MÓDULO 2: MÉTODOS ESTADÍSTICOS PARA IA

OBJETIVO GENERAL: INTRODUCIR LOS FUNDAMENTOS ESTADÍSTICOS NECESARIOS PARA ANALIZAR DATOS Y EVALUAR MODELOS DE INTELIGENCIA ARTIFICIAL, FAVORECIENDO LA TOMA DE DECISIONES BASADAS EN EVIDENCIA.

- CLASE 1: NOCIONES DE PROBABILIDAD. ESTADÍSTICA DESCRIPTIVA.
- CLASE 2: TEST DE HIPÓTESIS.
- CLASE 3: REGRESIÓN.
- CLASE 4: ESTADÍSTICA NO PARAMÉTRICA.

MÓDULO 2: MÉTODOS ESTADÍSTICOS PARA IA

OBJETIVO GENERAL: INTRODUCIR LOS FUNDAMENTOS ESTADÍSTICOS NECESARIOS PARA ANALIZAR DATOS Y EVALUAR MODELOS DE INTELIGENCIA ARTIFICIAL, FAVORECIENDO LA TOMA DE DECISIONES BASADAS EN EVIDENCIA.

- CLASE 1: NOCIONES DE PROBABILIDAD. ESTADÍSTICA DESCRIPTIVA.
- CLASE 2: TEST DE HIPÓTESIS.
- **CLASE 3: REGRESIÓN.**
- CLASE 4: ESTADÍSTICA NO PARAMÉTRICA.

Clase 3 - Parte 1: Métodos Estadísticos para IA

Regresión Lineal Simple

Objetivos de esta presentación

- En esta presentación aprenderemos el modelo de **regresión lineal simple** y algunos resultados obtenidos a partir de un modelado probabilístico:
  - **Modelo lineal con error de medición.**
  - **Propiedades del estimador de mínimos cuadrados.**
  - **Esperanza y varianzas estimadas de la pendiente y la ordenada al origen.**
- Realizar un test de hipótesis sobre la pendiente en un modelo lineal simple.

Introducción

- En muchos problemas existe una relación entre dos o más variables, y resulta de interés estudiar la naturaleza de esa relación.

- Esto podría suceder por ejemplo con los resultados medidos al realizar diferentes experimentos en el área de ingeniería, física, química, economía, climatología, entre otras.

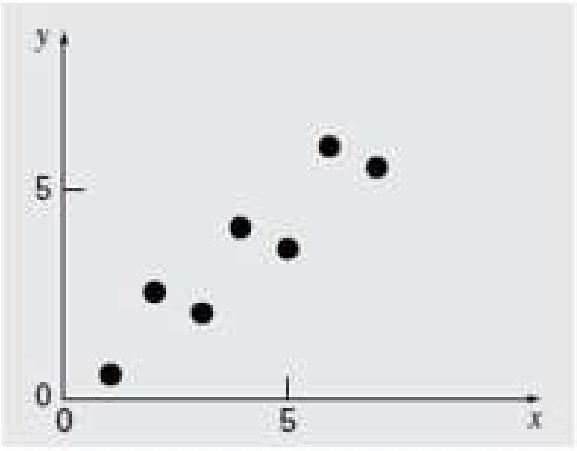
- El ***análisis de regresión*** es la técnica estadística para el modelado y la investigación de la relación entre dos o más variables.

- Entre otras cosas, el ***análisis de regresión*** permite obtener conclusiones acerca del comportamiento del sistema bajo estudio en aquellas situaciones que no han sido medidos directamente.

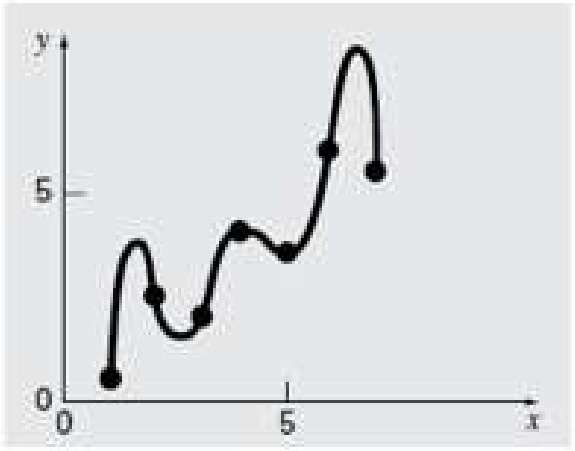
Introducción

- Cuando los datos tienen errores asociados, la interpolación polinomial es inapropiada y puede dar resultados poco satisfactorios para predecir valores intermedios.

