MODELADO DE SISTEMAS INFORMÁTICOS
Modelado empírico y Modelado físico

El objetivo de un modelo es reproducir (predecir) el comportamiento (funcionamiento) de un sistema bajo unas condiciones predefinidas

Un modelo "condensa" todo el conocimiento disponible sobre el sistema

Cualquier análisis o evaluación que se realice sobre un sistema informático se basa en una concepción abstracta, implícita o explícita, de cómo debe funcionar el sistema

Esta concepción abstracta se puede formalizar en forma de MODELO
De ahí, la gran importancia de los modelos en el análisis de sistemas

TIPOS DE MODELOS

EMPÍRICOS Basados exclusivamente en observaciones experimentales
Sólo modelan la relación entre las variables de entrada y las de salida

FÍSICOS Basados en caracterizar las propiedades físicas de los componentes del sistema y su interconexión (arquitectura del sistema)

En este tema $\xrightarrow[\text{técnicas básicas de}]{\text{Se presentan las}}$ Modelado Empírico

Modelos de regresión

Técnicas aplicables a series de datos con tendencias y valores estacionarios

Definición:

Un modelo de regresión permite estimar o predecir una variable aleatoria en función de otras variables

La variable estimada se denomina "**variable respuesta**" y las variables a partir de las que se estima: "**variables predictoras**", "predictores" o "factores"

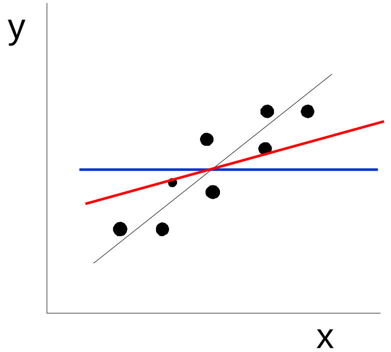
Los modelos de regresión más simples son los que presuponen una relación lineal entre la variable respuesta y las variables predictoras, son los **modelos de regresión lineal**

Si además se considera que solamente existe una variable predictora tendremos: **modelos de regresión lineal simple**

MODELADO EMPÍRICO DE SISTEMAS
Modelos de regresión lineal simple

Con esta técnica, se ajustan los puntos mediante una recta:

$$\hat{y} = b_0 + b_1 x$$



Donde:
$\hat{y}$ es el valor predicho
x es la variable predictora
b_0 y b_1 son los parámetros de regresión
y es el valor medido

Se ajusta la recta de forma que la suma de los errores cometidos sea nula si el error (o residuo) es $e_i = y_i - \hat{y}_i \Rightarrow \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - b_0 - b_1 x) = 0$

De entre todas las líneas que cumplen la condición anterior, se toma aquella que minimiza la suma de los cuadrados de los errores (SSE), este criterio se denomina, criterio de mínimos cuadrados

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x)^2$$

MODELADO EMPÍRICO DE SISTEMAS
Modelos de regresión lineal simple

Los parámetros de regresión se calculan:

$$\sum xy = \sum_{i=1}^{n} x_i y_i$$

$$b_0 = \bar{y} - b_1 \bar{x} \qquad b_1 = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$\sum x^2 = \sum_{i=1}^{n} x_i^2$$

Donde:

$\bar{x}$ e $\bar{y}$ Son los valores medios de la variable predictora y la variable respuesta.

Ejemplo. El número de E/S a disco y los tiempos de procesador medidos para siete programas, fueron: (14, 2), (16, 5), (27, 7), (42, 9), (39, 10), (50, 13), (83, 20) Se puede desarrollar un modelo lineal para predecir el tiempo de CPU en función del número de E/S a disco de la forma siguiente:

Tenemos:

nº de datos, $n = 7$; $\Sigma xy = 3375$; $\Sigma x = 271$; $\Sigma x^2 = 13885$; $\Sigma y = 66$; $\Sigma y^2 = 828$;

$\bar{x} = 38.71$; $\bar{y} = 9.43$;

Con estos datos y sustituyendo en las fórmulas de b_0 y b_1 nos quedarán:

$b_0 = -0.0083$; $b_1 = 0.2438$;

Modelos de regresión lineal simple

Coeficiente de determinación: R^2, es una medida de la calidad de la regresión ($0 \leq R^2 \leq 1$)

