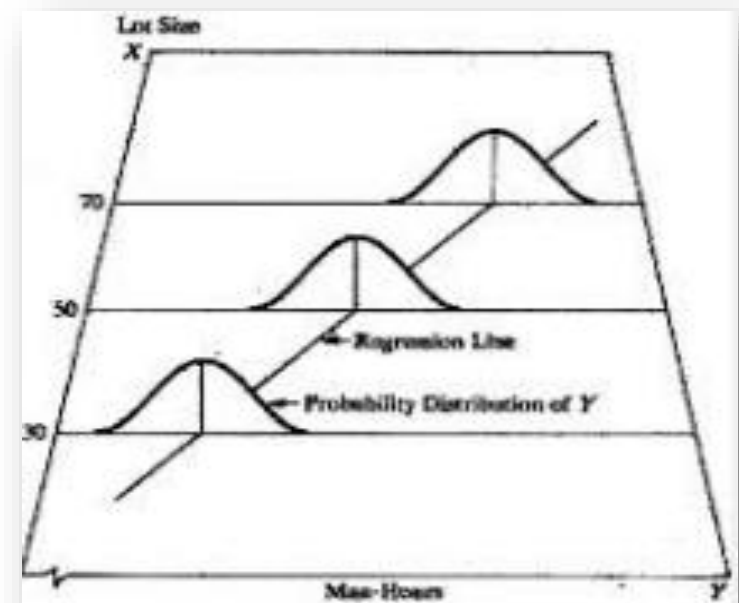
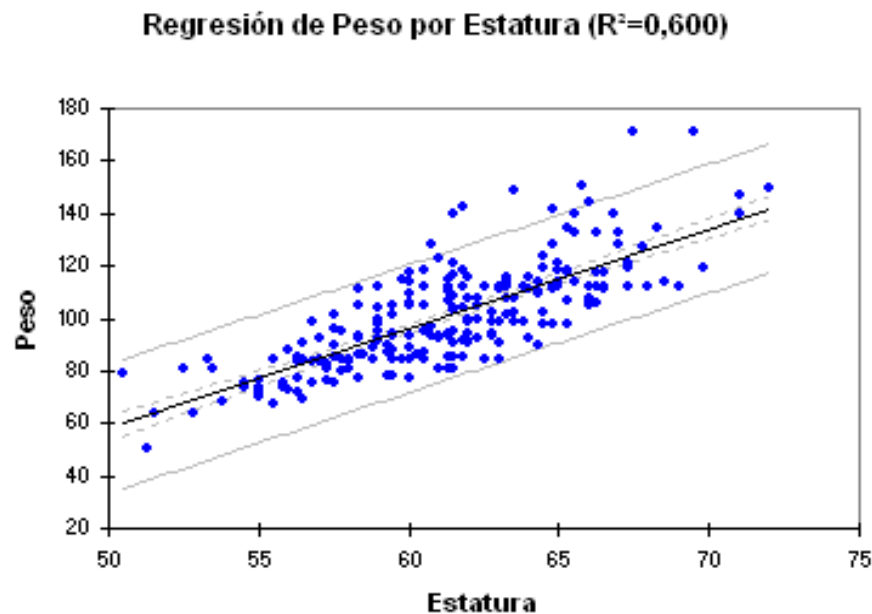


# Unidad V

## Regresión lineal Simple

### Parte I



# Análisis de Regresión

El **análisis de regresión** es una técnica estadística busca encontrar un modelo que permita describir el comportamiento de una v.a.  $Y$  ante modificaciones de una variable (*regresión simple*) o varias variables (*regresión múltiple*).

## Donde

**$Y$ : Variable Respuesta - dependiente - predicha - explicada.**

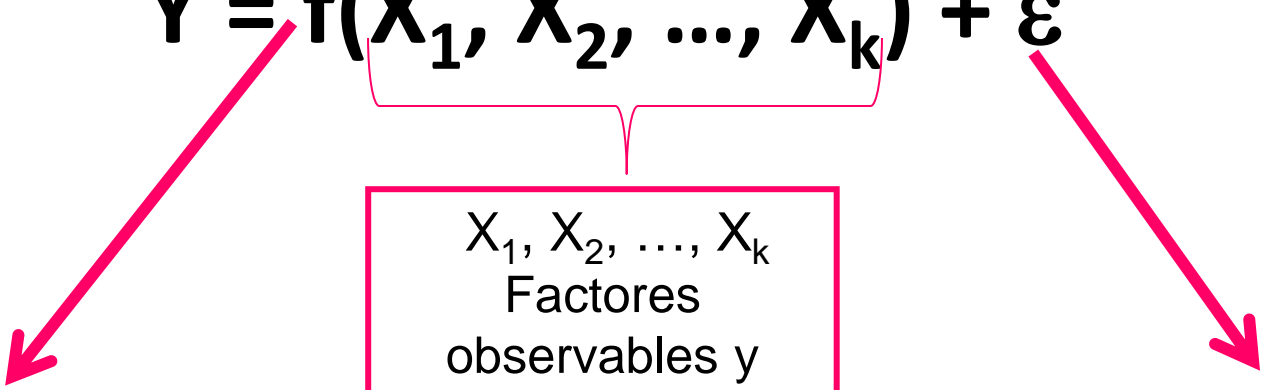
- Es la variable de interés en el problema. Es una variable aleatoria.

**$X_1, X_2, \dots, X_k$ : variables regresoras - independientes - explicativas - predictoras.**

- Aportan información sobre el comportamiento de  $Y$ . El experimentador controla sus valores y los cambia para ver el efecto que producen en  $Y$ .

# Modelo de Regresión

La variable respuesta  $Y$  se expresa:

$$Y = f(X_1, X_2, \dots, X_k) + \varepsilon$$
The diagram features the regression equation  $Y = f(X_1, X_2, \dots, X_k) + \varepsilon$  at the top center. A bracket underneath the function argument  $(X_1, X_2, \dots, X_k)$  points down to a central box. A red arrow points from the function  $f$  to a box on the left, and another red arrow points from the error term  $\varepsilon$  to a box on the right.

$X_1, X_2, \dots, X_k$   
Factores  
observables y  
controlables.

Establece el tipo de relación entre la variable respuesta y los regresores.

Error aleatorio o perturbación, representa el conjunto de factores no observables, desconocidos o incontrolables.

**EJEMPLO.** MODELO PARA EXPLICAR EL PRECIO DE ALQUILER DE DEPARTAMENTOS EN LA CIUDAD DE BAHIA BLANCA.

El *precio de alquiler de un departamento en la ciudad* cuya variación puede ser **predicha** por las **variables explicativas**: *superficie y antigüedad* del edificio.

$$Y = f(X_1, X_2) + \varepsilon$$

Y: Precio de alquiler.

$X_1$ : superficie.

$X_2$ : Antigüedad del edificio.

$\varepsilon$  = En el término del error aleatorio pueden incluirse el efecto del número de habitaciones, situación geográfica, altura, orientación, etc.

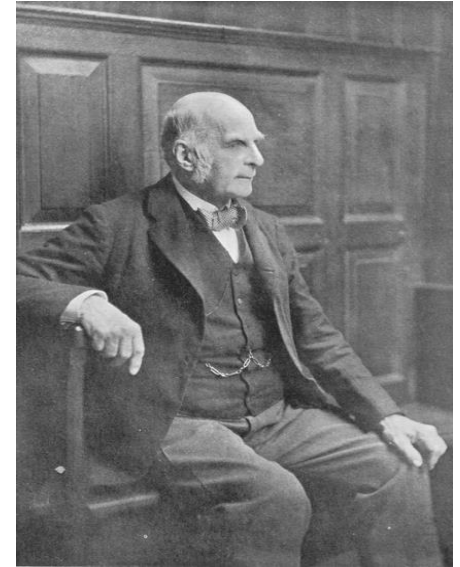
# Regresión

El término **regresión** fue introducido por **Francis Galton** en su libro “*Natural inheritance*” (1889) refiriéndose a la “ley de la regresión universal”.

Su trabajo se centraba en la descripción de los rasgos físicos de los descendientes (una variable) a partir de los de sus padres (otra variable).

