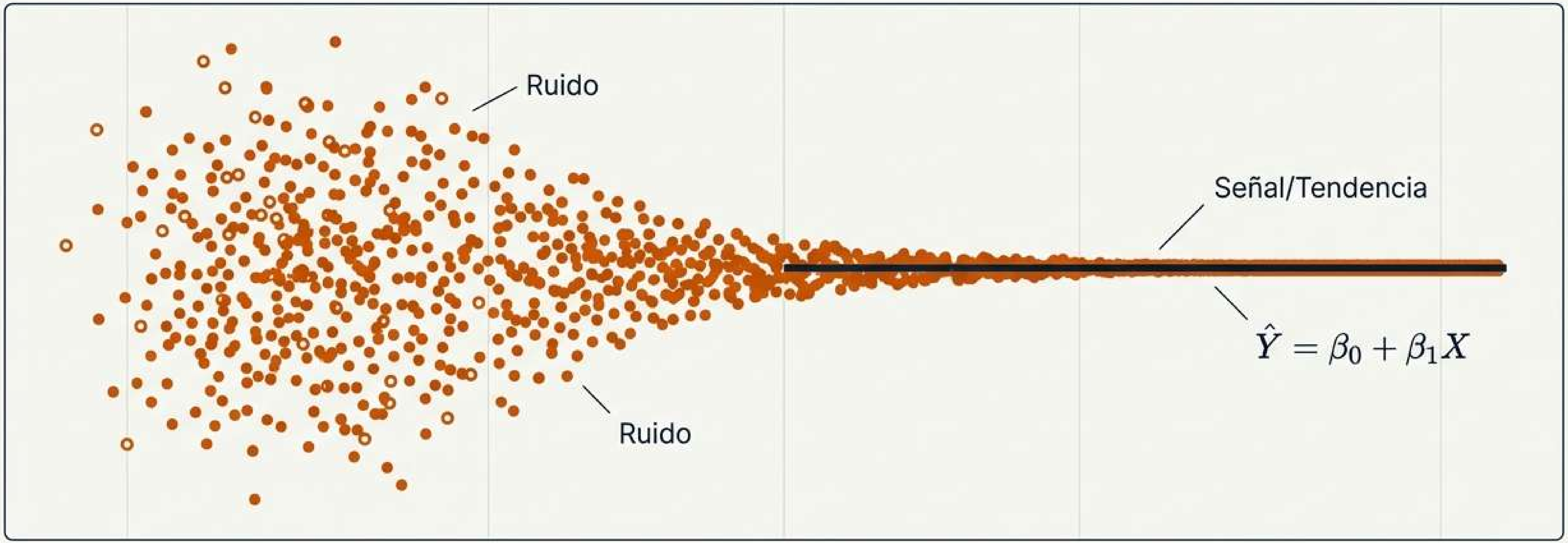
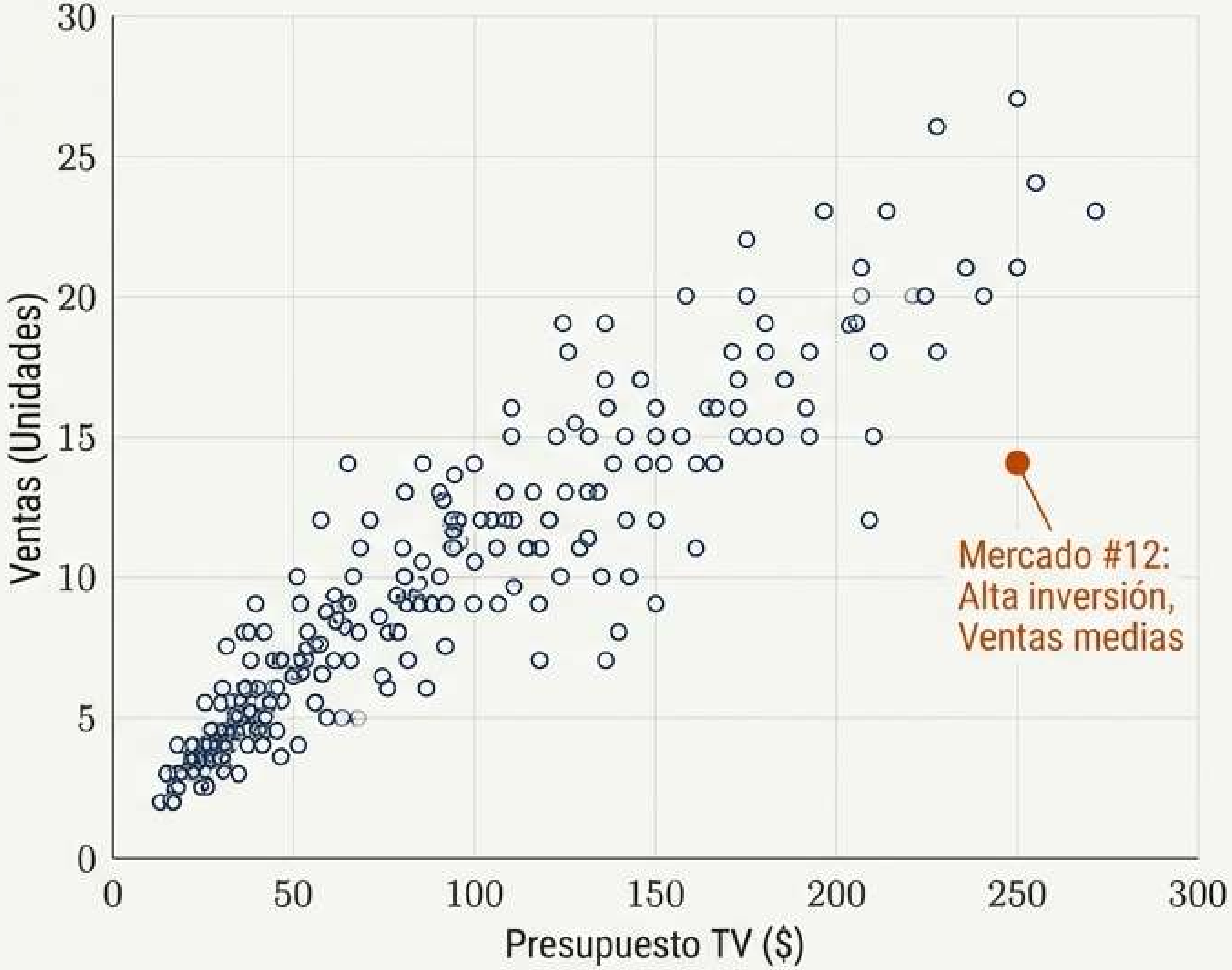
Más allá del Promedio: Entendiendo la Regresión Lineal Simple

De la intuición gráfica a la precisión matemática: una guía para descifrar el ruido.



