Universidad de los Andes

Educación Continua

Vicerrectoría Académica

Regresión lineal

- ¿Qué es un modelo?
- ¿Qué es un modelo lineal?
- Regresión lineal

En la clase de hoy

1. Entenderemos los conceptos básicos de regresión
2. Exploraremos cómo se ajusta una regresión lineal

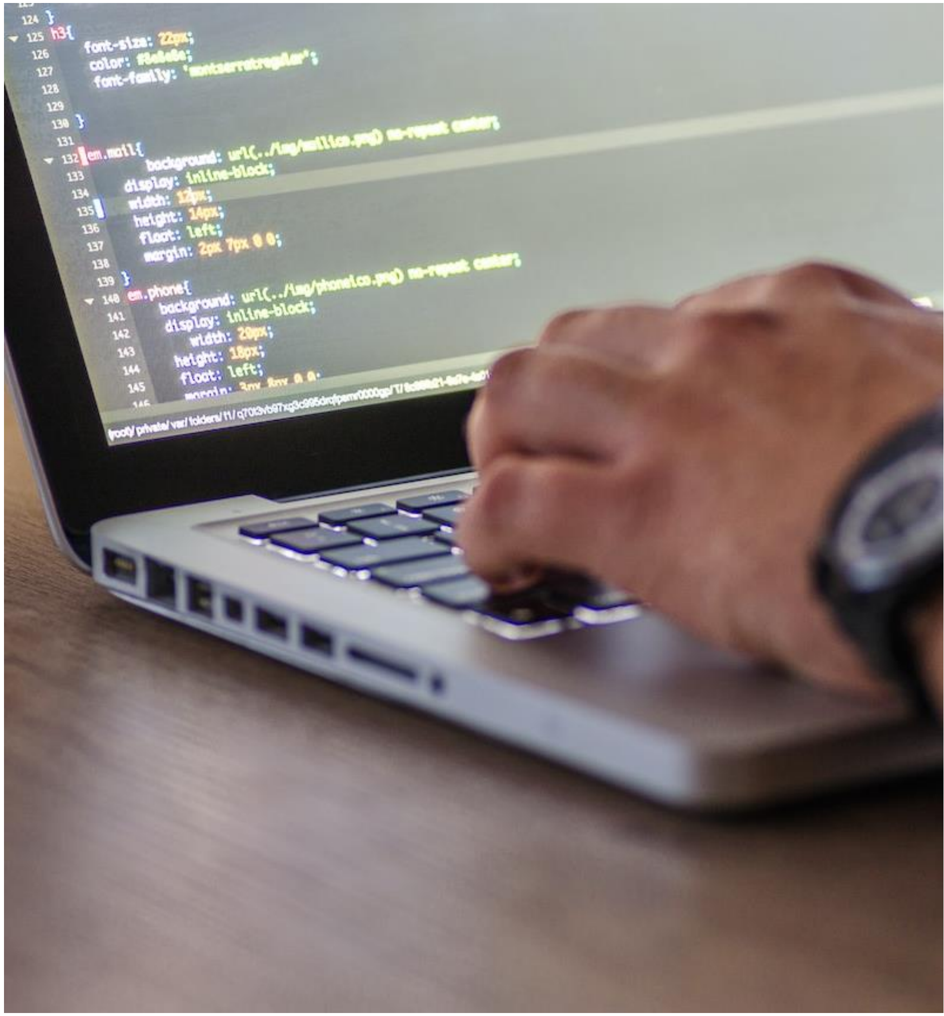


Photo by Lukas

Universidad de los Andes

¿QUÉ ES UN MODELO?

Todos los modelos son malos, pero algunos son útiles

George E.P. Box

Qué es un modelo

Para discutir (3 minutos)

- ¿Cómo definirían un modelo?
- ¿Algunos ejemplos de modelos?

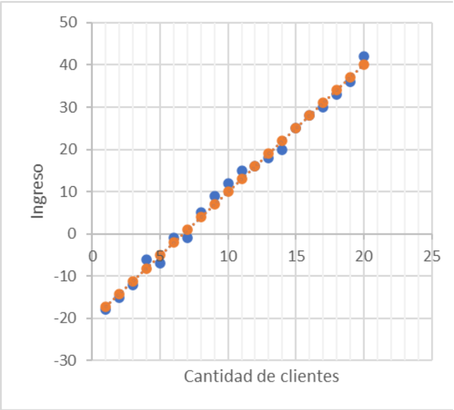
Qué es un modelo

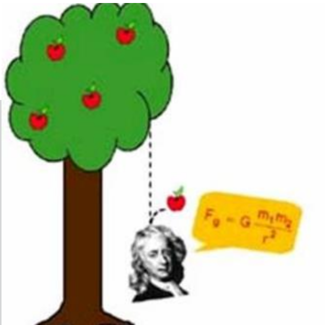
Modelo: “Arquetipo o punto de referencia para imitarlo o reproducirlo.” RAE

Nuestra definición: Representación simplificada de un sistema

¿Algunos ejemplos de modelos?