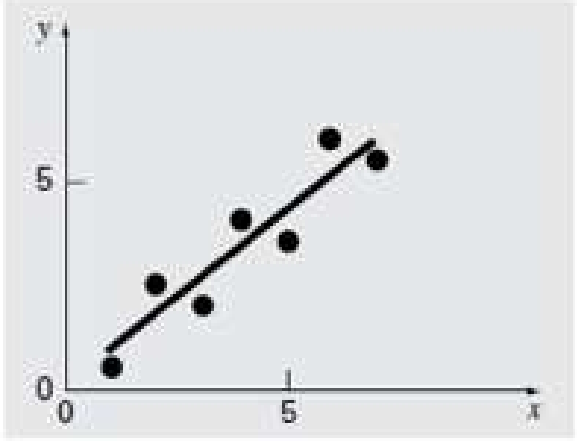
- Una estrategia más apropiada en tales casos consiste en obtener una función de aproximación que se ajuste a la forma o a la **tendencia general de los datos**, sin coincidir necesariamente en todos los puntos.



Datos con una variabilidad significativa



Polinomio de interpolación oscilando más allá del rango de los datos



Ajuste por mínimos cuadrados

Introducción: Ajuste de curvas

- Dado un número finito *n+1* de puntos conocidos en la forma de par ordenado (*x*,*y*), obtenidos a partir de un experimento,
  es decir: (x_0, y_0), (x_1, y_1),....., (x_n, y_n), siendo las x_k distintas entre sí (*k*=0,1,2,…n).

Ajustar una curva a los datos experimentales consiste en determinar una fórmula *y=f(x)* que relacione las variables *x* e *y*.

En general, se disponen de fórmulas previamente establecidas y se deben hallar los valores más adecuados de los coeficientes.

Introducción: Ajuste de curvas

Existen diferentes métodos de Ajuste de Curvas, cada uno con su correspondiente fórmula preestablecida, entre ellos:

- ***Ajuste por regresión lineal en mínimos cuadrados*** ($y=f(x)=Ax+B$).

- ***Ajuste por polinomio de grado m*** ($y=f(x)=A_m x^m+ \ldots + A_1 x + A_0$) **.**

- ***Ajuste Potencial*** ($y=Ax^M$).

- ***Ajuste Exponencial*** ($y=Ce^{Ax}$).

- ***Ajuste Cociente o Crecimiento Saturado*** $y = \frac{ax}{b+x}$

Resumen

Modelos determinísticos y probabilísticos

Método de mínimos cuadrados

Test de hipótesis

Resumen

Modelos determinísticos y probabilísticos

Método de mínimos cuadrados

Test de hipótesis

Ejemplo 16

- Supongamos que estamos interesados en verificar un modelo simple como el movimiento rectilíneo uniforme (MRU). Es decir, un móvil a velocidad constante. La mecánica elemental (Física 1) nos brinda una relación funcional entre la variable tiempo (x) y la variable posición (y):

$$y = \beta_0 + \beta_1 x$$

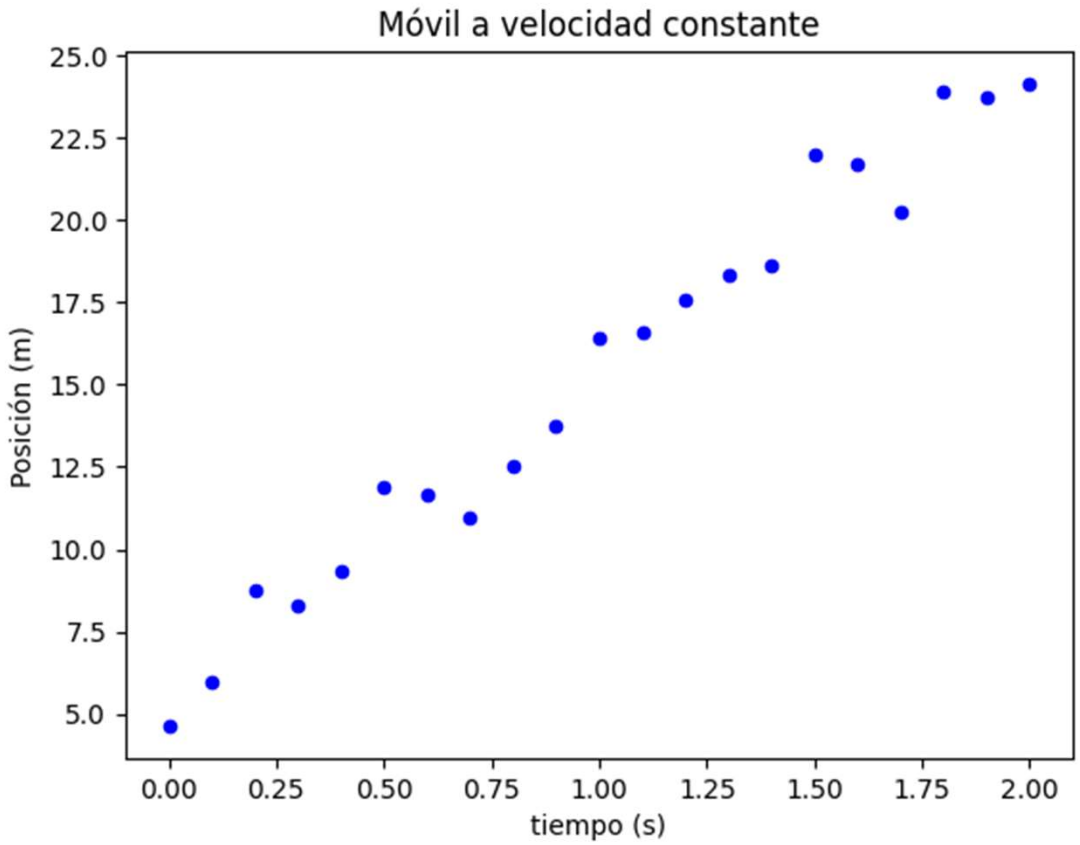
  donde β_0 es la posición inicial del móvil y β_1 es la velocidad inicial.