ESTRATEGIA DE ANÁLISIS DE DATOS | [NOMBRE DEL PRESENTADOR]

La Pregunta de Negocio: ¿La publicidad realmente impulsa las ventas?



Analizamos 200 mercados distintos. El objetivo es cuantificar el retorno de inversión.

Las 7 Preguntas del Detective:

1. ¿Cuál es la relación entre el presupuesto y las ventas?
2. ¿Existen otros factores que influyen?
3. ¿Cómo se comportan las ventas sin inversión?
4. **¿Con qué precisión podemos predecir las ventas futuras?**
5. ¿Cuál es el punto de saturación de la publicidad?
6. ¿Existen diferencias entre los mercados?
7. ¿Qué riesgos debemos considerar?

La Interrogante Principal: Si invertimos $1,000 adicionales en TV, ¿cuántas unidades exactas venderemos?

El Enfoque Detectivesco: Exploración de Datos

Antes de modelar, medimos la 'pista' inicial.

Covarianza (S_{xy})

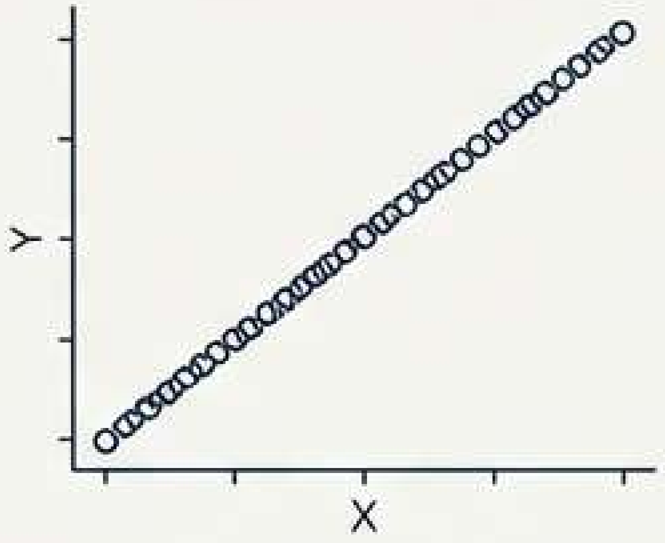
Indica la dirección.
Positiva = crecen juntas.
Negativa = inversa.
Defecto: Depende de las unidades de medida.

Correlación de Pearson (r_{xy})

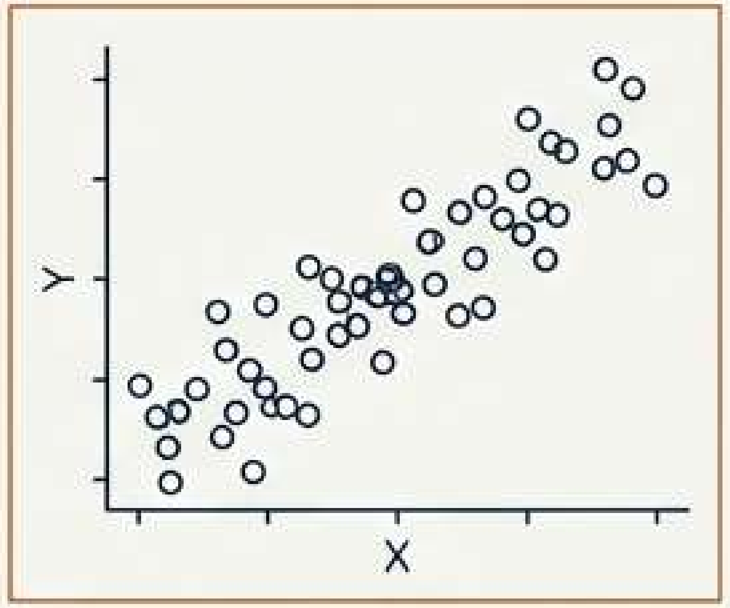
La medida estandarizada.
Elimina el efecto de escala.

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

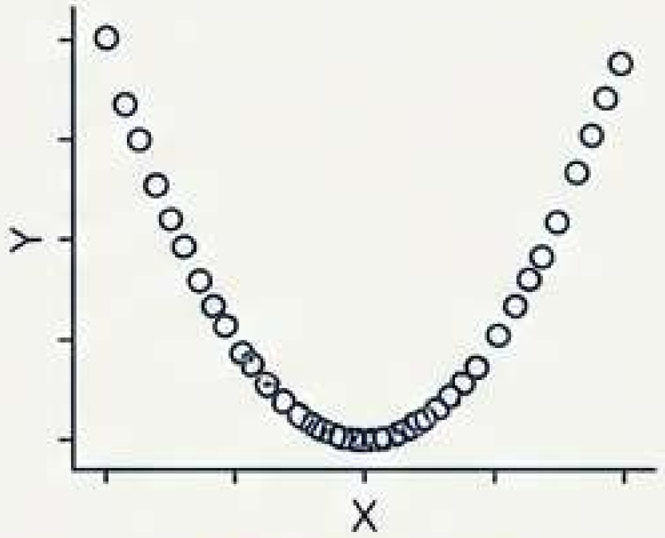
Reconocimiento de Patrones: Ver antes de Calcular



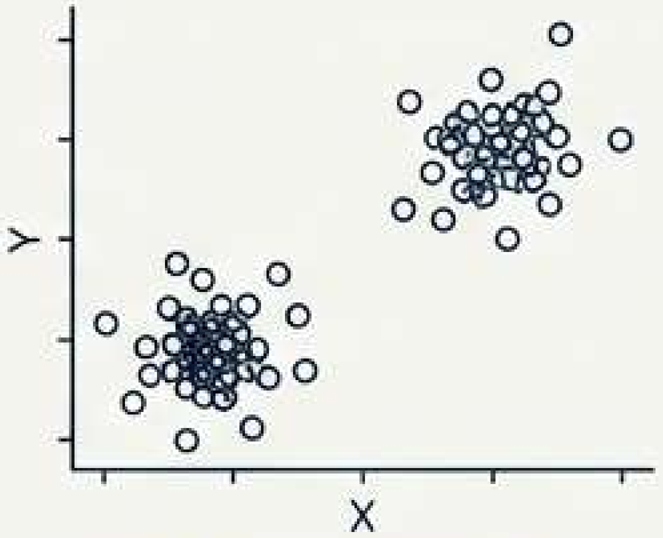
1. Determinístico
(Irreal)



2. Lineal Ruidoso
(Regresión Simple)



3. No Lineal
(Requiere Curvas)



4. Segmentado
(Requiere Clasificación)

La regresión lineal asume que una línea recta es la mejor descripción de la realidad.
Si los datos dicen 'curva', la línea miente.