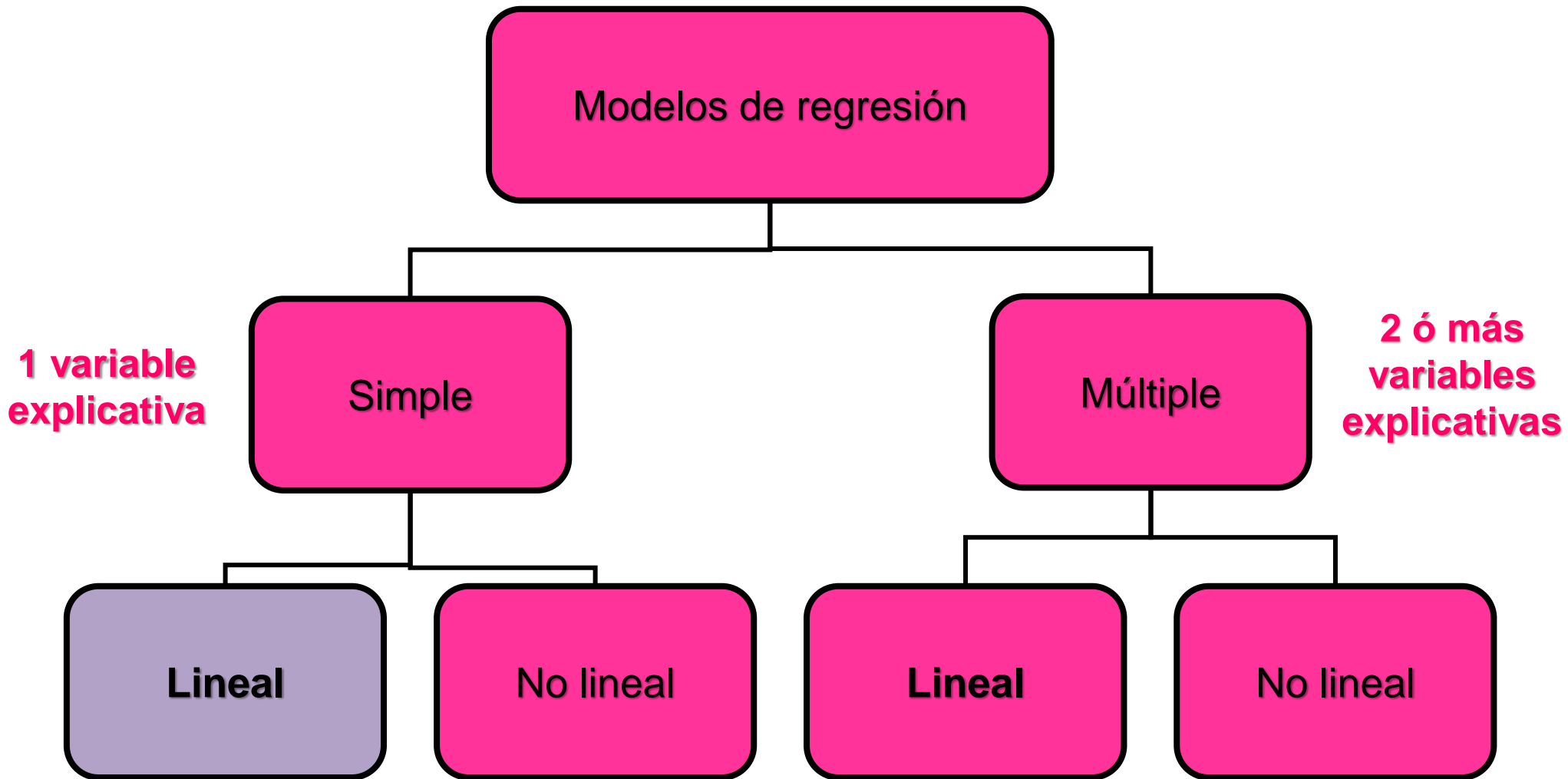
Él mostró en su estudio que los padres muy altos tienen tendencia a tener hijos cuya altura tiende a *retroceder* o “*regresar*” a la media de la población adulta. Lo mismo puede decirse de los padres muy bajos.

Hoy en día el sentido del término **regresión** es el proceso de predicción de una medida basándonos en el conocimiento de otra u otras.



- Primo de Darwin.
- Estadístico.
- Fundador (con otros) de la estadística moderna para explicar las teorías de Darwin.
- Antropólogo.
- Psicólogo.

# Modelos de análisis de regresión

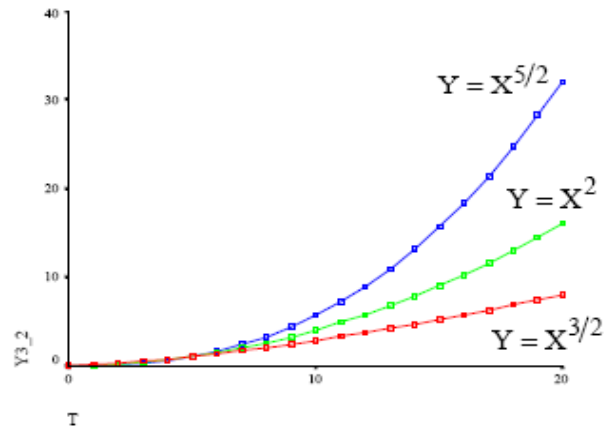


Nosotros en el curso nos centraremos en el modelo de **Regresión Lineal Simple**

# Modelos de regresión

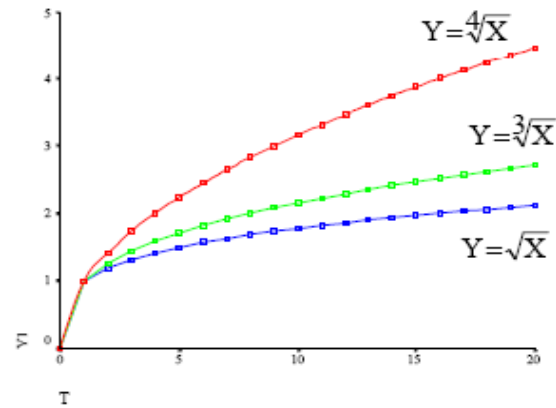
## FUNCIONES POTENCIALES

$$Y=a X^b \text{ con } b>1$$



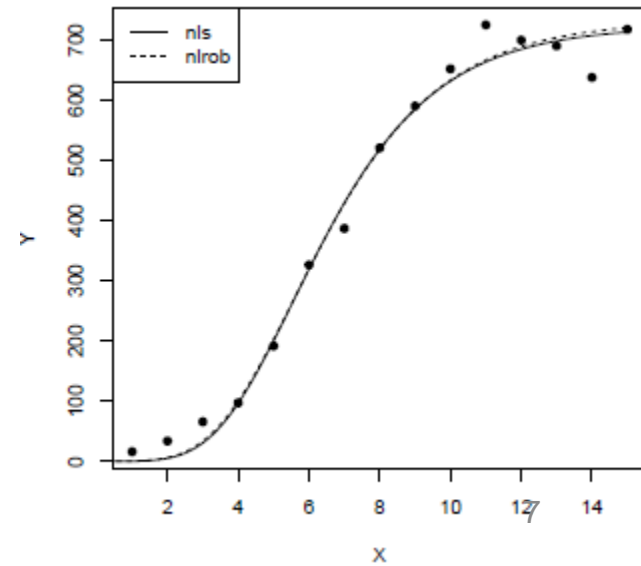
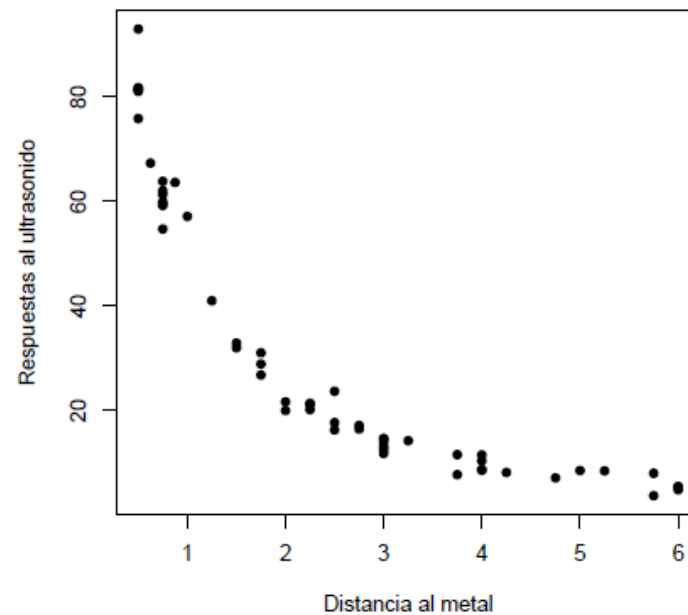
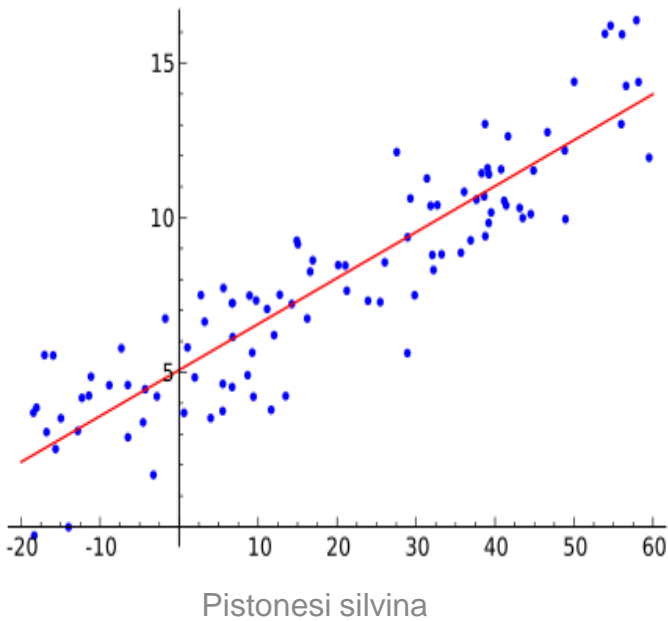
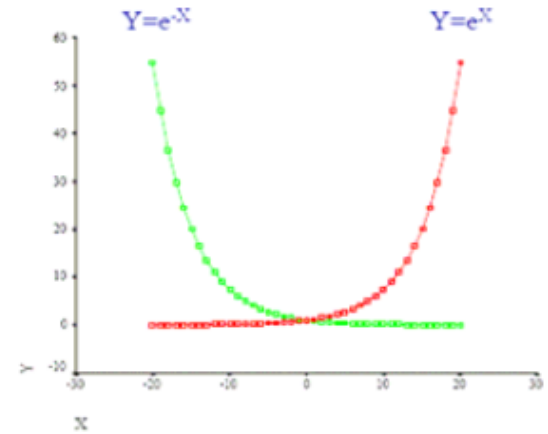
## FUNCIONES POTENCIALES