- Lo que significa que si conocemos el tiempo (**variable independiente**) podremos predecir la posición (**variable dependiente**).
- Si tenemos que realizar un experimento sobre esto, tendríamos que medir la posición para varios tiempos conocidos[1].

[1]A partir de un video se puede hacer con el software tracker https://physlets.org/tracker/

Ejemplo 16

- En el script ejemplo 16 se cargan datos de una medición del experimento mencionado.
- Se muestran 21 mediciones realizadas en el intervalo de 0 a 2 segundos (con un tiempo de muestreo de 0,1 segundos).




```
# Cargo los datos
archivo = "datosPosicion.csv"
datosPos = pd.read_csv(archivo)
print(datosPos.head())
print(f'Cantidad de datos n: {len(datosPos["tiempo"])}')
print(f'Media de la posición: {np.mean(datosPos["posicion"])}')
```

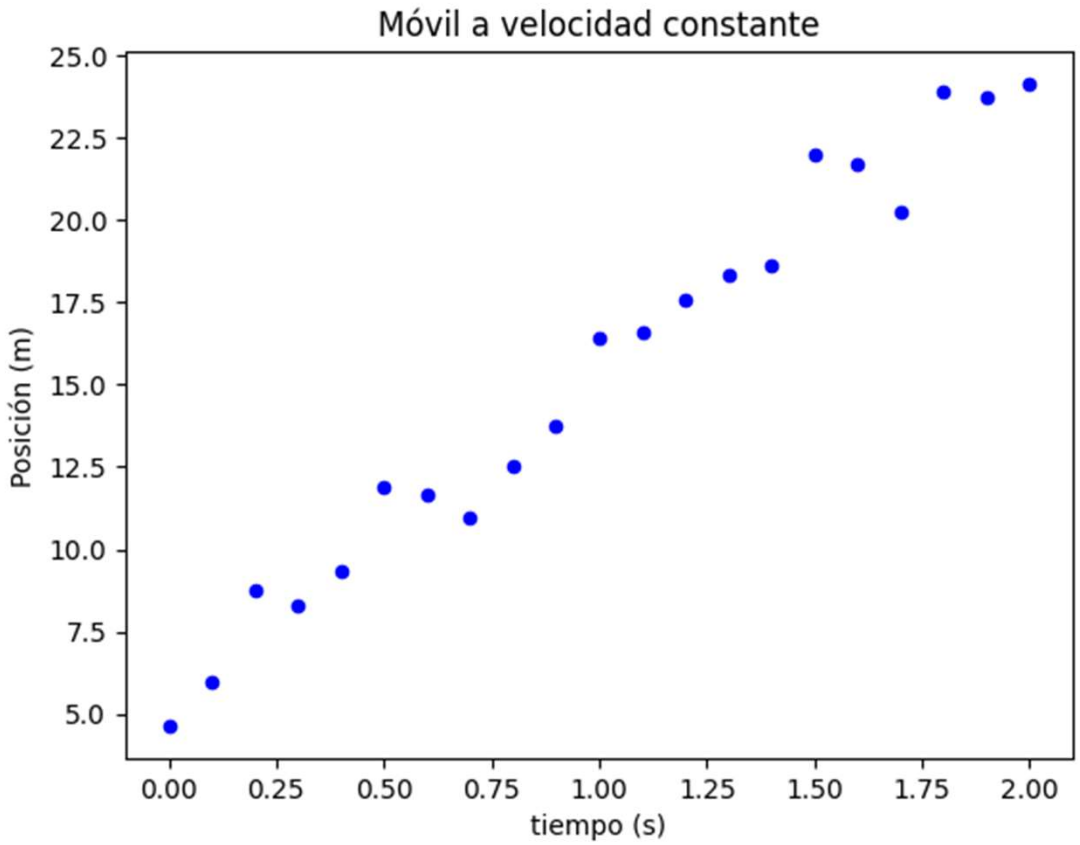
Ejemplo 16

- Todos los puntos no caen exactamente en una recta, pero el comportamiento lineal es evidente. Existe un error de medición (ruido aditivo), que se puede ver como una variable aleatoria (modelo **probabilístico**):

$$Y = \underbrace{\beta_0 + \beta_1 x}_{\text{determinístico}} + \underbrace{E}_{\text{probabilístico}} \qquad (1)$$

donde E es una variable aleatoria.