La Anatomía del Modelo

Variable Respuesta (Ventas).
Lo que predecimos.

$$\underbrace{Y}_{} = \underbrace{\beta_0 + \beta_1 X}_{} + \epsilon$$

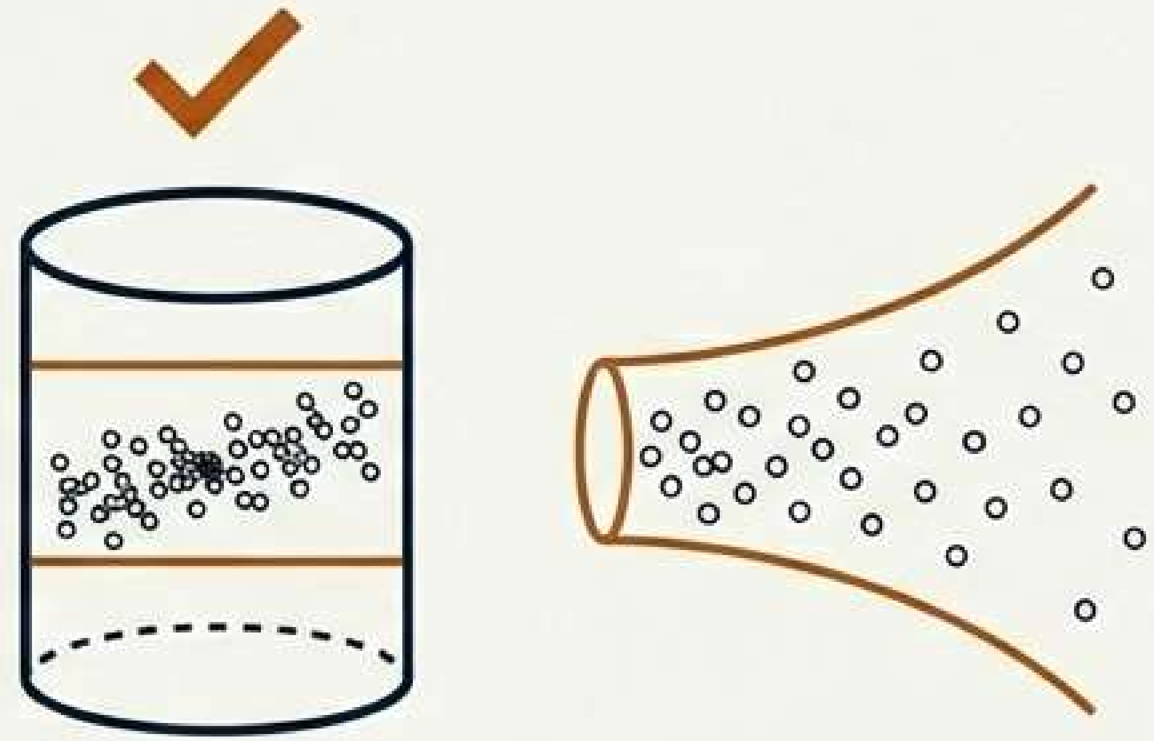
La Señal (Componente Determinístico).
Parámetros constantes desconocidos.

El Ruido (Error (Error Aleatorio).
Incertidumbre y variables ocultas.

El objetivo es encontrar los valores de β_0 y β_1 que separan la señal del ruido.

Las Reglas del Juego: Supuestos Fundamentales

Para que el modelo sea válido, el error (ϵ) debe comportarse de forma predecible.



1. Media Cero.

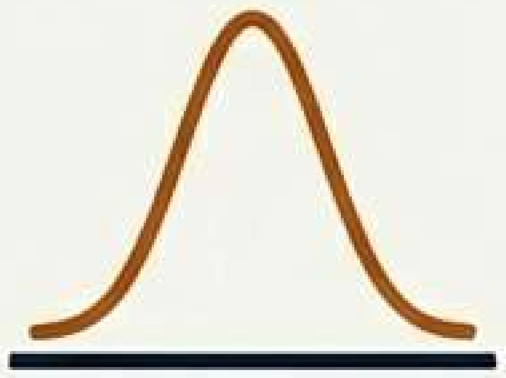
$E(\epsilon) = 0.$

Los errores no tienen sesgo; flotan equilibrados alrededor de la línea.

2. Homocedasticidad.

Varianza constante (σ^2).

El error no crece cuando X crece.



3. Normalidad.

$\epsilon \sim N(0, \sigma^2).$

El ruido sigue una distribución normal.

4. Independencia.

El error de hoy no predice el error de mañana.

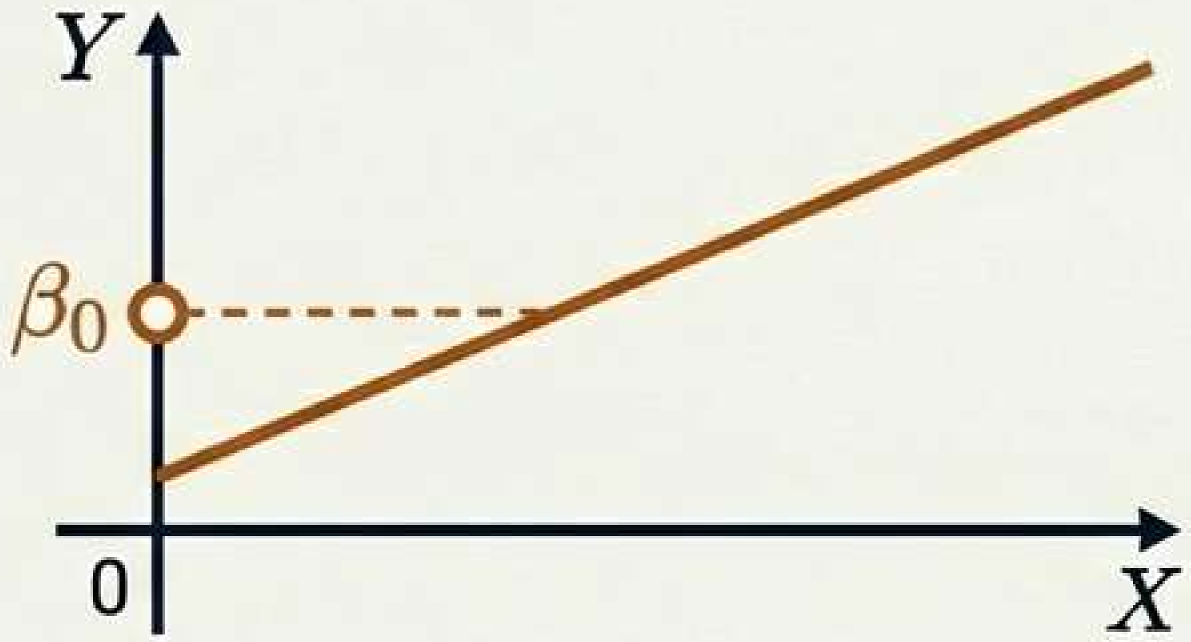
Interpretando la Realidad: ¿Qué significan los parámetros?