$$Y=a X^b \text{ con } 0 < b < 1$$



## FUNCIONES EXPONENCIALES

$$Y=a b^X$$



# Regresión lineal Simple

Pretende explicar una **variable respuesta** en mediante una **variable explicativa** a través de una función lineal.

$$Y = f(X) + \varepsilon$$

- la variable ***dependiente o respuesta***: es la que se desea explicar o predecir.
- Hay ***una sola variable explicativa*** (o variable independiente o predictora). Por tal motivo el término «**Simple**».
- La función ***f(X)*** es ***Lineal***.
- $\varepsilon$  : es el ***error aleatorio o perturbación***.



# Modelo de regresión lineal simple

La variable respuesta  $Y$  se relaciona con la variable independiente  $X$  a través del modelo:

$$(1) \quad Y = \beta_0 + \beta_1 X + \varepsilon$$

modelo de  
regresión lineal  
simple

donde  $\beta_0$  y  $\beta_1$  son los parámetros del modelo.

$\beta_0$  Es la ordenada al origen y representa el valor de  $Y$  cuando  $X = 0$ , (tiene las mismas unidades que  $Y$ .)

$\beta_1$  Es la pendiente e indica cuanto aumenta  $Y$  (si es positiva) ó disminuye  $Y$  (si es negativa) cuando  $X$  aumenta en una unidad; (tiene las unidades de  $Y$  sobre las de  $X$ ).

$\varepsilon$  Es el término de error aleatorio, variable aleatoria, con  $E(\varepsilon) = 0$  y  $V(\varepsilon) = \sigma^2$ . Es independiente de  $X$ .

El modelo (1) se dice **simple** y **lineal** en la variable explicativa.

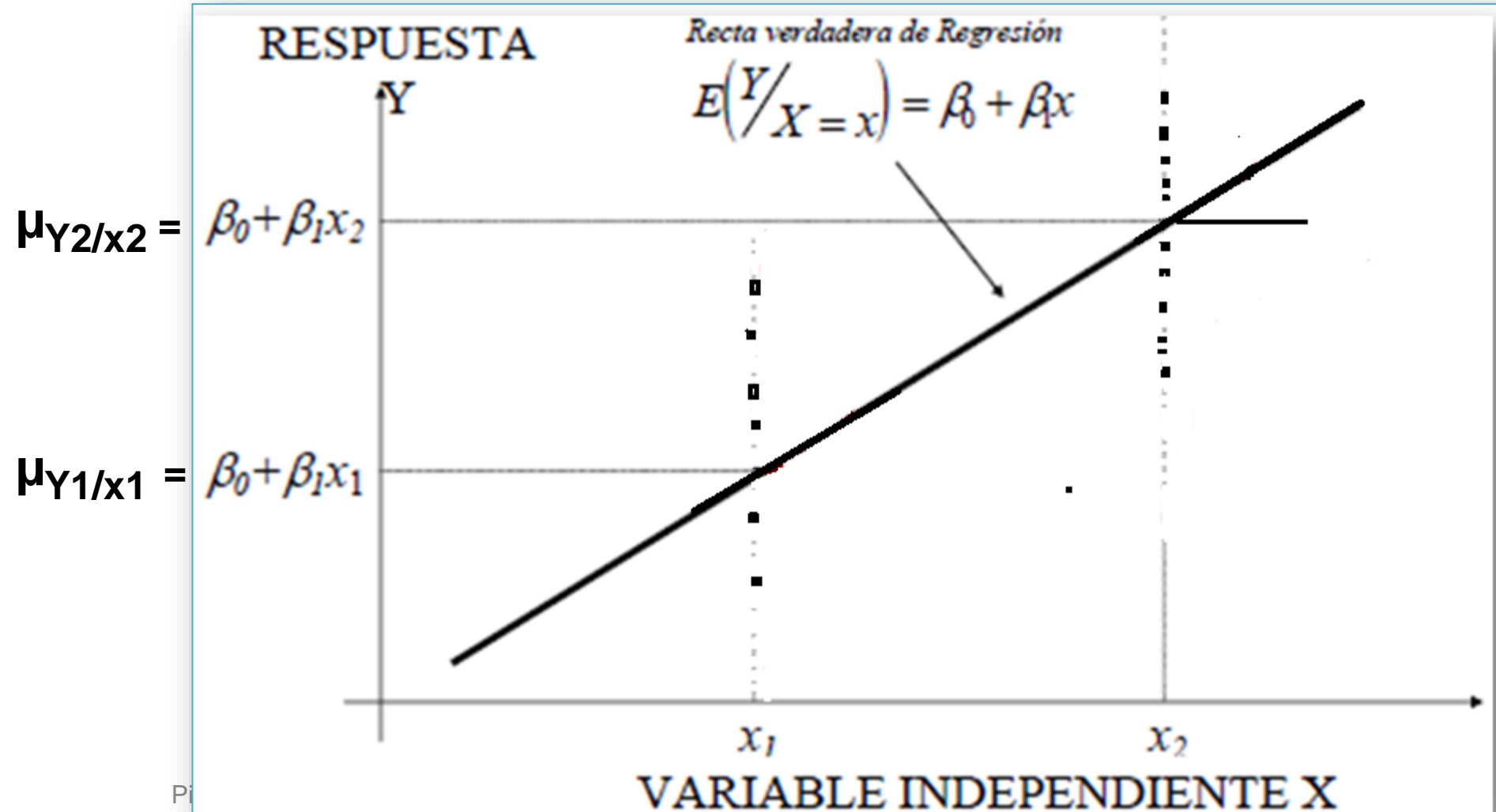
# Ejemplos de Regresión Lineal Simple

- 1) El **consumo de nafta** de un vehículo, cuya variación puede ser explicada por la **velocidad media** del mismo. Podemos incluir en el término del error aleatorio el efecto del conductor, del tipo de rutas, las condiciones ambientales, etc.
- 2) El **presupuesto de una universidad**, cuya variación puede ser predicha por la variable explicativa **número de alumnos**. En el término del error aleatorio pueden incluirse el efecto del número de profesores, del número de laboratorios, la superficie disponible de instalaciones, del número de personal de administración, etc.