- Un mapa
- Una miniatura
- Ley de gravedad de Newton





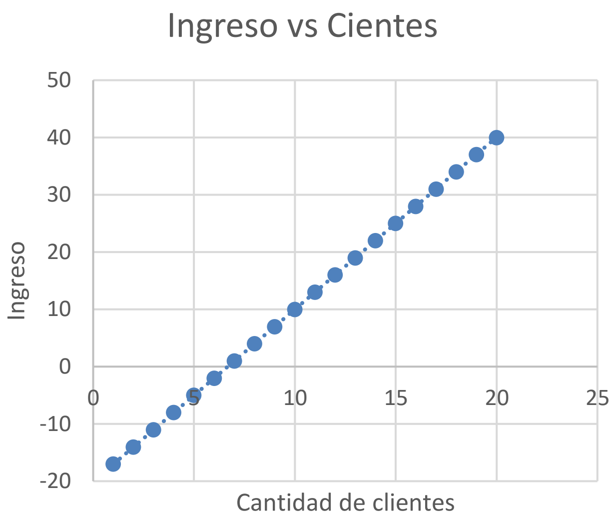
¿QUÉ ES UN MODELO LINEAL?

Qué es un modelo lineal

Determinístico - Es un modelo de la forma:

$$\boldsymbol{Y = B_0 + B_1X_1 + B_2X_2 + \cdots + B_nX_n}$$

Ejemplo: El ingreso generado por un servicio de suscripción dependerá del número de clientes $\boldsymbol{X_1}$, el valor de la suscripción $\boldsymbol{B_1}$.



Ejercicio en Excel: 1

Represente el modelo de utilidad de una compañía de servicios de suscripción, en la forma

$\boldsymbol{Y = B_0 + B_1 X_1}$ donde:

- Se genera un ingreso de 5 USD por cada cliente.
- Se genera un costo de 2 USD por cada cliente.
- Se genera un costo fijo de 20 USD independiente de la cantidad de clientes

Calcule en Excel la utilidad (Y) para cantidad de clientes (X) de 1 a 20. **5 Minutos.**

Qué es un modelo lineal

No determinístico - Es un modelo de la forma:

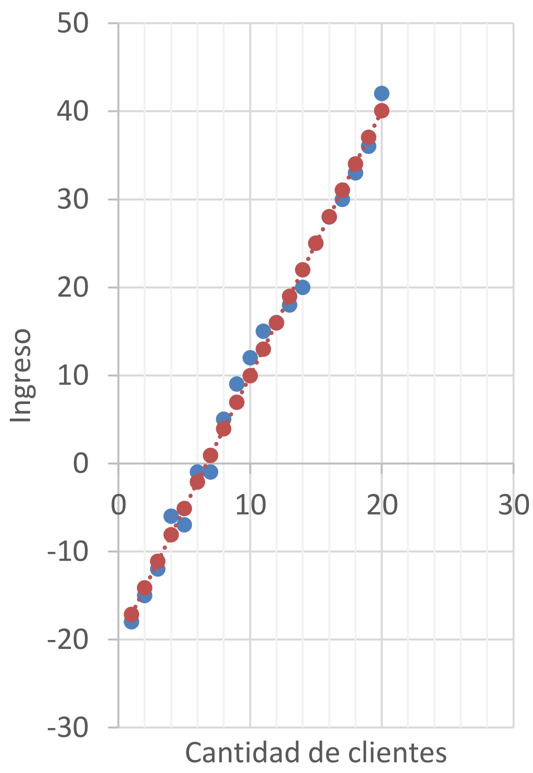
$$Y = B_0 + B_1X_1 + B_2X_2 + \cdots + B_nX_n + \boldsymbol{\varepsilon}$$

Donde $\boldsymbol{\varepsilon}$ se llama error y usualmente asumimos que:

- Es una variable aleatoria con una distribución conocida (usualmente normal)
- Tiene valor esperado cero E($\boldsymbol{\varepsilon}$) = 0

Ejemplo: Ventas del día en una tienda con la cantidad de clientes que visitan una tienda (tráfico)

Aún si conoces el ticket promedio y la cantidad de visitas en un día, es difícil *adivinar* exactamente cuáles fueron las ventas.



¿QUÉ ES UN MODELO DE REGRESIÓN LINEAL?

Ahora imaginemos

- No podemos observar directamente el valor de la suscripción ni el costo fijo.
- En el ingreso y los costos, se presentan variaciones al azar todos los meses. (Sabemos que son “pequeñas”)
- No conocemos necesariamente el proceso que genera el ingreso.
- Sabemos cuántos clientes hay cada mes
- ¿Qué podríamos hacer, si queremos predecir la utilidad para una cantidad de clientes? – **Lo formalizaremos hoy**

Ejercicio en Excel: 2

- Vamos a suponer que los datos de ingreso son proporcionales a la cantidad de clientes reportada por mes.
- Vamos a probar diferentes valores de B0 y B1 y ver gráficamente cómo se comporta.
- Vamos a buscar manualmente los valores que hagan más cercana nuestra predicción a los valores observados.