β_0

El Intercepto / Punto de Partida

Valor esperado de Y cuando $X = 0$.

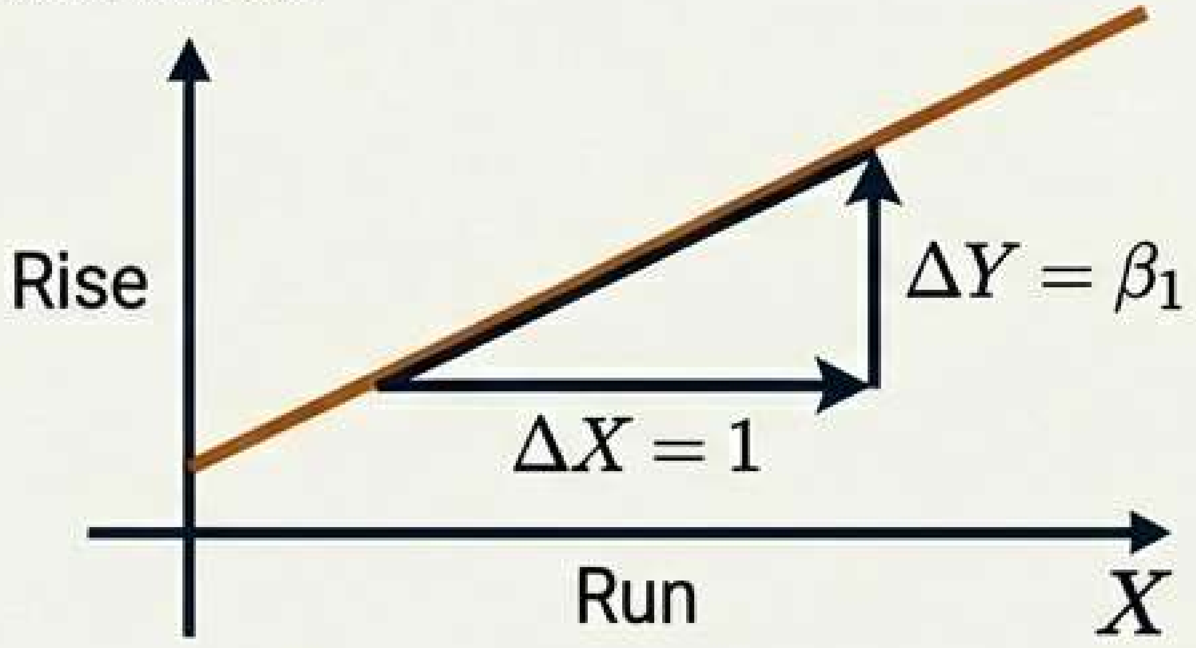
Ejemplo: Ventas base sin inversión en publicidad. (A veces es teórico: ¿Estatura con edad 0?).



β_1

La Pendiente / Tasa de Cambio

Cambio promedio en Y por cada unidad extra de X.



Positiva (+): Inversión sube, Ventas suben.

El Reto de la Estimación

¿Cómo encontramos la “línea verdadera” entre infinitas posibilidades?

Datos de la Muestra (x, y)

Estimación Heurística

El Ojímetro. Subjetivo.

Máxima Verosimilitud (MLE)

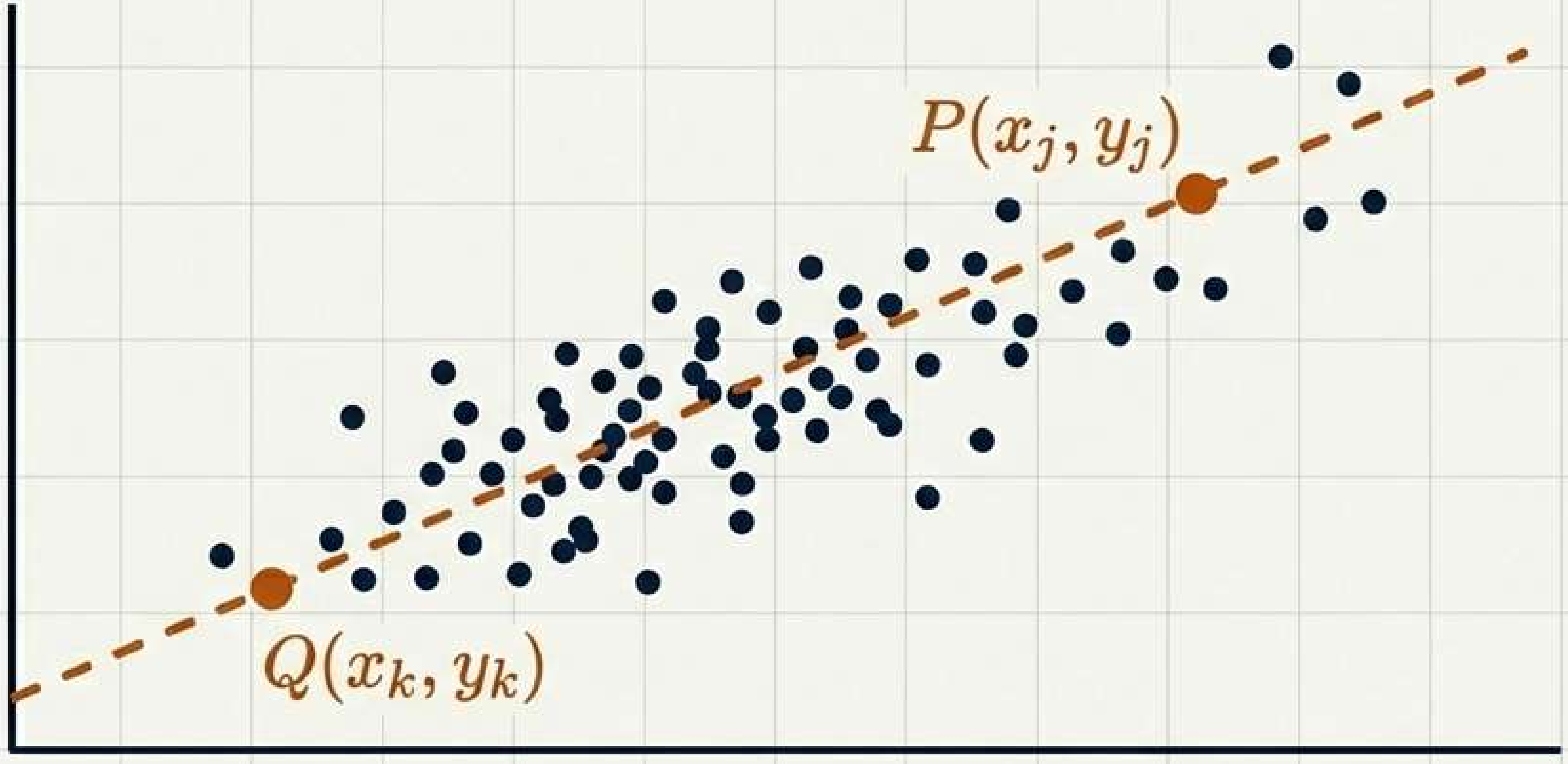
Probabilístico. Maximizar chance de observar los datos.

