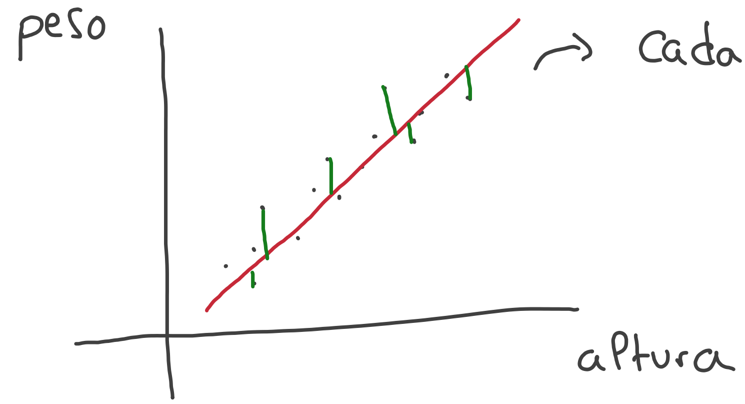
Regresión lineal Rep

jueves, 28 de abril de 2022 19:44

Supongamos que tenemos la siguiente repre
de dos variables:



→ Cada punto es una per

- ¿Existe alguna relación entre las dos va
- ¿Se puede encontrar una aproximación e
  datos?.

La respuesta a las dos preguntas es sí

La aproximación que vamos a buscar co
es una recta (aprox. lineal)

Ecuación de una recta

$$y = a + bx \implies y = \beta_0 + \beta_1 x$$

Desarrollemos el ejemplo:

Se toma la altura y el peso a

(Ejemplo R)

Resultado :

$y = -98.23 + 95.24X$

↳ Esta sería la recta de aproximación respecto a la altura

Ahora supongamos que se quiere hacer

físico y algunas personas rellenan el

pero olvidan el peso. Se decide utilizar

lineal antes utilizada:

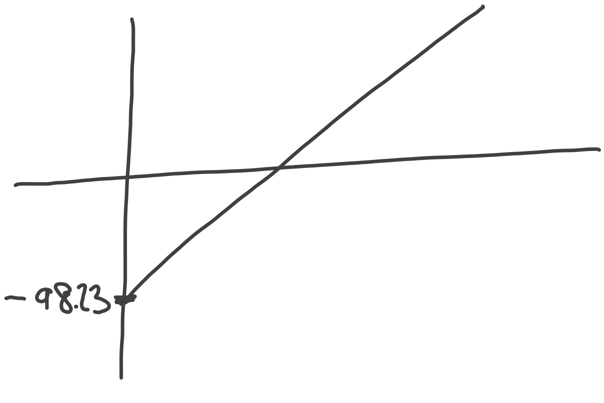
Competidor 1 ⇒ altura = 1.88 ⇒ peso = −98.23

Competidor 2 ⇒ altura = 1.61 ⇒ peso = −98.23 +

Esta recta $y = -98.23 + 95.24X$

↓

pendiente de la re



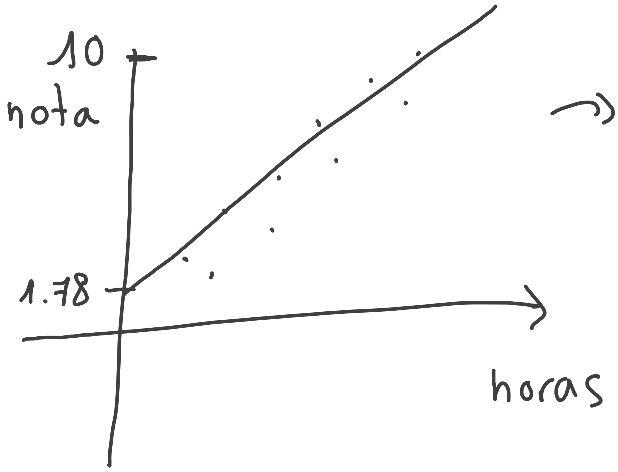
β_0 = valor de y a

β_1 = pendiente de

Otro ejemplo:

$y =$ nota de la materia AED

$x =$ horas de estudio a la semana



→ De nuevo correlació

A más horas de es

más nota :)

En este caso obtenemos nota $= 1.78$

¿Cómo se interpretan los β_i's?

$\beta_0 =$ la nota obtenida en el caso de que estudio sean 0.

$\beta_1 =$ cuanto aumentará y (nota) en caso aumente 1 unidad.

↳ es decir, a cada hora extra aumenta 0.79 la nota de AED.

Ejemplo regresión negativa: Temperatura

Ahora bien, esto es mucho más útil cua

con más de 2 variables.

Supongamos que queremos predecir 1

partir de otras. El concepto es el mi

$$\underbrace{y}_{\text{variable respuesta}} = \beta_0 + \underbrace{\beta_1 X_1 + \beta_2 X_2}_{\text{variables predictoras}} + \dots$$

Por ejemplo:

$$\underbrace{y}_{\text{valor propiedad inmobiliaria}} = \underbrace{\beta_1 X_1}_{\text{zona}} + \underbrace{\beta_2 X_2}_{m^2} + \underbrace{\beta_2 X_3}_{\text{antigüedad}} + \dots$$

Cuanta más información añadamos en

mejor ajustará.