# Modelo de regresión lineal simple

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i = E(y/x_i) + \varepsilon_i = \mu_{Y_i/X_i} + \varepsilon_i$$



# Modelo de regresión lineal simple

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i = E(y/x_i) + \varepsilon_i = \mu_{Y_i/X_i} + \varepsilon_i$$

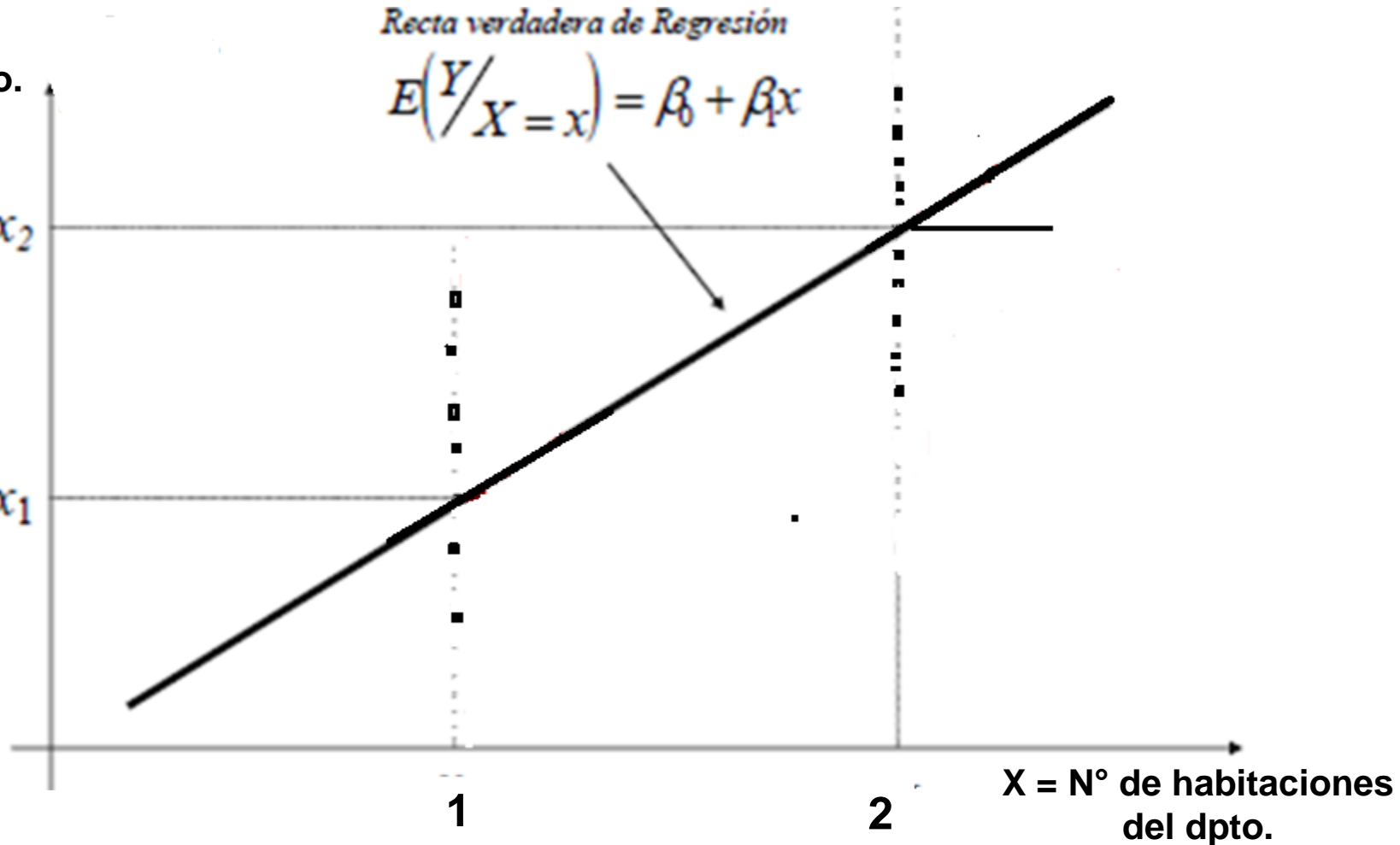
Y = Precio del alquiler del dpto.

*Recta verdadera de Regresión*

$$E\left(\frac{Y}{X=x}\right) = \beta_0 + \beta_1 x$$

$$\mu_{Y2/x2} = \beta_0 + \beta_1 x_2$$

$$\mu_{Y1/x1} = \beta_0 + \beta_1 x_1$$



# Modelo de regresión lineal simple

Desafortunadamente, los valores de los parámetros,  $\beta_0$  y  $\beta_1$ , **no se conocen**, y deben estimarse a través de una muestra.

- Dadas  $n$  observaciones  $(x_1, y_1), \dots, (x_n, y_n)$ , el modelo de regresión se **estima** mediante la siguiente recta, que permite a través de valores de  $X$  predecir los valores medios de  $Y$ ,

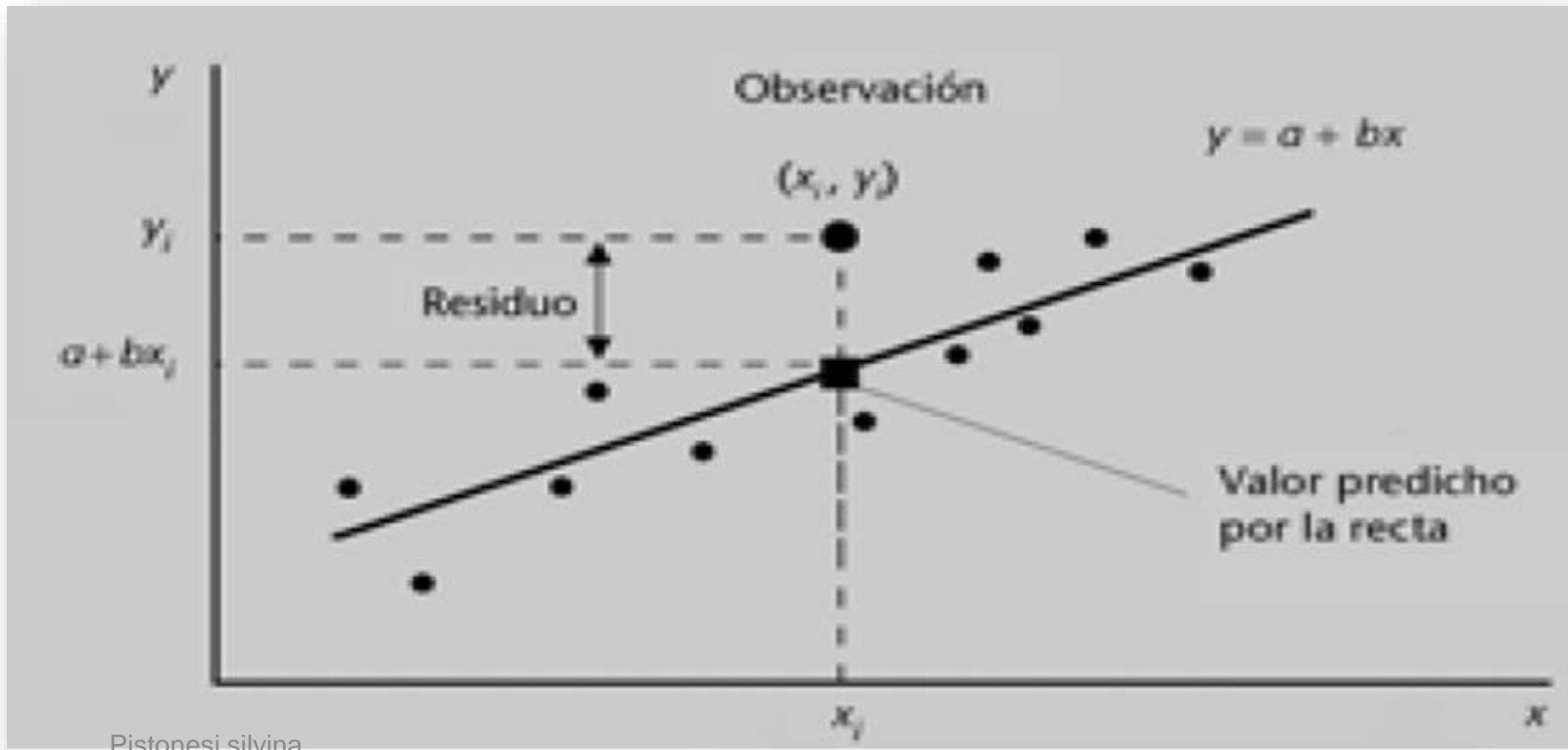
$$\hat{Y} = a + b X$$

- **a** (ordenada en el origen ó también llamada constante) estimador de  $\beta_0$
- **b** (pendiente de la recta) estimador de  $\beta_1$

# Modelo de regresión lineal simple

$Y_i$  e  $\hat{Y}_i$  rara vez coincidirán por muy bueno que sea el modelo de regresión. A la cantidad:

$\mathbf{e_i = Y_i - \hat{Y}_i}$  se le denomina **residuo** o **error residual**.