Mínimos Cuadrados (OLS)

Geométrico. Minimizar distancia física.

Exploraremos los tres caminos.

Método 1: Estimación Heurística (El "Ojímetro")



Seleccionamos dos puntos representativos.

Calculamos la pendiente manualmente:

$$\tilde{\beta}_1 = \frac{y_j - y_k}{x_j - x_k}$$

**Veredicto: Rápido e intuitivo, pero inestable. Depende subjetivamente de qué puntos elija el observador.

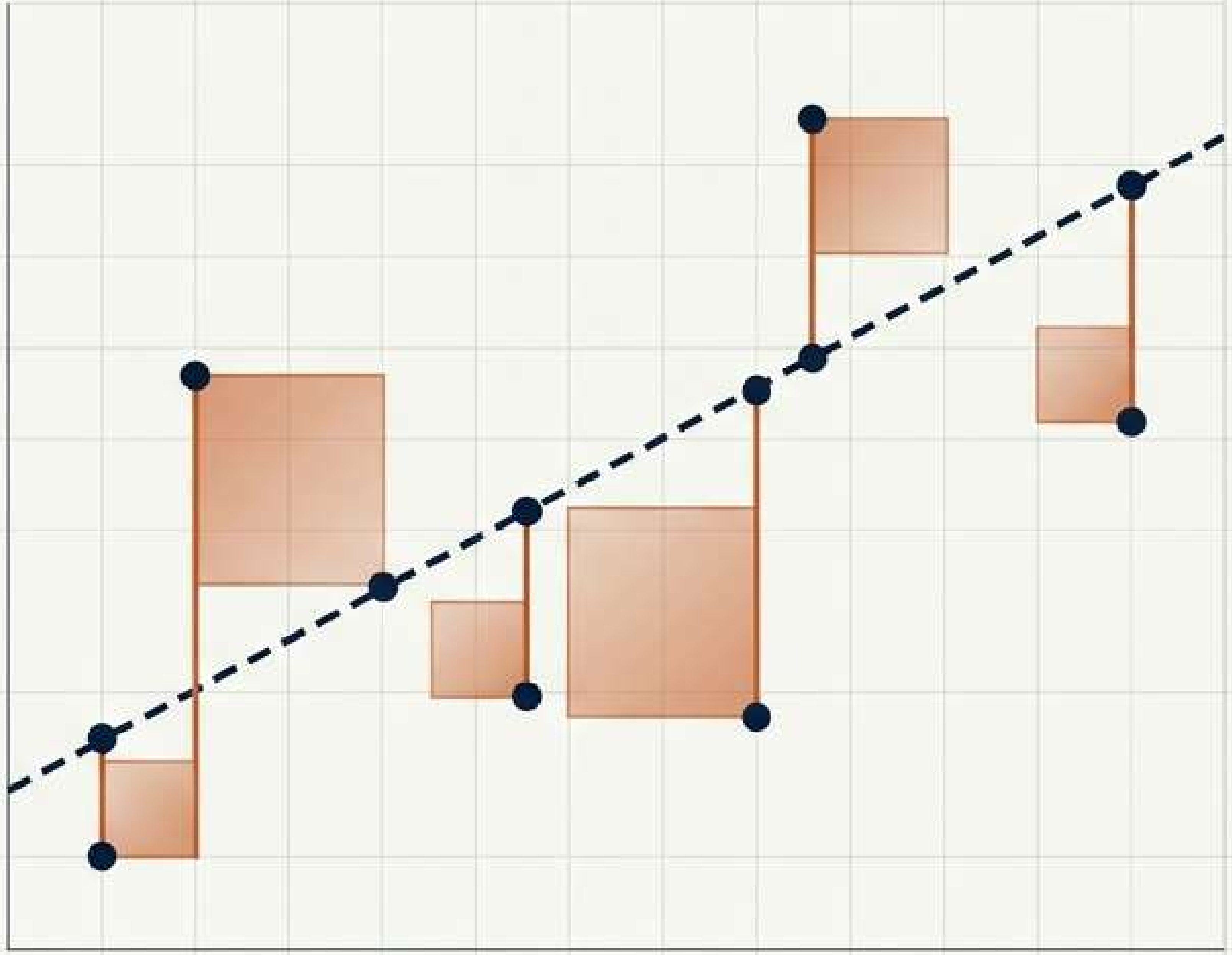
Método 2: Máxima Verosimilitud (Probabilístico)

Buscamos los parámetros (β) que hacen matemáticamente más probable haber observavo estos datos específicos.

$$l(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

1. Derivamos parcialmente respecto a β_0 y β_1.
2. Igualamos a cero para encontrar el máximo.
3. Resultado: Genera las **Ecuaciones Normales**.