- Debe notarse que el tiempo x se escribe con minúscula porque lo conozco (variable **determinística**).



Comentarios

- Notemos que este es un modelo que proviene de una relación **física (teórica)** que depende de una sola variable independiente.
- Se cumplen las siguientes relaciones[2]:

$$\mathbf{E}\{Y|x\} = \mathbf{E}\{\beta_0 + \beta_1 x + E\} = \beta_0 + \beta_1 x + \mathbf{E}\{E\} \tag{2}$$

$$\mathbf{V}\{Y|x\} = \mathbf{V}\{\beta_0 + \beta_1 x + E\} = \mathbf{V}\{E\} \tag{3}$$

- Se supone en general $\mathbf{E}\{E\} = 0$ y $\mathbf{V}\{E\} = \sigma^2$.

  - También es posible pensar la variable x como aleatoria. En ese caso, podríamos escribir:

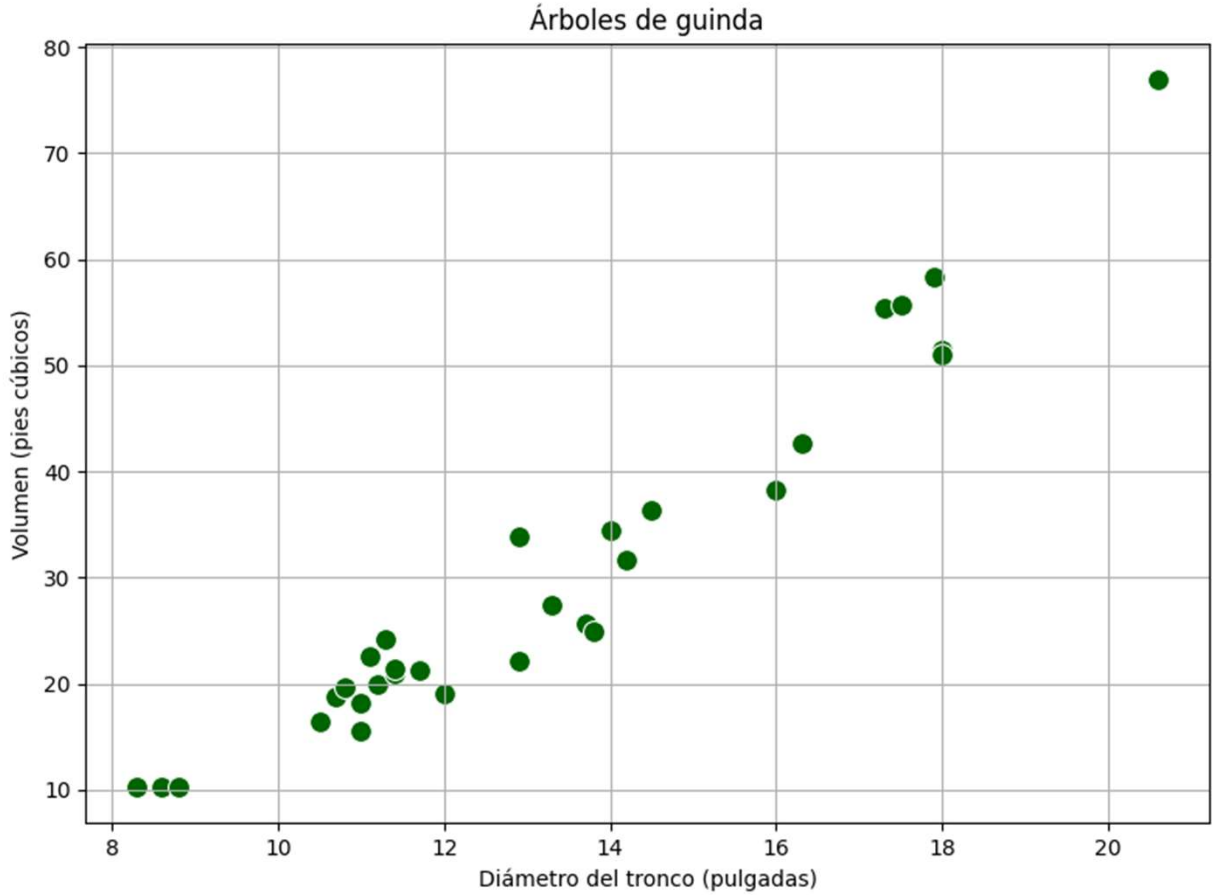
$$Y = \beta_0 + \beta_1 X$$

    Además, existen otros modelos de ruido (no sólo aditivo), pero en este curso solo nos quedaremos con la Ec. (1).