Se calcula:

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$$

Suma total de cuadrados

$$SST = \sum_{i=1}^{n} \left(y_i - y\right)^2 = \left(\sum_{i=1}^{n} y_i^2\right) - ny^2 = SSY - SS0$$

Se tiene:

$$SST = SSR + SSE$$

Parte explicada por la regresión

Parte **NO** explicada por la regresión

Entonces:

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST}$$

$R^2 = 1$, ajuste perfecto

$R^2 = 0$, ningún ajuste

Grados de libertad de las sumas de cuadrados

GL = Numero de elementos independientes de información usados para estimar un parámetro

GL(*SSE*) = n-2 Porque se obtiene después de calcular 2 parámetros a partir de n datos

GL(*SSY*) = n Porque se obtiene de n datos independientes sin estimar ningún parámetro

GL(*SS0*) = 1 Porque se obtiene simplemente de la media de las observaciones

GL(*SST*) = n-1 Porque hay que calcular la media para obtenerlo

Los grados de libertad se suman igual que las sumas de cuadrados correspondientes:

$$SST = SSY - SS0 = SSR + SSE$$

(n-1) = n - 1 = 1 + (n-2)

Intervalos de confianza para las predicciones

Se define la desviación estándar de los errores: $s_e = \sqrt{\frac{SSE}{n-2}}$

Se calcula la desviación estándar para las predicciones de m observaciones futuras:

$$s_{\hat{y}mp} = s_e \left[\frac{1}{m} + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum x^2 - n\bar{x}^2} \right]^{1/2}$$

Casos particulares:
m = 1 → (1/m) = 1 y
m = ∞ → (1/m) = 0

El intervalo de confianza con la muestra usada para calcular la regresión será:

$$Intervalo = \hat{y}_p \pm t_{1-\alpha/2;n-2} \cdot s_{\hat{y}mp}$$

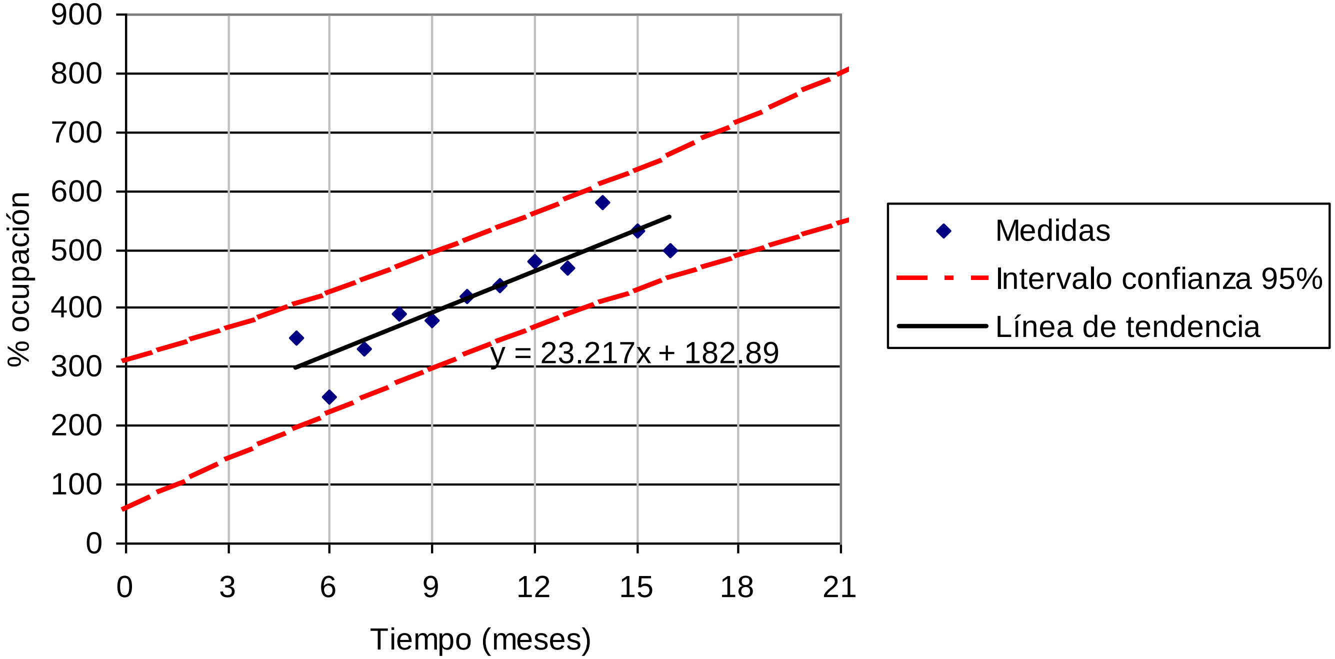
Valor predicho por el modelo de regresión