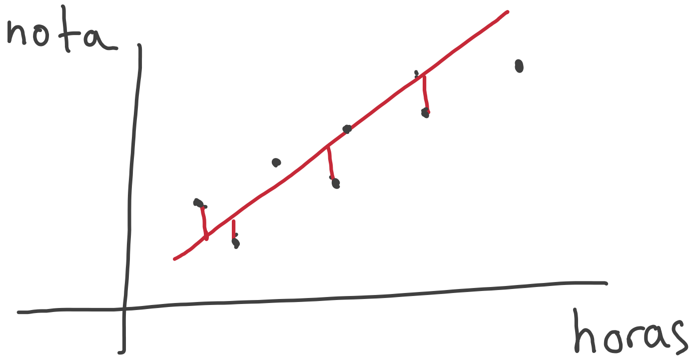
Modelo final:

$$y = \beta_0 + \beta_1 X_1 + \dots + \beta_r X_r + \underline{\underline{\varepsilon}} \rightarrow \text{err}$$

a

Para ajustar lo mejor posible este mo

minimizar los errores.



2 dimensiones

Error: Diferencia entre valor real y

Objetivo: Minimizar la suma de todos

Por tanto: Buscamos los β_i's del

que la recta sea óptima, es decir, qu

los errores.

Un poco de teoría:

Sea el modelo de regresión lineal:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_r X_r + \varepsilon$$

Si hay n observaciones independientes

expandirlo como:

$$Y_1 = \beta_0 + \beta_1 \cdot X_{11} + \dots + \beta_r \underbrace{X_{1r}}^{\text{obs. 1 var. r}} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 X_{21} + \dots + \beta_r X_{2r} + \varepsilon_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 X_{n1} + \dots + \beta_r X_{nr} + \varepsilon_n$$

Estos errores son una variable aleatori
se asume que:

1) $E(\varepsilon_i) = 0 \Rightarrow E(y_i) = \beta_0 + \beta_1 X_1$

2) $Var(\varepsilon_i) = \sigma^2$ (cte)

3) $Cov(\varepsilon_i, \varepsilon_j) = 0 \ \forall i \neq j$ (Indeps)

Esto lo veremos más adelante.

Podemos escribir en notación matricial:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & \dots & X_{1r} \\ \vdots & & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{nr} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_r \end{pmatrix} +$$

$$\Rightarrow \mathbb{Y} = \mathbb{X} \cdot \beta + \mathcal{E}$$

donde cada col. de $\mathbb{X}$ son las n

var. predictoras y cada fila i de
valores de las var. predictoras en

Estimador de mínimos cuadrados

Vamos a estudiar cómo se calculan
manera que devuelva los valores má
con relación lineal de $\mathbb{Y}$ y $\mathbb{X}$.

Es decir, los B_i's que minimicen los
residuos) $\Rightarrow$ β_i's que minimicen la diferen
el valor real y el valor estimado.

Por tanto, intentamos minimizar la
cuadrados:

$$S(\mathbb{b}) = \sum_{i=1}^{n} (\underbrace{y_i}_{\text{real}} - \underbrace{b_0 - b_1 x_{i1} - \dots - b_r x_{ir}}_{-\text{ estimado}})^2$$

$$= (\mathbb{Y} - \mathbb{b}\mathbb{X})' (\mathbb{Y} - \mathbb{b}\mathbb{X})$$

Definición: Las entradas de $\mathbb{b}$ que mi
se obtienen con mínimos cuadrados y

denotamos $\hat{\beta}$.

Teorema: El estimador de mínimos cuadr

está dado por:

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

Además, los residuos estimados se pueden calcul

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

con $H = X(X'X)^{-1}X'$

Ejemplo:

Example 7.3 (Calculating the least squares estimates, the residuals, and the residual sum of squares)

Calculate the least square estimates $\hat{\boldsymbol{\beta}}$, the residuals $\hat{\boldsymbol{\varepsilon}}$, and the residual sum of squares for a straight-line model

$$Y_j = \beta_0 + \beta_1 \cancel{z}_{j1} + \varepsilon_j$$

fit to the data

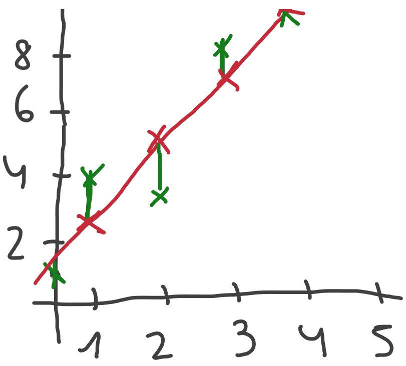
$\cancel{z_1}$	0	1	2	3	4
y	1	4	3	8	9

We have

$$\begin{array}{ccccc} \mathbf{Z}' & \mathbf{y} & \mathbf{Z}'\mathbf{Z} & (\mathbf{Z}'\mathbf{Z})^{-1} & \mathbf{Z}'\mathbf{y} \\ \begin{array}{r} 1's \rightarrow \\ x_1 \rightarrow \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} & \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} & \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix} & \begin{bmatrix} .6 & -.2 \\ -.2 & .1 \end{bmatrix} & \begin{bmatrix} 25 \\ 70 \end{bmatrix} \end{array}$$

10+

Consequently



reales

estimados

Consequently,

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \begin{bmatrix} .6 & -.2 \\ -.2 & .1 \end{bmatrix} \begin{bmatrix} 25 \\ 70 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

and the fitted equation is

$$\hat{y} = 1 + 2z \rightarrow \text{Recta de reg}$$

The vector of fitted (predicted) values is

$$\text{est} \rightarrow \hat{\mathbf{y}} = \mathbf{Z}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix}$$

so

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

The residual sum of squares is

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = [0 \quad 1 \quad -2 \quad 1 \quad 0] \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} = 0^2 + 1^2 + (-2)^2 + 1^2 + 0^2 = 6 \qquad \blacksquare$$