[2] $\mathbf{E}\{\{Y|x\}\}$ se debe leer como esperanza de Y dado un valor x.

Comentarios

- Si no existe una relación teórica (o es muy compleja) es posible pensar en un modelo **empírico**. En el ejemplo 17 se grafica el volumen del árbol de guinda en función del diámetro del tronco.



*A este tipo de diagramas se los denomina diagramas de dispersión y son útiles para visualizar correlaciones entre variables.

Resumen

Modelos determinísticos y probabilísticos

Método de mínimos cuadrados

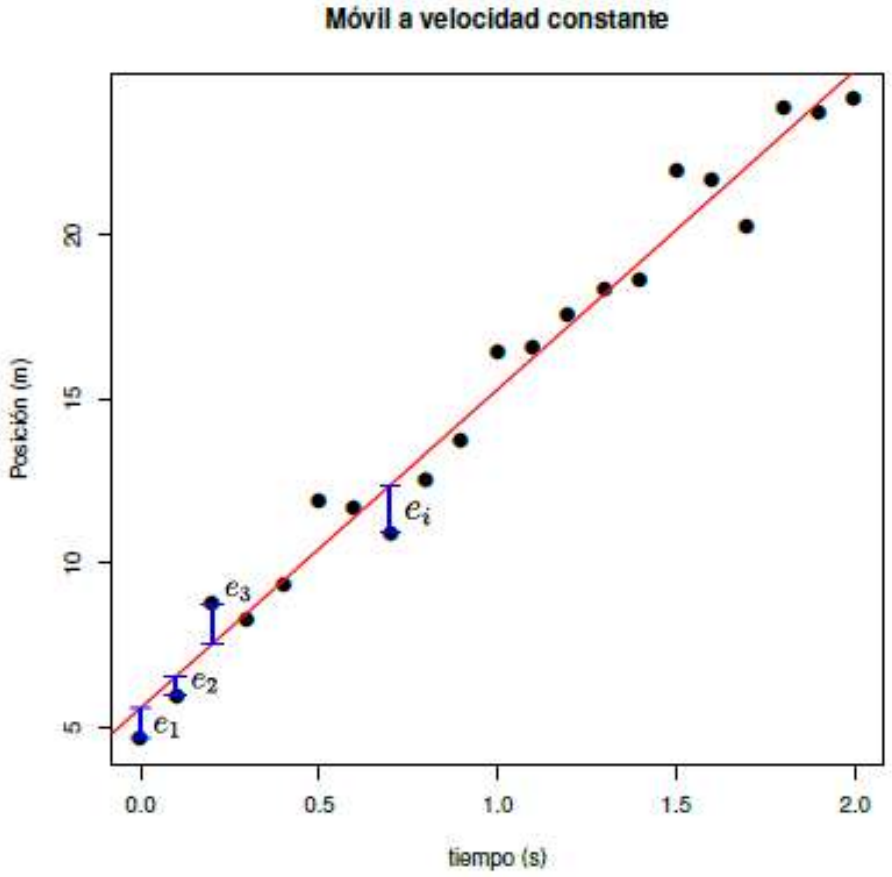
Test de hipótesis

Método de mínimos cuadrados

- Volvamos al modelo de la Ec. (1), pero supongamos que tenemos una muestra aleatoria de n pares de valores (x_i, Y_i), entonces para el i-ésimo dato la ecuación se modifica a:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \qquad (4)$$

- Estamos interesados en estimar la mejor recta ("mejor" en algún sentido) que explica los puntos del diagrama de dispersión.
- Un enfoque puede ser minimizar las distancias a la recta $e_i = y_i - (\beta_0 + \beta_1 x_i)$.
- Note que todos los términos están con minúsculas, porque representan una realización (salida) de la variable aleatoria.



Método de mínimos cuadrados

- Si sumamos todos los errores (al cuadrado para evitar cancelaciones) entonces tenemos:

$$L = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right]^2$$

- Minimizando L respecto de los parámetros β_0 y β_1 se obtienen las siguientes estimaciones:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \text{ y } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{5}$$

donde:

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \tag{6}$$

y

$$S_{xy} = \sum_{i=1}^{n} y_i (x_i - \bar{x})^2 = \sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} \tag{7}$$