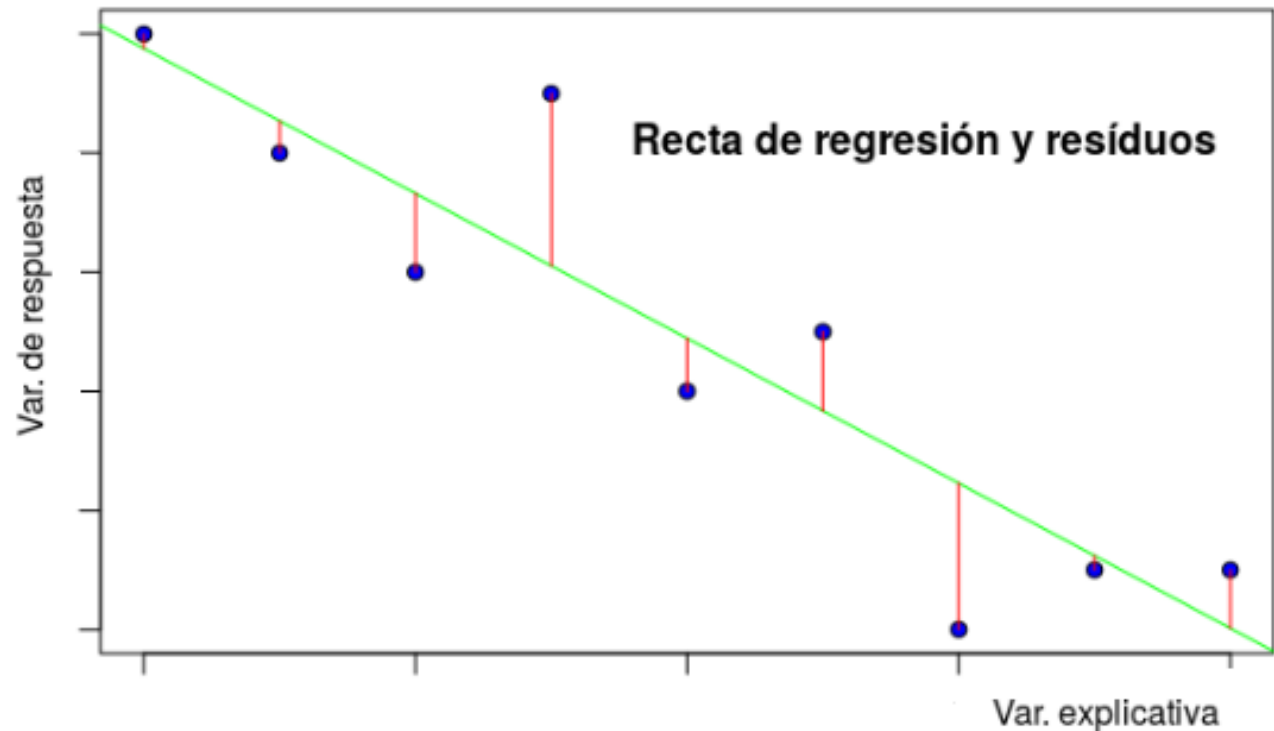


# Método de Mínimos cuadrados

Se determinarán las estimaciones, **a** y **b**, de los parámetros  $\beta_0$  y  $\beta_1$ , de tal forma que la suma de los cuadrados de los residuos sea mínima.

Este procedimiento de minimización para estimar los parámetros se llama **método de Mínimos cuadrados**:

Hallar las expresiones de **a** y **b** de tal manera que se minimice la cantidad  $\phi(\mathbf{a}, \mathbf{b})$ :



$$\phi(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n (e_i)^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \mathbf{a} - \mathbf{b}X)^2$$

Hallar **a**, **b** de tal manera que se **minimice** la cantidad:

$$\phi(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n (e_i)^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \mathbf{a} - \mathbf{b}X)^2$$

Derivando parcialmente  $\phi$  respecto de **a** y de **b**, e igualando a **0**, y resolviendo las ecuaciones para a y b, se obtiene:

$$\frac{\partial \phi}{\partial \mathbf{a}} = -2 \sum_{i=1}^n (y_i - \mathbf{a} - \mathbf{b}x_i) = 0$$

$$\frac{\partial \phi}{\partial \mathbf{b}} = -2 \sum_{i=1}^n (y_i - \mathbf{a} - \mathbf{b}x_i)x_i = 0$$

**Ecuaciones Normales**

Resolviendo el sistema se obtiene:

$$\mathbf{b} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{S_{xy}}{S_x^2} \quad \mathbf{a} = \bar{y} - b \bar{x}$$

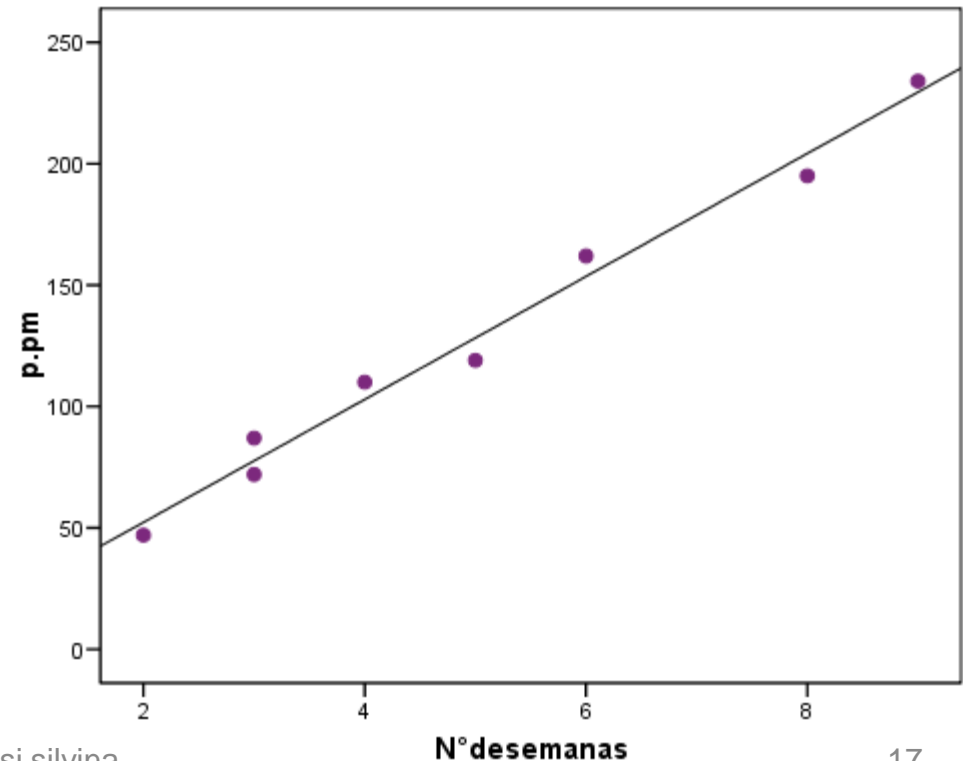


# Ejemplo

El departamento de personal de la empresa informática dedicada a la introducción de datos ha llevado a cabo un programa de formación inicial del personal. La siguiente tabla indica el progreso en pulsaciones por minuto (p.p.m.) obtenido en mecanografía de ocho empleados que siguieron el programa y el número de semanas que hace que lo siguen:



Nº de semanas (X)	Ganancia en velocidad (p.p.m.) (Y)
3	87
5	119
2	47
8	195
6	162
9	234
3	72
4	110



Medias muestrales:  $\bar{x} = 5$  e  $\bar{y} = 128.25$

Varianza muestral:  $S^2_x = 6,286$

Covarianza muestral:  $S_{xy} = 159.143$

$$b = \frac{S_{xy}}{S_x^2} = \frac{159.143}{6.286} = \mathbf{25.318} \qquad a = \bar{y} - b \bar{x} = 128.25 - 25.318 \cdot 5 = \mathbf{1.659}$$

