



POLITÉCNICA

# Proyecto 5

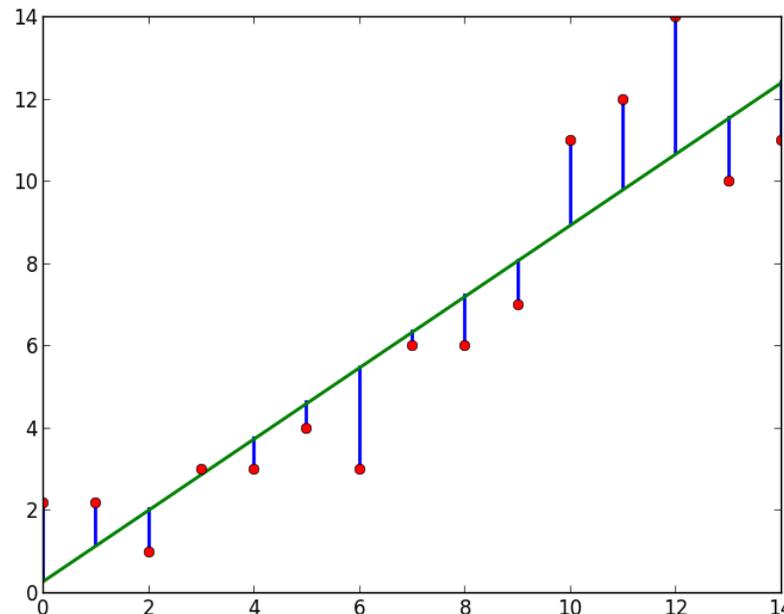


## □ Regresión lineal:

- El modelo lineal relaciona una variable *dependiente*  $Y$  con una variable *independiente*  $X$  mediante la siguiente expresión lineal:

$$Y = \beta_1 + \beta_2 X$$

## □ Interpretación: minimización de la suma de los errores al cuadrado





POLITÉCNICA

# Proyecto 5

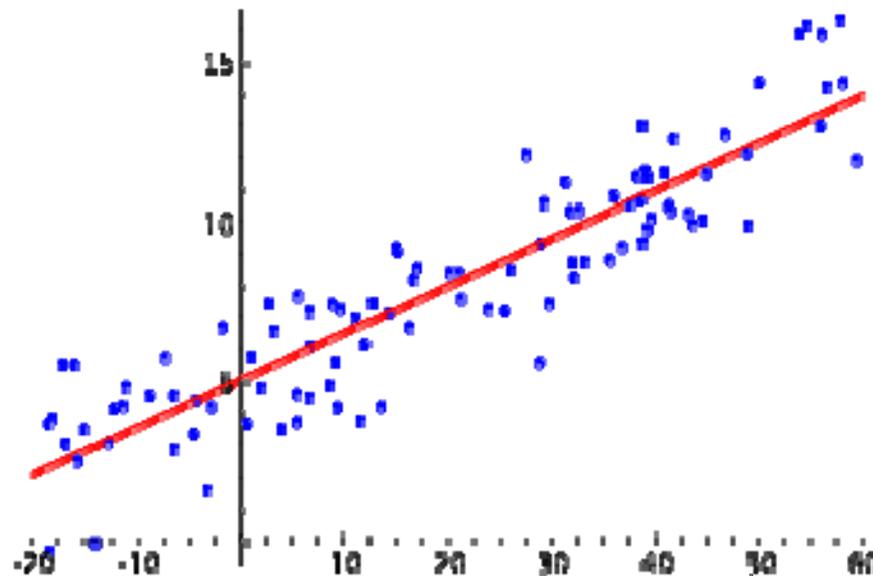


## □ Regresión lineal:

- El modelo lineal relaciona una variable *dependiente*  $Y$  con una variable *independiente*  $X$  mediante la siguiente expresión lineal:

$$Y = \beta_1 + \beta_2 X$$

## □ Ejemplo 1: existe una relación lineal clara entre $X$ e $Y$





POLITÉCNICA

# Proyecto 5

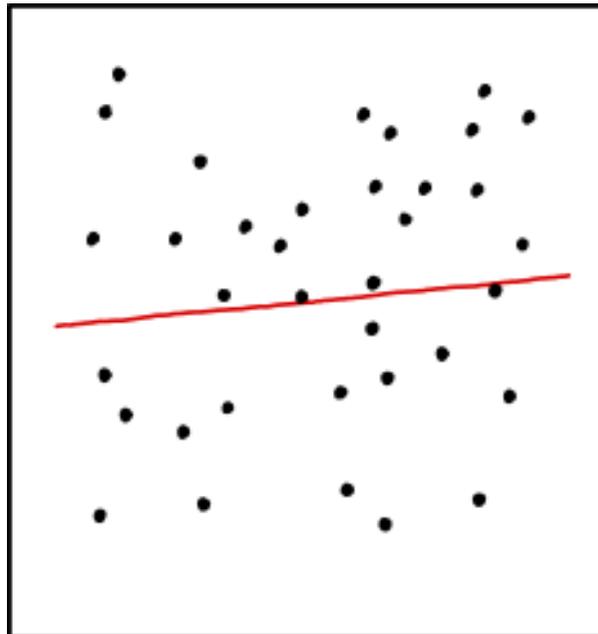


## □ Regresión lineal:

- El modelo lineal relaciona una variable *dependiente*  $Y$  con una variable *independiente*  $X$  mediante la siguiente expresión lineal:

$$Y = \beta_1 + \beta_2 X$$

## □ Ejemplo 2: no existe una relación lineal clara entre $X$ e $Y$





POLITÉCNICA

# Proyecto 5



## □ Regresión lineal:

- El modelo lineal relaciona una variable *dependiente*  $Y$  con una variable *independiente*  $X$  mediante la siguiente expresión lineal:

$$Y = \beta_1 + \beta_2 X$$

- El problema de la regresión consiste en elegir unos valores determinados para los parámetros desconocidos  $\beta_1$  y  $\beta_2$  de modo que la ecuación quede completamente especificada:

$$y = \bar{y} + \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} (x - \bar{x})$$



POLITÉCNICA

# Proyecto 5



□ Coeficiente de correlación:

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

- El coeficiente de correlación de la recta determinará la calidad del ajuste. Si  $r$  es cercano o igual a 1, el ajuste será bueno, y si es cercano o igual a 0, se tratará de un ajuste malo.



POLITÉCNICA

# Proyecto 5



□ Input: archivo con el siguiente formato

c <número de casos>  
n <número de puntos del primer caso>  
x <coordenada x del primer punto del primer caso>  
y <coordenada y del primer punto del primer caso>  
...  
x <coordenada x del último punto del primer caso>  
y <coordenada y del último punto del primer caso>  
...  
n <número de puntos del primer caso>  
x <coordenada x del primer punto del último caso>  
y <coordenada y del primer punto del último caso>  
...  
x <coordenada x del último punto del último caso>  
y <coordenada y del último punto del último caso>



POLITÉCNICA

# Proyecto 5



□ Output: archivo con el siguiente formato

$\beta_1$	<parámetro $\beta_1$ de la primera recta de regresión>
$\beta_2$	<parámetro $\beta_2$ de la primera recta de regresión>
r	<coeficiente de correlación de la primera regresión>
...	
$\beta_1$	<parámetro $\beta_1$ de la última recta de regresión>
$\beta_2$	<parámetro $\beta_2$ de la última recta de regresión>
r	<coeficiente de correlación de la última regresión>



POLITÉCNICA

# Proyecto 5-2



- Inventar un método para eliminar el punto que peor encaja con el resto (**punto espurio**).
- Implementarlo en un método que toma como entrada la lista de puntos y los parámetros  $\beta_1$ ,  $\beta_2$  y  $r$  de la recta de regresión.
- Aplicar el método anterior para suprimir el 5% de los puntos y devolver como resultado la lista de puntos resultante del proceso de borrado y los parámetros  $\beta_1$ ,  $\beta_2$  y  $r$  de la recta de regresión de la lista final



POLITÉCNICA

# Proyecto 5-2



□ Input: archivo con el siguiente formato

n            <número de puntos>  
x            <coordenada x del primer punto>  
y            <coordenada y del primer punto>  
...  
x            <coordenada x del último punto>  
y            <coordenada y del último punto>



POLITÉCNICA

# Proyecto 5-2



□ Output: archivo con el siguiente formato

$\beta_1$	<parámetro $\beta_1$ de la recta de regresión>
$\beta_2$	<parámetro $\beta_2$ de la recta de regresión>
r	<coeficiente de correlación de la recta de regresión>
m	<número de puntos>
x	<coordenada x del primer punto>
y	<coordenada y del primer punto>
...	
x	<coordenada x del último punto>
y	<coordenada y del último punto>