Ejemplo 16: con tablas

Muestra	tiempo (X)	posición (Y)	X.Y	X^2
1	0	4,639524353	0	0
2	0,1	5,969822511	0,596982251	0,01
3	0,2	8,758708314	1,751741663	0,04
4	0,3	8,270508391	2,481152517	0,09
5	0,4	9,329287735	3,731715094	0,16
6	0,5	11,91506499	5,957532493	0,25
7	0,6	11,66091621	6,996549724	0,36
8	0,7	10,93493877	7,654457136	0,49
9	0,8	12,51314715	10,01051772	0,64
10	0,9	13,75433803	12,37890423	0,81
11	1	16,4240818	16,4240818	1
12	1,1	16,55981383	18,21579521	1,21
13	1,2	17,60077145	21,12092574	1,44
14	1,3	18,31068272	23,80388753	1,69
15	1,4	18,64415887	26,10182241	1,96
16	1,5	21,98691314	32,98036971	2,25
17	1,6	21,69785048	34,71656077	2,56
18	1,7	20,23338284	34,39675083	2,89
19	1,8	23,9013559	43,02244062	3,24
20	1,9	23,72720859	45,08169633	3,61
21	2	24,13217629	48,26435259	4
SUMA	**21**	**320,9646523**	**395,6882364**	**28,7**

$$S_{xy} = 395.69 - \frac{21 \times 320.96}{21} \approx 74.73$$

$$S_{xx} = 28.7 - \frac{21^2}{21} \approx 7.7$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{74.73}{7.7} = 9.70$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{320.96}{21} - 9.70 \times \frac{21}{21} = 5.58$$

$$y = 5.58 + 9.70x$$

Ejemplo 16: con Python

- Volviendo al ejemplo 16, calculemos las estimaciones de β_0 y β_1, debemos utilizar las ecuaciones 5, 6 y 7.

```
#Regresión lineal simple
n = len(datosPos['tiempo'])
xi = datosPos['tiempo']
yi = datosPos['posicion']
Sxx = np.sum(xi**2) - (np.sum(xi)**2) / n
Sxy = np.sum(xi * yi) - (np.sum(xi) * np.sum(yi)) / n
beta1_est = Sxy / Sxx
beta0_est = np.mean(yi) - beta1_est * np.mean(xi)
```

$\hat{\beta}_0$ = 5,58 m (posición inicial, ordenada al origen)

$\hat{\beta}_1$ = 9,70 m/s (velocidad inicial, pendiente)

- Python tiene una función que realiza una estimación lineal y algunas cosas más que veremos luego:

```
# Usando sklearn
model = LinearRegression().fit(xi.values.reshape(-1, 1), yi)
print(f'Intercept: {model.intercept_}, Slope: {model.coef_[0]}')
```

Verifique las salidas con los resultados obtenidos).

Método de mínimos cuadrados: error y residuos

- A la realización del error E_i se lo denomina **residuo**:

$$e_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x\right) \tag{8}$$

- Si se supone una distribución Normal para el error, con media 0 y varianza desconocida σ^2, ésta se puede estimar a partir de la suma de los errores cuadráticos:

$$SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = SS_T - \hat{\beta}_1 S_{xy} \tag{9}$$

donde $SS_T = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$. Si se calcula la esperanza:

$$\mathbf{E}\{SS_E\} = (n-2)\,\sigma^2$$

- Por lo tanto un estimador insesgado de σ^2 es:

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} \tag{10}$$

Se dice que un estimador es **insesgado** si la Media o Valor Esperado de la distribución del estimador es igual al parámetro.

Más comentarios y conclusiones preliminares

- La estimación de mínimos cuadrados se podría pensar de manera determinística, pero
  ¿qué nos puede ofrecer pensarlo como un problema probabilístico?
  - Se puede demostrar que si el error es Normal las estimaciones $\hat{\beta}_0$ y $\hat{\beta}_1$ de la Ec. (5) son además estimadores de máxima verosimilitud (ver pag. 487 de Referencia [2]).
  - Se puede verificar que sus esperanzas y varianzas son:

$$\mathbf{E}\{\hat{\beta}_1\} = \beta_1 \text{ y } \mathbf{V}\{\hat{\beta}_1\} = \frac{\sigma^2}{S_{xx}}$$
$$\mathbf{E}\{\hat{\beta}_0\} = \beta_0 \text{ y } \mathbf{V}\{\hat{\beta}_0\} = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right] \qquad (11)$$

  por lo tanto, son insesgados. Si se utiliza la estimación de la varianza de la Ec. (10) entonces se pueden calcular las desviaciones típicas de ambos estimadores[3]: $\sqrt{\mathbf{V}\{\hat{\beta}_1\}}$ y $\sqrt{\mathbf{V}\{\hat{\beta}_0\}}$.
  - Se pueden calcular intervalos de confianza y test de hipótesis.

[3]Verifique estas ecuaciones en el script del ejemplo 16 y compare con las salidas calculadas.