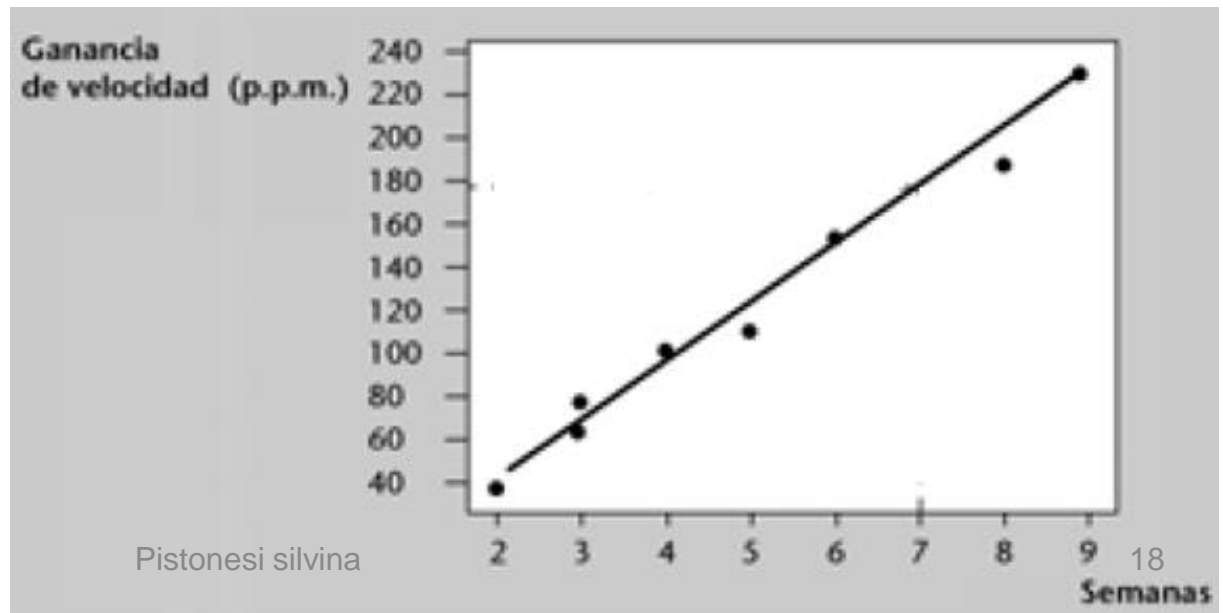
$$\hat{Y} = 1.659 + 25.318 X$$



Recta de ajuste

- En este caso la **ordenada al origen** no debe interpretarse, (no esta incluido el 0 en el rango de **X**, representaría la ganancia de velocidad con **0** semanas de clases (la velocidad que posee el empleado al comenzar el curso).

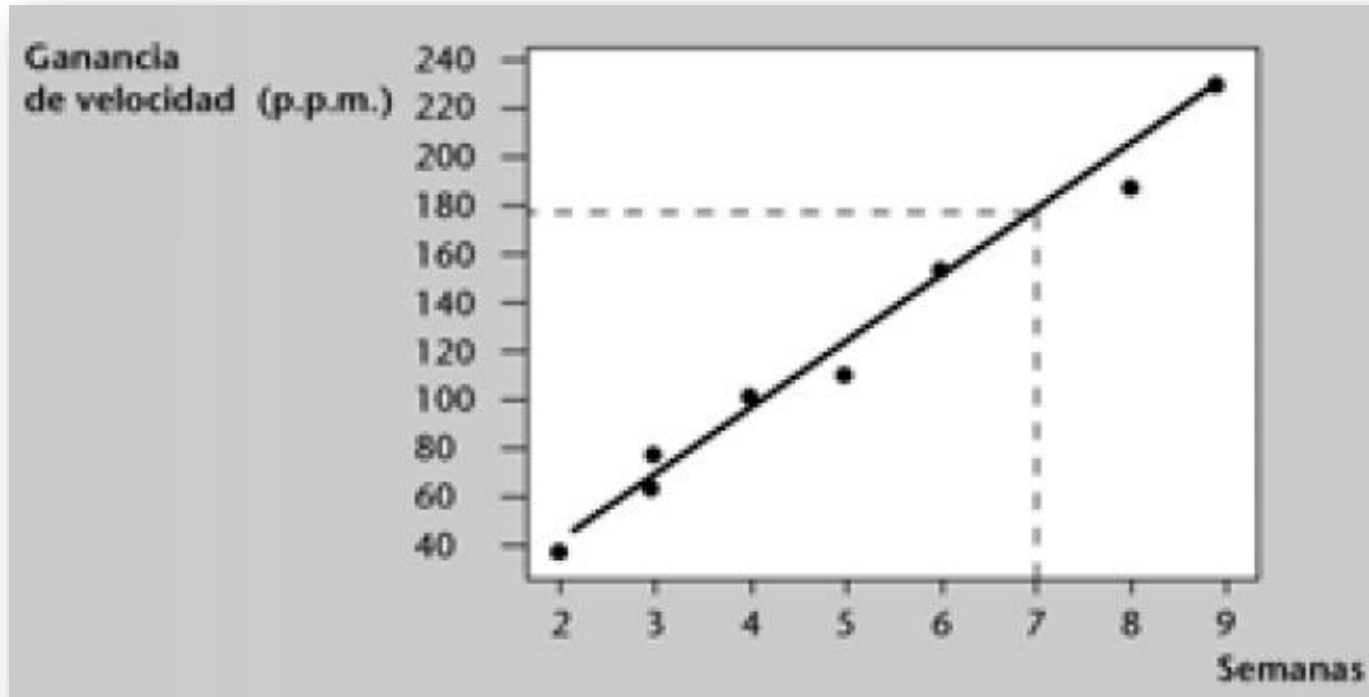
- La **pendiente** indicaría que, por cada semana de clase se tiene una ganancia de velocidad de aproximadamente 25 p.p.m.



$$\hat{Y} = 1.659 + 25.318 X$$

→ Recta de ajuste

¿Cuál es la **ganancia promedio** de una persona que asistió **7 semanas** a clase?



$$178.885 = 1.659 + 25.318 \cdot 7$$

Una persona que asistió 7 semanas a clase, la ganancia promedio de velocidad es de aprox. **178.885 p.p.m.**

# Uso de la Calculadora

Las calculadoras permiten calcular la pendiente y la ordenada al origen, entre otras cantidades.

Por ejemplo: en una calculadora CASIO fx – 82MS, se debe ir:

Mode >> 3 (Reg)

>> 1 (Lin)

Luego, ingresar los datos.



Explicado en el Power point de **Análisis de Correlación**

# Interpolación y extrapolación

Uno de los objetivos más importantes de la regresión es la aplicación del modelo para el **pronóstico del valor promedio de la variable dependiente (Y)** para un valor de la variable independiente (**X**) no observado en la muestra.

**Ejemplo:** La siguiente tabla corresponde a observaciones de los pesos (kg) (Y) y las alturas (cm) (X) de un conjunto de 10 personas.

Individuos (y)	1	2	3	4	5	6	7	8	9	10
Altura ( $x_i$ )	161	152	167	153	161	168	167	153	159	173
Peso ( $y_i$ )	63	56	77	49	72	62	68	48	57	67

$$b = \frac{S_{xy}}{S_x^2} = 0.979 \quad a = \bar{y} - b \bar{x} = -96.11$$

$$\hat{Y} = -96.11 + 0.979 X$$

# Interpolación y extrapolación

Uno de los objetivos más importantes de la regresión es la aplicación del modelo para el **pronóstico del valor promedio de la variable dependiente (Y)** para un valor de la variable independiente (**X**) no observado en la muestra.

**Ejemplo:** La siguiente tabla corresponde a observaciones de los pesos (kg) (Y) y las alturas (cm) (X) de un conjunto de 10 personas.

Individuos (y)	1	2	3	4	5	6	7	8	9	10
Altura ( $x_i$ )	161	152	167	153	161	168	167	153	159	173
Peso ( $y_i$ )	63	56	77	49	72	62	68	48	57	67



$$\hat{Y} = - 96.11 + 0.979 X$$

¿Cuál es el peso promedio de una persona de altura 160 cm?  
(interpolación)

# Interpolación y extrapolación

**Ejemplo:** La siguiente tabla corresponde a observaciones de los pesos (kg) ( $Y$ ) y las alturas (cm) ( $X$ ) de un conjunto de 10 personas.

Individuos ( $y$ )	1	2	3	4	5	6	7	8	9	10
Altura ( $x_i$ )	161	152	167	153	161	168	167	153	159	173
Peso ( $y_i$ )	63	56	77	49	72	62	68	48	57	67

$$\hat{Y} = - 96.11 + 0.979 X$$

¿Cuál es el peso promedio de una persona de altura 160 cm?

(interpolación)

$$\hat{Y} = - 96.1121 + 0.979009 \cdot 160 = 60.53 \text{ kg.}$$

**Extrapolación:** cuando se quiere conocer el valor promedio que presentará la variable  $Y$  para un determinado valor de  $X$  que se encuentre fuera del intervalo de valores de la  $X$ , con el que se construyó el modelo de ajuste. En este caso, **no se puede predecir** el valor promedio de  $Y$  con la recta de regresión, pues se **desconoce** si fuera del rango de las  $X$ , la **relación** entre  $X$  e  $Y$  sigue siendo **lineal**.

# La calidad del ajuste

La recta de regresión obtenida por **mínimos cuadrados** minimiza la suma de los cuadrados de los residuos.



## ¿CÓMO EVALUAMOS SI EL AJUSTE LINEAL LOGRADO ES BUENO?

El diagrama de dispersión nos permite ver si los puntos experimentales quedan muy cerca o no de la recta de regresión obtenida. Nos da una idea de si la recta se ajusta o no a los datos, pero nos hace falta un **valor numérico** que nos ayude a precisarlo.