$$\hat{y}_p = b_0 + b_1 x$$

Distribución t para un nivel de confianza 100(1-α)% y n-2 grados de libertad

Modelos de regresión lineal simple

Ejemplo:

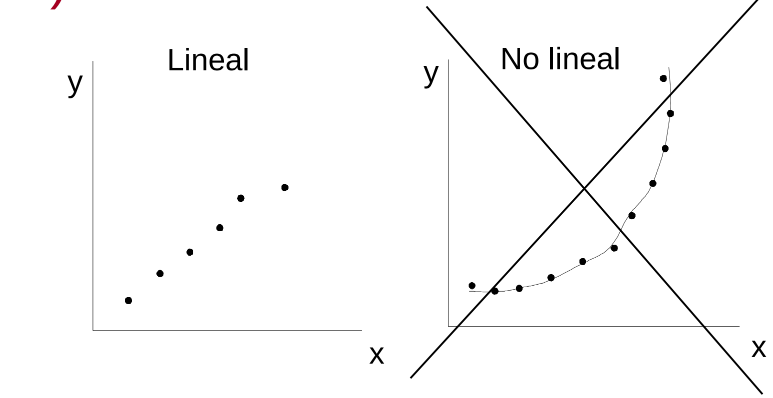


Los intervalos son más estrechos en la parte central y se abren a medida que nos alejamos del centro.

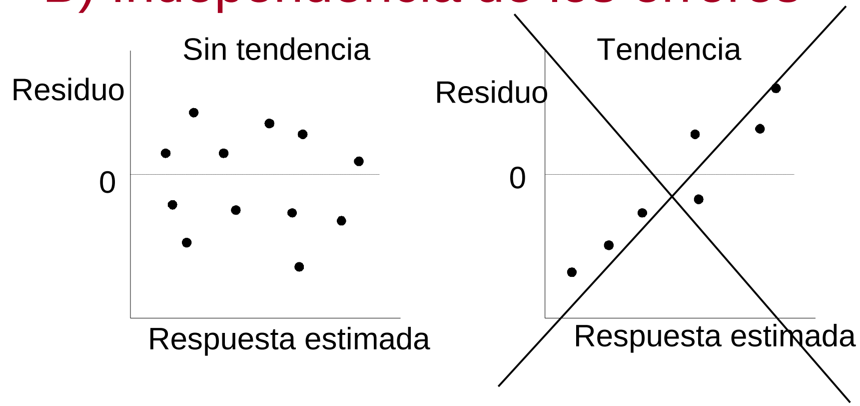
MODELADO EMPÍRICO DE SISTEMAS
Modelos de regresión lineal simple