Resumen

Modelos determinísticos y probabilísticos

Método de mínimos cuadrados

Test de hipótesis

Test de hipótesis sobre β_1

- Se puede realizar un test de hipótesis sobre β_1.

$$H_0 : \beta_1 = \beta_{1,0} \text{ Hipótesis nula}$$
$$H_1 : \beta_1 \neq \beta_{1,0} \text{ Hipótesis alternativa} \tag{12}$$

para realizar el test, es necesario un estadístico de prueba con una distribución conocida. El estadístico es:

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2 / S_{xx}}}. \tag{13}$$

Para deducir la distribución, debemos notar que $\hat{\beta}_1$ es suma de variables aleatorias Y_i (ver Ec. (5)) y éstas son independientes. Entonces Y_i tiene distribución $\mathbf{N}\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$ (porque el error E es Normal) y, por lo tanto, $\hat{\beta}_1$ tiene distribución $\mathbf{N}\left(\beta_1, \sigma^2 / S_{xx}\right)$. El denominador de la Ec. (13) es una suma de Normales al cuadrado que tienen distribución Chi-cuadrado, por lo tanto T tiene distribución t-student $n - 2$ grados de libertad[4].

[4]La Ec. (10) indica que los grados de libertad de la v.a. T será $n - 2$.

Test de hipótesis en el Ejemplo 16

- Volviendo al ejemplo 16, un test de hipótesis muy habitual es determinar si exite una relación lineal entre las variables x e Y, para esto se puede proponer:

$H_0: \beta_1 = 0$ Hipótesis nula

$H_1: \beta_1 \neq 0$ Hipótesis alternativa

```
#Valor de estadístico de prueba
T0 = beta1_est / np.sqrt(varSigma_est / Sxx)
print(f'T0: {T0}')
p_valor = 2 * (1 - stats.t.cdf(T0, n - 2))
print(f'p-valor de beta1: {p_valor}')
```

- En este caso, el p-valor es muy pequeño, por lo tanto, se puede concluir que la H_0 se rechaza. Esto puede significar que existe una relación lineal entre las variables o que con un polinomio de mayor orden se obtenga un mejor resultado.

Test de hipótesis en el Ejemplo 16

- En el ejemplo 16, había una relación teórica entre ambas variables y sabíamos que sería lineal. Un test de hipótesis más interesante sería probar si $\beta_1 = 10$ m/s (es decir, si la velocidad del móvil es o no 10 m/s).

$H_0: \beta_1 = 10$ Hipótesis nula

$H_1: \beta_1 \neq 10$ Hipótesis alternativa

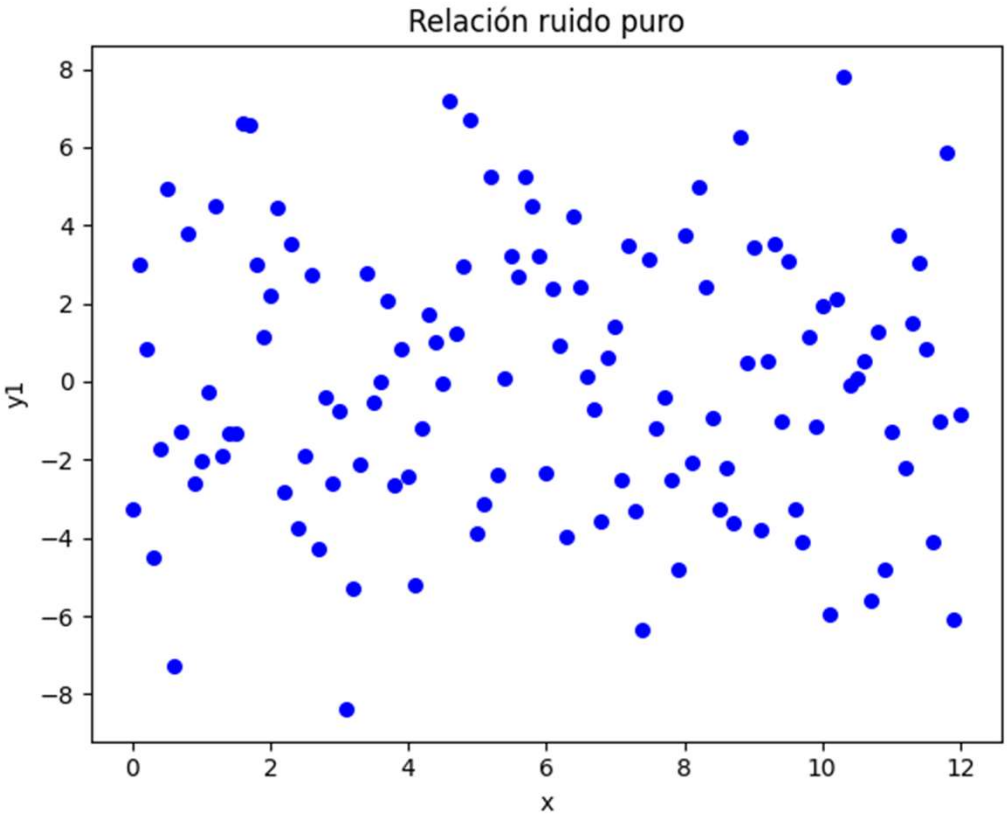
```
#Prueba de H0: beta1 = 10 o distinto
t0 = (beta1_est - 10) / np.sqrt(varSigma_est / Sxx)
print(f't0: {t0}')
p_valor = 2 * (stats.t.cdf(t0, n - 2))
print(f'p-valor de beta1: {p_valor}')
```