# Bondad del ajuste: *coeficiente de determinación*

La **bondad de un ajuste** de un modelo de regresión se mide a través del **coeficiente de determinación  $R^2$**

$$R^2 = \frac{\text{Varianza explicada por la recta de regresión}}{\text{varianza de los valores observados } y_i}$$



- $R^2$  representa la **proporción de variabilidad de la variable Y explicada** por el modelo de regresión lineal.
- $R^2$  es un coeficiente nos indica el **grado de ajuste** de la recta de regresión a los valores de la muestra.
- $R^2$  es una cantidad **adimensional** que sólo puede tomar valores en  **$[0, 1]$** ,
- Cuando un **ajuste es bueno**,  $R^2$  será cercano a **uno**.
- Cuando un **ajuste es malo**  $R^2$  será cercano a **cero**.

# Bondad del ajuste: coeficiente de determinación ( $R^2$ )

Vamos a hallar el  $R^2$ :

$$y_i - \hat{y}_i = e_i \quad \longrightarrow \quad y_i = \hat{y}_i + e_i$$

Si restamos a ambos miembros de la igualdad la media de las observaciones  $\bar{y}$

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + e_i$$

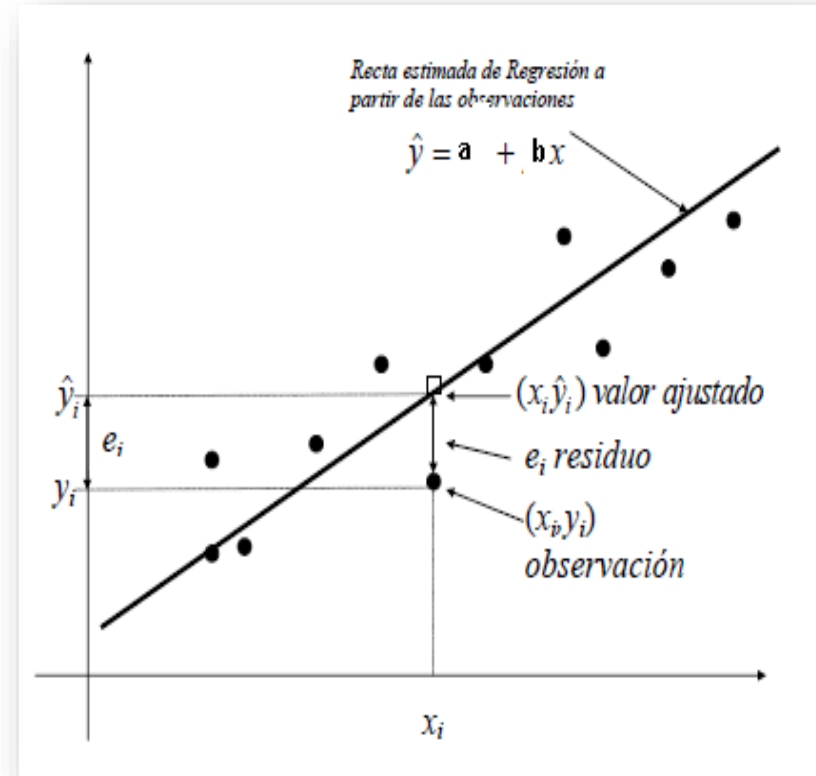
Elevando al cuadrado y sumando todos los valores, se demuestra que

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

Suma de  
Cuadrados Totales

Suma de Cuadrados de  
la Regresión

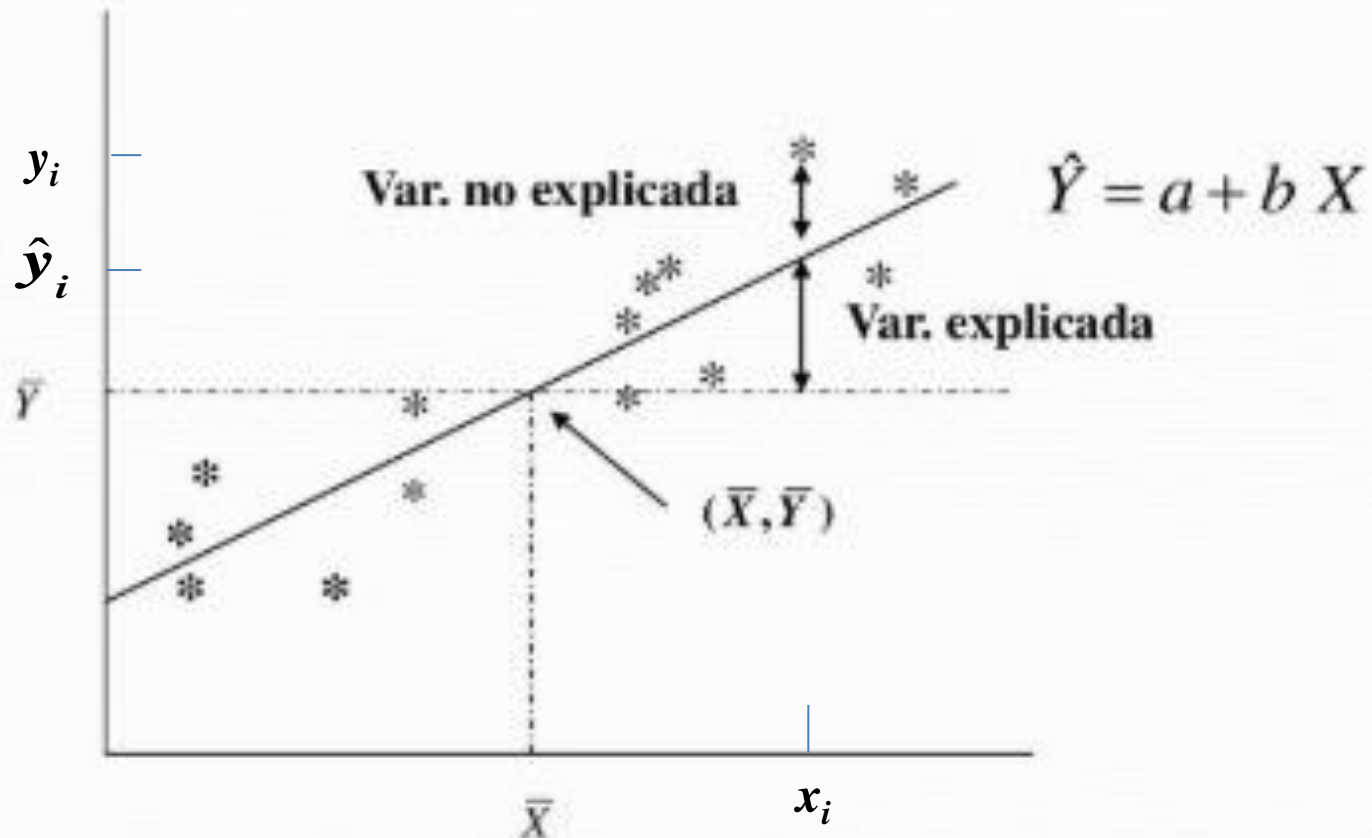
Suma de Cuadrados de  
los errores o residual



# Sumas de cuadrados

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

Variación total = Var. explicada + Var. no explicada



$$\mathbf{SCT = SCR + SCE}$$

$$SCT = \sum_{i=1}^n (y_i - \bar{y})^2$$

**Suma de Cuadrados Totales**

(numerador de la variabilidad muestral de y)

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

**Suma de Cuadrados de la Regresión**

(numerador de la variabilidad del modelo de ajuste)

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

**Suma de Cuadrados de los errores**

(numerador de la variabilidad residual)

**El  $R^2$ , coeficiente de determinación:**

$$\mathbf{SCT = SCR + SCE}$$

$$\frac{\mathbf{SCT}}{\mathbf{SCT}} = \frac{\mathbf{SCR}}{\mathbf{SCT}} + \frac{\mathbf{SCE}}{\mathbf{SCT}}$$

$$\mathbf{1 = \frac{SCR}{SCT} + \frac{SCE}{SCT}}$$

$$\mathbf{R^2 = \frac{SCR}{SCT} = \frac{\sum_{i=1}^n (\hat{y} - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_{i=1}^n e^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

**$R^2$**  se obtiene como el cociente entre la **SCR** y la **SCT**

# El $R^2$ , coeficiente de determinación se obtiene:

I) Una forma de determinar el  $R^2$  es:

$$\mathbf{R}^2 = \frac{\mathbf{SCR}}{\mathbf{SCT}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

II) Otra forma de determinar el  $R^2$  es:

$$\mathbf{R}^2 = 1 - \frac{\mathbf{SCE}}{\mathbf{SCT}} = 1 - \frac{\sum_{i=1}^n e^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

III) Otra forma de determinar el  $R^2$  es:

$$R^2 = \frac{b^2 S^2_x}{S^2_y}$$

¿Cuál uso?