Comprobación visual de la idoneidad del modelo de regresión lineal

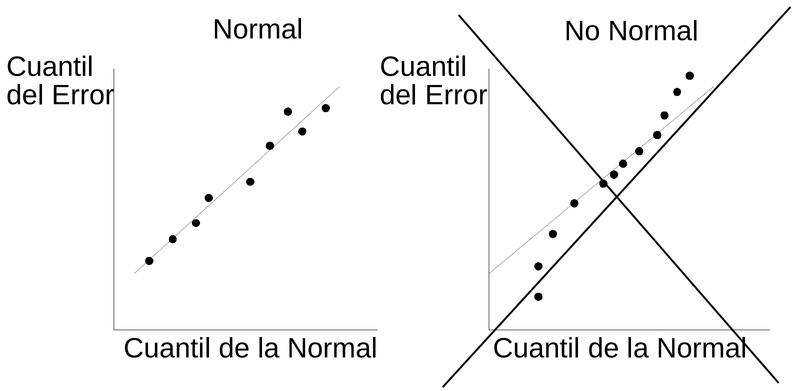
A) Relación lineal



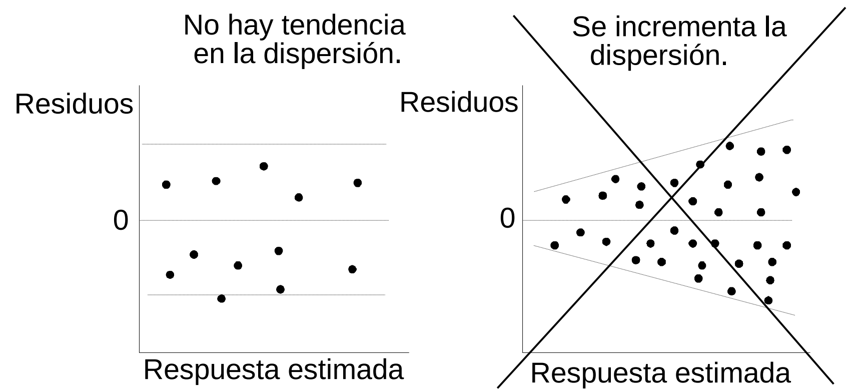
B) Independencia de los errores



C) Errores distribuidos normalmente



D) Desviación de errores constante



MODELADO EMPÍRICO DE SISTEMAS
Modelos de regresión lineal multivariable

Modelos de regresión lineal multivariable: se utiliza más de una variable predictora

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k + e \Longrightarrow [y] = [X] \cdot [b] + [e]$$

$$\begin{aligned} y_1 &= b_0 + b_1 x_{11} + b_2 x_{21} + \cdots + b_k x_{k1} + e_1 \\ y_2 &= b_0 + b_1 x_{12} + b_2 x_{22} + \cdots + b_k x_{k2} + e_2 \\ &\vdots \\ y_n &= b_0 + b_1 x_{1n} + b_2 x_{2n} + \cdots + b_k x_{kn} + e_n \end{aligned} \qquad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ & \vdots & & & \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

y = Vector columna con n observaciones de Y
X = Matriz de n filas · k+1 columnas con n·k observaciones de X
b = Vector columna con k+1 parámetros a estimar
e = Vector columna con n errores de estimación

Estimación de parámetros: $b = (X^T X)^{-1} (X^T y)$

MODELADO EMPÍRICO DE SISTEMAS
Regresión curvilínea

Las técnicas de regresión lineal solo se pueden usar si observa una relación lineal entre los predictores y la respuesta al representarlos gráficamente

Si la función NO lineal observada se puede transformar en una función lineal entonces se pueden aplicar las técnicas de regresión lineal (simple o múltiple)
Y a esta regresión se la denomina regresión curvilínea

EJEMPLOS DE RELACIONES NO LINEALES QUE PUEDEN TRANSFORMARSE EN LINEALES

NO lineal	Lineal
$y = a + b/x$	$y = a + b(1/x)$
$y = 1/(a + bx)$	$(1/y) = a + bx$
$y = x/(a + bx)$	$(x/y) = a + bx$
$y = ab^x$	$\ln y = \ln a + (\ln b)x$
$y = a + bx^n$	$y = a + b(x^n)$

MODELADO EMPÍRICO DE SISTEMAS
Utilidades de regresión en EXCEL

La hoja de cálculo EXCEL permite trabajar con modelos de regresión lineal de forma sencilla: Herramientas → Análisis de datos → Regresión
Se puede realizar el análisis de regresión lineal simple de los datos

---

Aparte provee **funciones** para calcular coeficientes de regresión:

- ESTIMACION.LINEAL → Coeficientes de la recta de regresión
- ESTIMACION.LOGARITMICA → Coeficientes que mejor describen los datos según una curva exponencial, calculado mediante un análisis de regresión
- TENDENCIA → Permite predecir los valores que resultan de una regresión lineal simple

---

Por último, en el **análisis gráfico** permite calcular líneas de tendencia:

- Lineal → se supone una relación lineal
- Logarítmica → Log y = Log x + C
- Polinomial → Mediante un polinomio de grado ≥ 2
- Potencial → relación en forma de potencia, $y = x^a$
- Exponencial → relación en forma exponencial, $y = e^{ax}$