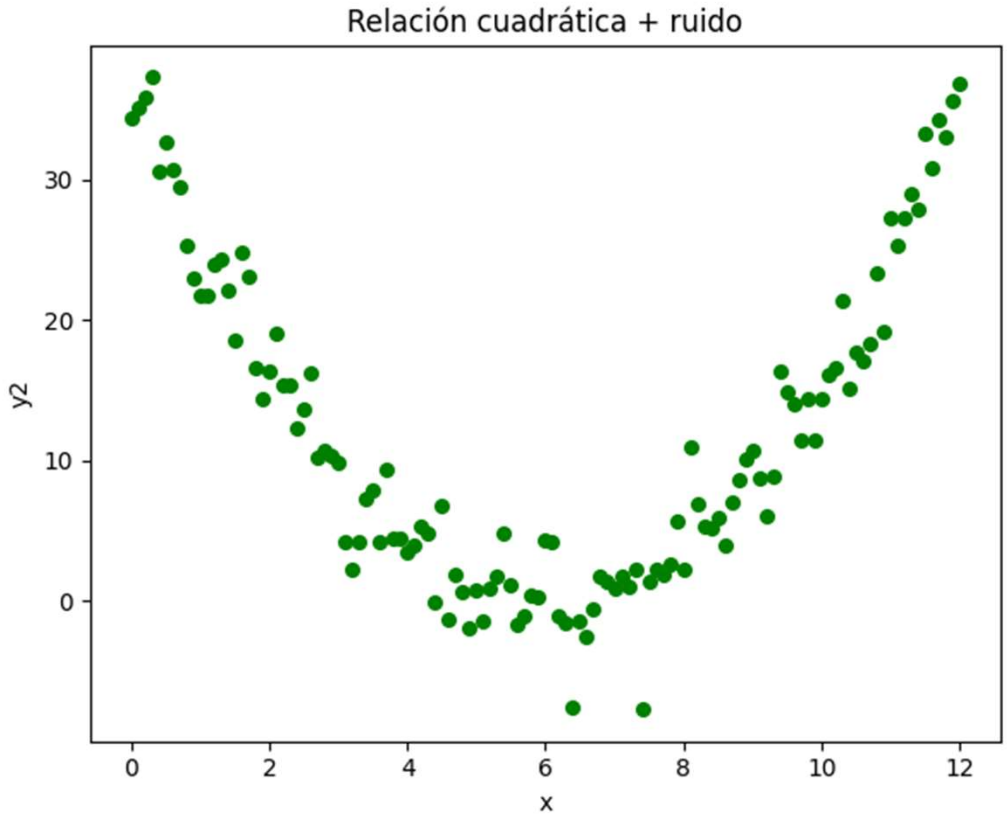
- En este caso, el p-valor es: 0,418. Por lo tanto, no podemos rechazar que la velocidad sea 10 m/s.

Test de hipótesis sobre β_1

- En el contraste de $\beta_1 = 0$ contra distinto, es posible que exista relación entre las variables estudiadas y no se rechace H_0.

- En el ejemplo 18 se ven dos casos en que se acepta H_0 y sin embargo en la figura derecha hay una relación cuadrática.





Test de hipótesis sobre β_0

- De manera similar a como se hizo con β_1, se puede plantear un test de hipótesis con la ordenada al origen. En este caso:

$$H_0 : \beta_0 = \beta_{0,0} \text{ Hipótesis nula}$$

$$H_1 : \beta_{0,0} \neq \beta_{0,0} \text{ Hipótesis alternativa}$$

- El estadístico de prueba será:

$$T = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}}. \tag{14}$$

que tendrá también distribución t-student con $n - 2$ grados de libertar.

Conclusiones de la Parte I

- El modelo probabilístico de ruido aditivo permite realizar algunas cuentas útiles a la hora de encontrar buenos estimadores de los parámetros de un regresor lineal.
- Aprendimos a hacer un test de hipótesis sobre la pendiente suponiendo independencia y distribución Normal del error.
- Definimos residuo y algunas otras expresiones útiles para el resto de la primera parte.

Bibliografía

1- D. C. Montgomery and G. C. Runger, “Applied Statistics and Probability for Engineers Third Edition”, John Wiley & Sons, Inc.

2- A. Mc. Mood, F. A. Graybill, D. C. Boes, “Introduction to the theory of Statistics”, Mc. Graw-Hill, 1974.