Método 3: Mínimos Cuadrados Ordinarios (Geométrico)



Goal: Minimizar la Suma de los Residuos al Cuadrado (**RSS**).

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Why Squares?

- Penaliza severamente los errores grandes.
- Evita que errores positivos y negativos se cancelen.

La Convergencia: El Veredicto Matemático

Bajo el supuesto de normalidad, Máxima Verosimilitud y Mínimos Cuadrados producen **exactamente el mismo resultado**.

Pendiente ($\hat{\beta}_1$): $$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = r_{xy}\frac{S_y}{S_x}$$

Covarianza normalizada por varianza de X

Intercepto ($\hat{\beta}_0$): $$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

La línea pivota sobre el promedio $(\bar{x}, \bar{y})$

Aplicación: Resultados del Caso Advertising

Datos: 200 Mercados (TV vs Ventas)

Cálculo de Parámetros:

$$\hat{\beta}_0 = 7.03$$
$$\hat{\beta}_1 = 0.0475$$

$$Ventas = 7.03 + 0.0475 \times TV\ Budget$$

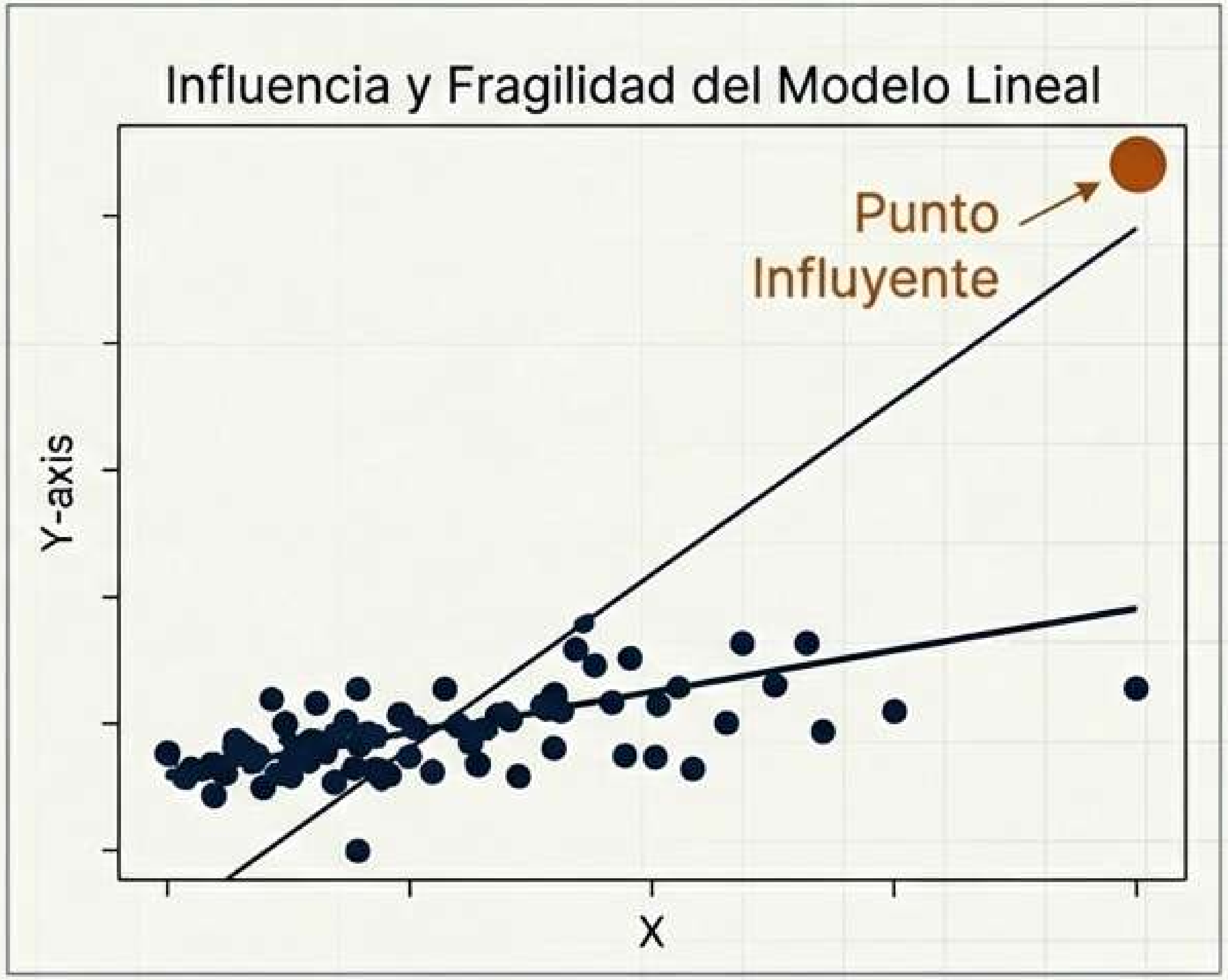
Interpretación

El Intercepto:
Sin publicidad, vendemos **7,030** unidades base.

La Pendiente:
Por cada **$1,000** extra en TV, vendemos **47.5** unidades adicionales.

Diagnóstico y Cautela: No confíe ciegamente

1. **Extrapolación:** Peligroso predecir fuera del rango observavo. No sabemos si la línea continúa.
2. **Datos Influyentes:** Un solo 'outlier' puede distorsionar toda la pendiente (ver gráfico).
3. **Análisis de Residuos**: Verifique los errores. Si ve patrones (curvas, embudos), el modelo lineal no es suficiente.



NotebookLM

Resumen Ejecutivo:
El Poder de la Simplificación

Explorar

Visualizar nube de puntos y correlación.

Modelar

$Y = \beta X + \epsilon$

Definir ecuación $Y = \beta X + \epsilon$.

Estimar

Minimizar cuadrados para hallar la línea.

Interpretar

$\beta \longrightarrow$ Impacto en Ventas

Traducir β a impacto en Ventas.