¿Qué es, entonces, la regresión lineal?

Un método para encontrar la línea "teórica" que mejor se ajusta los datos observados.

Para esto vamos a hablar de dos tipos de regresión

Simple, si hablamos de:

$$Y = B_0 + \ B_1 X_1 + \varepsilon$$

Múltiple, si hablamos de:

$$Y = B_0 + \ B_1 X_1 + B_2 X_2 \ + \cdots \ + B_n X_n + \varepsilon$$

con n mayor a 1

Vamos a trabajar con la regresión lineal simple

$$Y = B_0 + \ B_1 X_1 + \varepsilon$$

Cuando tengamos una "sospecha" de los valores $B_0 \ y \ B_1$ podemos generar una predicción de Y.

A esta predicción la llamaremos $\hat{Y}$

$$\hat{Y} \ = B_0 + \ B_1 X_1$$

Ahora remplazamos

$$Y = (B_0 + B_1 X_1) + \varepsilon$$

Remplazamos por $\hat{Y}$ y tenemos

$$Y = \hat{Y} + \varepsilon$$

Y tendremos si despejamos ε

$$\varepsilon = Y - \hat{Y}$$

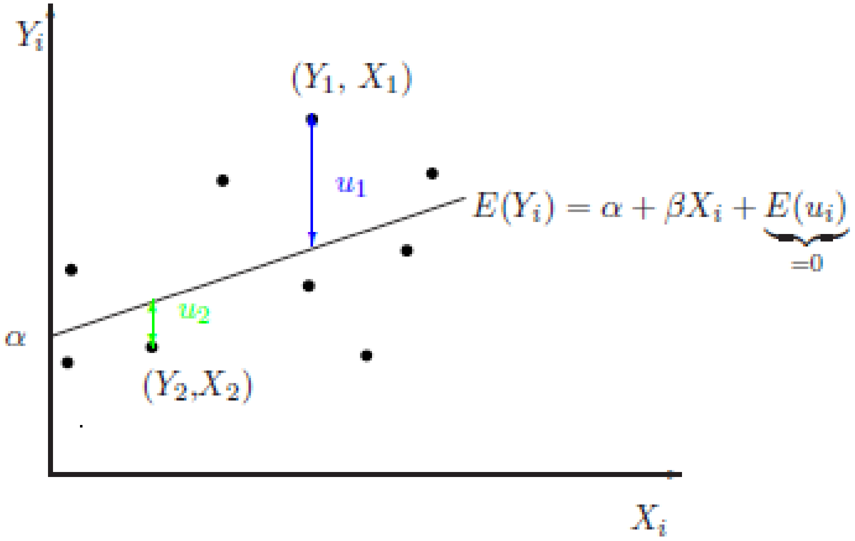
Y lo podríamos leer como: el error de observación es la diferencia entre el valor observado (Y) y el predicho ($\hat{Y}$)

Ejercicio en Excel 3:

Vamos a calcular la sumatoria del error con los datos el ejercicio 2:

$$\sum_{i=1}^{n} \varepsilon_i$$

Usemos buscar objetivo para cambiar el valor de B0, de tal forma que la sumatoria del error sea igual a cero.



REGRESIÓN LINEAL OLS
(MÍNIMOS CUADRADOS ORDINARIOS)

Un modelo de regresión lineal

… de <u>mínimos cuadrados ordinarios</u>, busca los valores B0 y B1, que hagan lo más pequeña posible la sumatoria de

$$\sum_{i=1}^{n} (\varepsilon_i)^2$$

En general este valor es mínimo cuando $\sum_{i=1}^{n} (\varepsilon_i) = 0$

Ejercicio en Excel 4: Solver

Vamos a usar solver para encontrar los valores de B0 y B1 que minimizan:

$$\sum_{i=1}^{n} \varepsilon_i$$

REGRESIÓN LINEAL OLS (ESTIMADORES Y R2)

Estimadores de OLS

Los estimadores de Mínimos cuadrados ordinarios se definen como:

$$B_1 = S_{xx}{}^{-1}\, S_{xy}$$

$$S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_{i-}\bar{y})$$

$$B_0 = \bar{y} - B_1\bar{x}$$

Estadísticas de desempeño: R2

$$R^2 = \frac{var(\hat{y})}{var(y)}$$

- Se interpreta como la proporción de la variabilidad, explicada por el modelo.
- Toma valores entre cero y uno, cuanto más cercano a uno más variabilidad es explicada por el modelo

Ejercicio en Excel 5: estimadores

Vamos a estimar B0 y B1 usando los estimadores de regresión lineal.

Vamos a calcular R2.

Con todo esto…

1. Entendemos los conceptos básicos de regresión y mínimos cuadrados.
2. Podemos estimar los parámetros B de una regresión.

¡Gracias!

Aprendiendo juntos a lo largo de la vida

educacioncontinua.uniandes.edu.co

Síguenos: **EdcoUniandes**