IV) Otra forma de determinar el  $R^2$  es mediante el coeficiente de correlación muestral,  $r$ .

coeficiente de determinación =  $R^2 = r^2 = (\text{coeficiente de correlación})^2$

## Ejemplo

La siguiente tabla corresponde a observaciones de los pesos (kg) ( $Y$ ) y las alturas (cm) ( $X$ ) de un conjunto de diez personas.



Individuos ( $y$ )	1	2	3	4	5	6	7	8	9	10
Altura ( $x_i$ )	161	152	167	153	161	168	167	153	159	173
Peso ( $y_i$ )	63	56	77	49	72	62	68	48	57	67

A partir de la recta de regresión:  $Y = -96,1121 + 0,979009 X$ , determinar el *coeficiente de determinación*.

Primero calcularemos los **valores estimados** y los **residuos**. Es muy conveniente, disponer de los datos y los cálculos en forma de tabla para determinar **el coeficiente de determinación**:

i	$x_i$	$y_i$	$\hat{y}_i$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$\hat{y}_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$	$e_i$	$e_i^2$
1	161	63	61,51	1,10	1,21	-0,39	0,15	1,49	2,23
2	152	56	52,70	-5,90	34,81	-9,20	84,69	3,30	10,91
3	167	77	67,38	15,10	228,01	5,48	30,06	9,62	92,50
4	153	49	53,68	-12,90	166,41	-8,22	67,63	-4,68	21,87
5	161	72	61,51	10,10	102,01	-0,39	0,15	10,49	110,07
6	168	62	68,36	0,10	0,01	6,46	41,75	-6,36	40,47
7	167	68	67,38	6,10	37,21	5,48	30,06	0,62	0,38
8	153	48	53,68	-13,90	193,21	-8,22	67,63	-5,68	32,22
9	159	57	59,55	-4,90	24,01	-2,35	5,52	-2,55	6,50
10	173	67	73,26	5,10	26,01	11,36	128,97	-6,26	39,14
$\Sigma$		619			812,90		456,61		356,29

I) Tenemos que:

$$SCR = 456,61$$

$$SCT = 812,90$$

Por tanto, el coeficiente de determinación:

$$R^2 = 456,61 / 812,90 = 0.5617$$

II) A partir de la suma de los cuadrados de los residuos:

$$SCE = 356,29$$

el coeficiente de determinación:

$$R^2 = 1 - ( 356,29 / 812,90 ) = 1 - 0,4383 = 0.5617$$



# Coeficiente de determinación

III) Otra forma

$$R^2 = \frac{b^2 S^2_X}{S^2_Y}$$

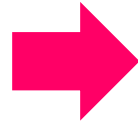
$$R^2 = 0.979009^2 \cdot 52.93 / 90.32 = \mathbf{0.5617}$$

IV) Otra forma  $r^2 = 0.749444476^2 = \mathbf{0.5617} = R^2$

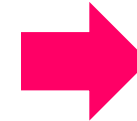
**Interpretación:** el **56.17 %** de la variabilidad de los pesos está explicada por el modelo de regresión lineal simple, ie, por la altura de las personas. El **42.84%** se debe a otros factores no contemplados.

# Analizar el ajuste lineal

Para evaluar la  
bondad del  
ajuste

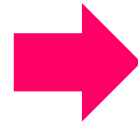


*coeficiente de  
determinación*  
 **$R^2$**

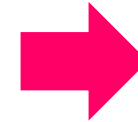


$$R^2 = \frac{b^2 S^2_X}{S^2_Y}$$

Para medir la  
dispersión de  
los puntos  
alrededor de la  
recta estimada,



*El Desvío  
residual*  
 **$S$**

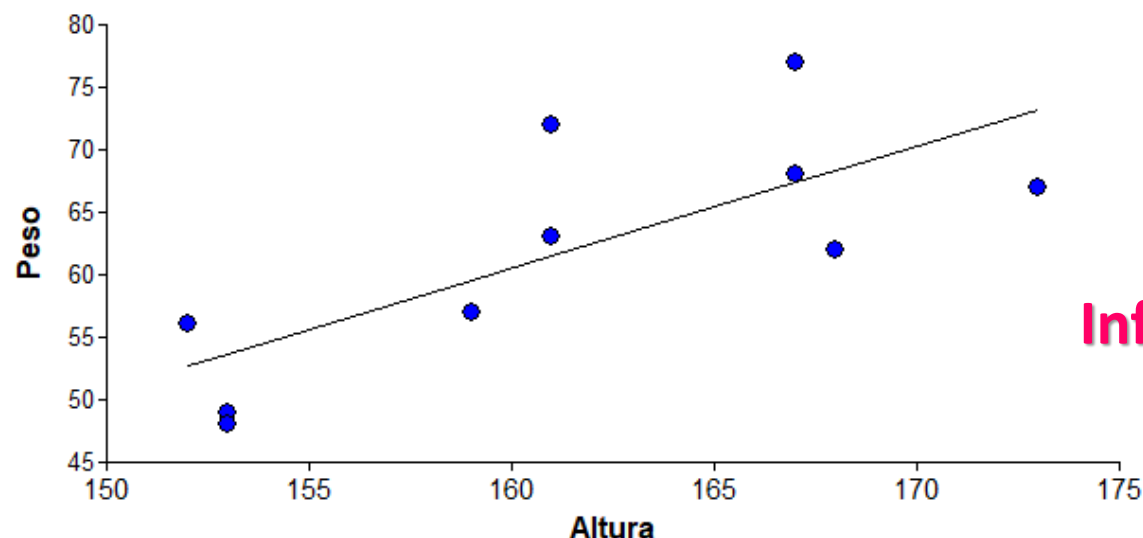


$$s = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-2}}$$



¿Cuál uso?

# Ejemplo (altura-peso) Infostat



**Variable explicativa:**

**X=** Altura de una persona (cm)

**Variable a predecir:**

**Y =** Peso de una persona (kg)

**Infostat:** Estadísticas <<Medidas resumen



## Análisis de regresión lineal

Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	ECMP	AIC	BIC
Peso	10	0,56	0,51	70,26	70,11	71,02

## Coefficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI(95%)	LS(95%)	T	p-valor	CpMallows	VIF
const	-96,11	49,39	-210,01	17,79	-1,95	0,0876		
Altura	0,98	0,31	0,27	1,68	3,20	0,0126	10,25	1,00

**Infostat:** Estadísticas << Regresión lineal

# Relación entre $R^2$ y $r$



## Importante!!!

Diferencia entre el coeficiente de correlación y el coeficiente de determinación:

- $R^2$ : mide la proporción de variación de la variable dependiente explicada por la variable independiente o por el modelo.
- $r$ : mide el grado de asociación entre las dos variables .

**En regresión lineal simple  $R^2 = r^2$**

**OJO!!!!!!**

El coeficiente de correlación se interpreta sólo si  $X$  e  $Y$  son v.a.



## Observación Importante!!

Al igual que el coeficiente de **correlación muestral**,  **$r$** , la pendiente de la recta de ajuste,  **$b$** , es **positiva** cuando  **$X$**  e  **$Y$**  están positivamente correlacionadas y **negativa** si  **$X$**  e  **$Y$**  están negativamente correlacionadas.

La diferencia entre ambos es que  **$r$**  es **adimensional** mientras que la pendiente  **$b$** , se mide en **unidades** de  **$Y$**  sobre las **unidades** de  **$X$** . Por lo tanto, **su valor** por sí mismo **no** indica si la dependencia es **débil** o **fuerte**. Depende de las unidades, de la escala de  **$X$**  e  **$Y$** .

$$b = \frac{S_{xy}}{S_x^2} = r \frac{S_y}{S_x}$$

# Fórmulas de la Regresión Lineal Simple:

$$b = \frac{S_{xy}}{S_x^2} \quad a = \bar{y} - b \bar{x}$$

Pendiente y ordenada al origen de la recta de ajuste.

$$SCT = S_Y^2 (n-1)$$

Suma de cuadrados total.

$$SCE = S_Y^2 (n-1) - b S_{XY} (n-1)$$

Suma de cuadrados error

$$SCR = S_Y^2 (n-1) - SCE$$

Suma de cuadrados de la regresión

$$s^2 = \frac{SCE}{n-2} = \frac{S_Y^2 (n-1) - b S_{XY} (n-1)}{n-2}$$

Varianza residual

Coeficiente de correlación

$$r = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}}$$

$$R^2 = \frac{b^2 S_X^2}{S_Y^2} = r^2$$

Coeficiente de determinación