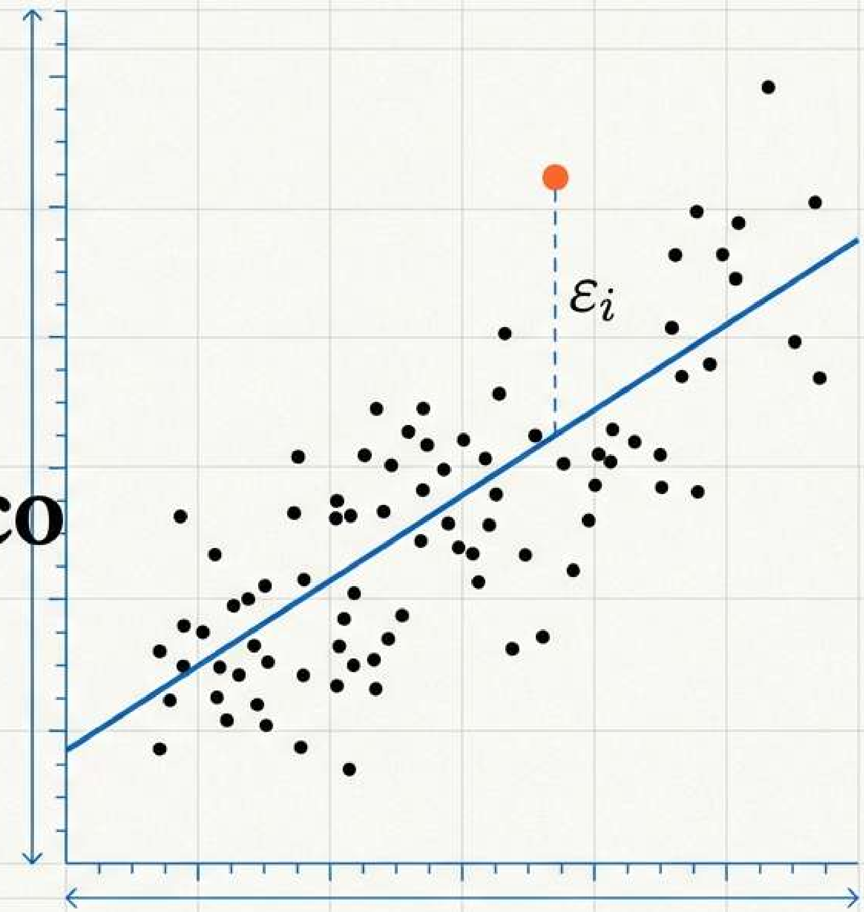
Los modelos lineales son herramientas poderosas no porque sean perfectos, sino porque simplifican la complejidad para permitir la toma de decisiones.

ESTADÍSTICA COMPUTACIONAL AVANZADA

Editorial Mathematics

Modelos de Regresión Simple: Estimación y Diagnóstico

De la intuición geométrica a la formulación matricial.



El Objetivo de la Predicción

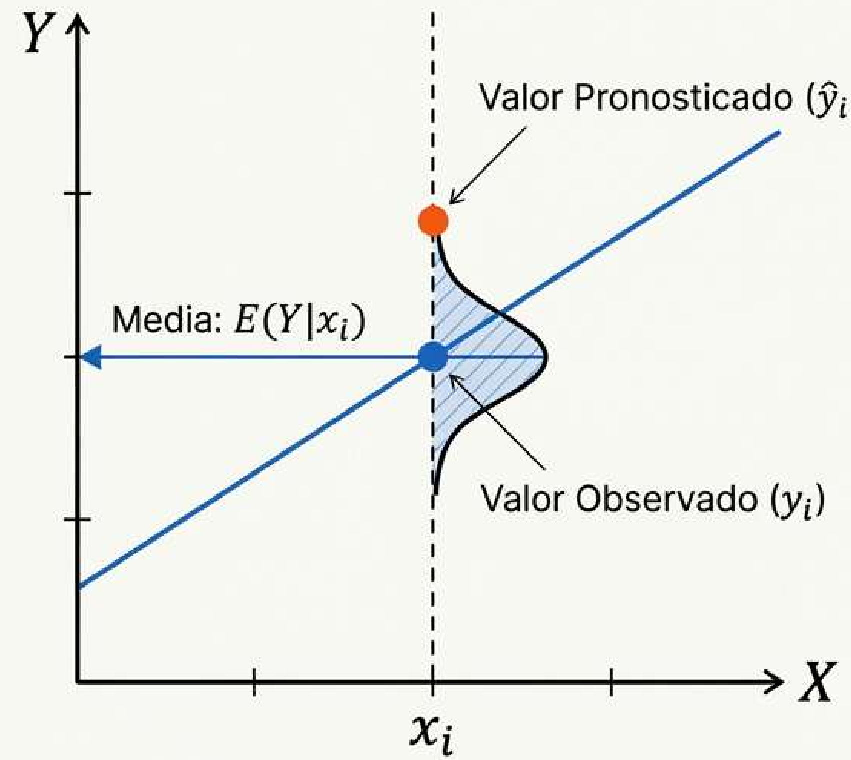
Buscamos un estimador puntual de la media de Y para un x fijo.

La ecuación de predicción:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Forma centrada:

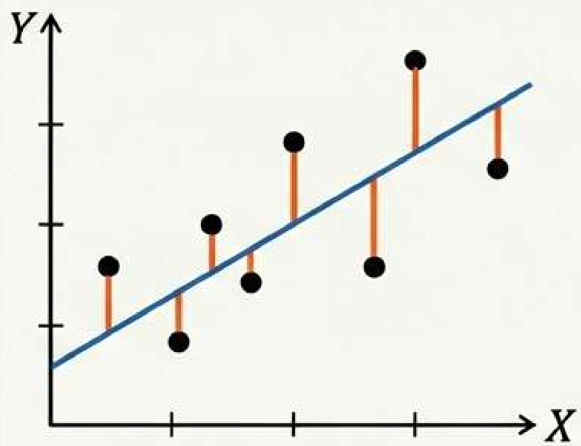
$$\hat{y}_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x})$$



NotebookLM

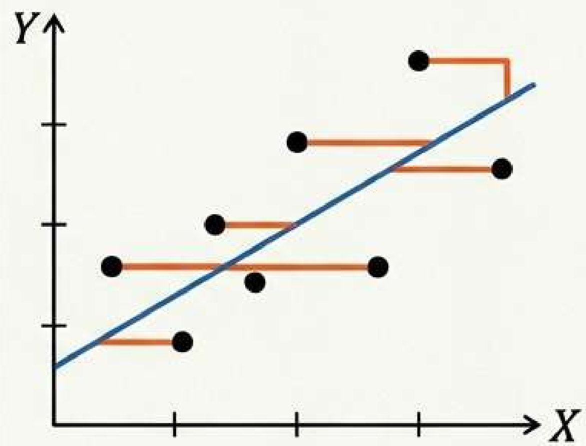
Geometría del Error: ¿Qué distancia minimizamos?

Distancia Vertical
(Regresión Clásica)



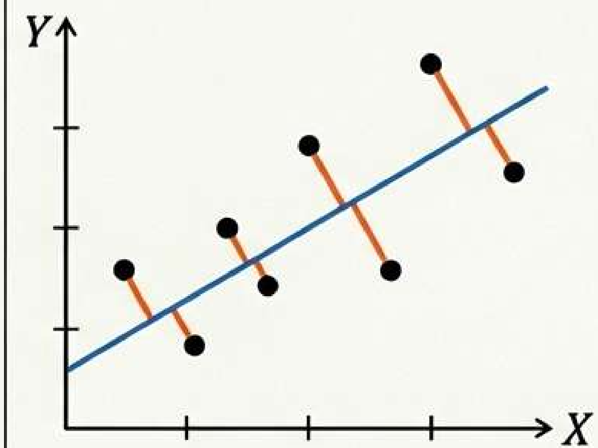
Minimiza error en Y.

Distancia Horizontal
(Regresión Inversa)



Minimiza error en X.

Distancia Perpendicular
(Ortogonal)



Componentes Principales.

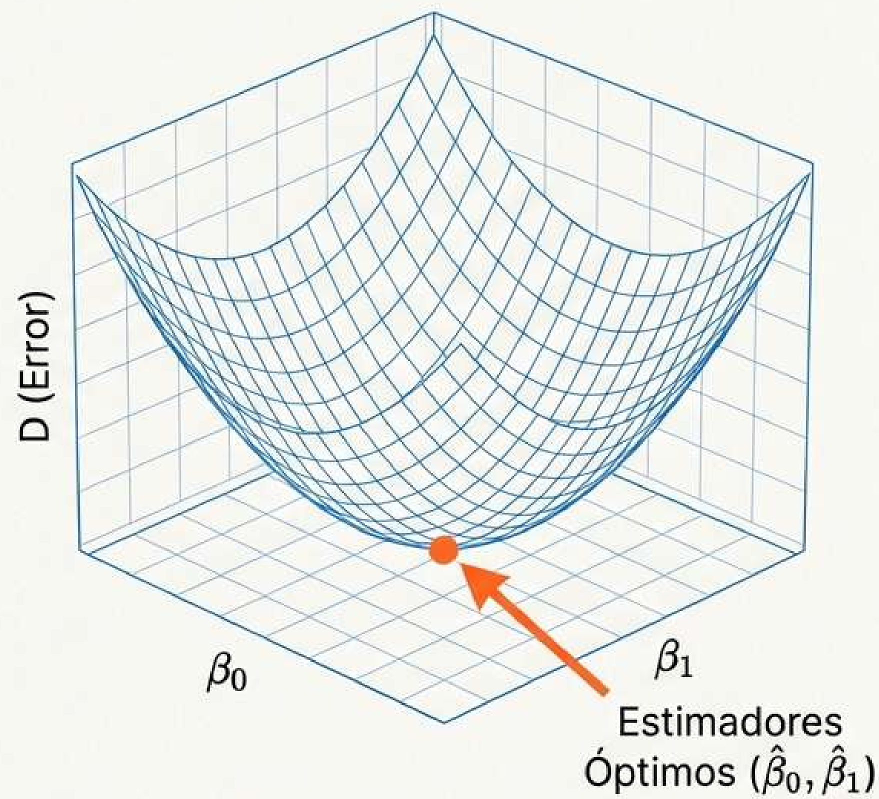
Para estimar la media de Y dado x, el método de Mínimos Cuadrados Ordinarios (MCO) utiliza exclusivamente la **Distancia Vertical**.

El Método de Mínimos Cuadrados Ordinarios (MCO)

Función a minimizar (D):

$$D(\beta_0, \beta_1) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Elevamos al cuadrado para penalizar las grandes desviaciones y evitar que los errores positivos y negativos se cancelen.



La Derivación de los Estimadores

1. Derivadas Parciales = 0

$$\frac{\partial D}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial D}{\partial \beta_1} = -2\sum_{i=1}^{n}x_i(y_i - \beta_0 - \beta_1 x_i) = 0$$

La matriz Hessiana es definida positiva, asegurando un mínimo global.

2. Ecuaciones Normales

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum y_i$$
$$\hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i$$

3. Solución Escalar

$$\boxed{\hat{\beta}_1 = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}}$$

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

Traducción al Lenguaje Matricial

Anatomy of the Equation

$$Y = X\beta + \epsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$n \times 1$
(Respuesta)

Intercepto →

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$n \times 2$
(Matriz de Diseño)

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

2×1
(Parámetros)

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$n \times 1$
(Errores)

La Solución Elegante

Minimizar la suma de cuadrados en forma matricial:

$$D(\beta) = (Y - X\beta)'(Y - X\beta)$$

Derivando respecto a β:

$$-2X'Y + 2X'X\beta = 0$$

Ecuaciones Normales:

$$X'X\beta = X'Y$$

$$\boxed{\hat{\beta} = (X'X)^{-1}X'Y}$$

Condición: La matriz (***X'X***) debe ser no singular (invertible).

Coeficiente de Determinación (R^2)

Variabilidad Total (SCT)

$$SCT = SCM + SCE$$

Variabilidad Explicada (SCM) Ruido / Error (SCE)

$$R^2 = \frac{SCM}{SCT} = 1 - \frac{SCE}{SCT}$$

Proporción de la variabilidad de Y explicada por el modelo lineal. $0 \leq R^2 \leq 1$.

R^2 Ajustado vs. Estándar

R^2 Estándar

El problema: Nunca disminuye al agregar variables, aunque sean irrelevantes.

Estatura Hijo-Padre
R^2 = **0.9415**

R^2 Ajustado

La solución: Penaliza la complejidad (número de parámetros p).

$$R^2_{adj} = 1 - \frac{SCE/(n-p)}{SCT/(n-1)}$$

Estatura Hijo-Padre
R^2_{adj} = **0.9332**

El ajuste corrige los grados de libertad. Si el modelo es parsimonioso, ambos valores deben ser cercanos.

Análisis de los Residuos: La Huella del Modelo

Residuo: $e_i = y_i - \hat{y}_i$

Suma Cero

$$\sum e_i = 0$$

Ortogonalidad

$$\sum e_i x_i = 0$$

Key Metric (Varianza Residual)

$$\hat{\sigma}_R^2 = \frac{\sum e_i^2}{n - 2}$$

Grados de libertad (perdemos **2** al estimar β_0 y β_1).