Laboratorio 1C

Clase 4

Regresiones – Cuadrados mínimos

30/04/22

- Supongamos que queremos medir la velocidad de un cuerpo en MRU.
- En forma **directa**, por ejemplo usando una pistola de velocidad, se obtiene:

$$v = v_o + \Delta v$$

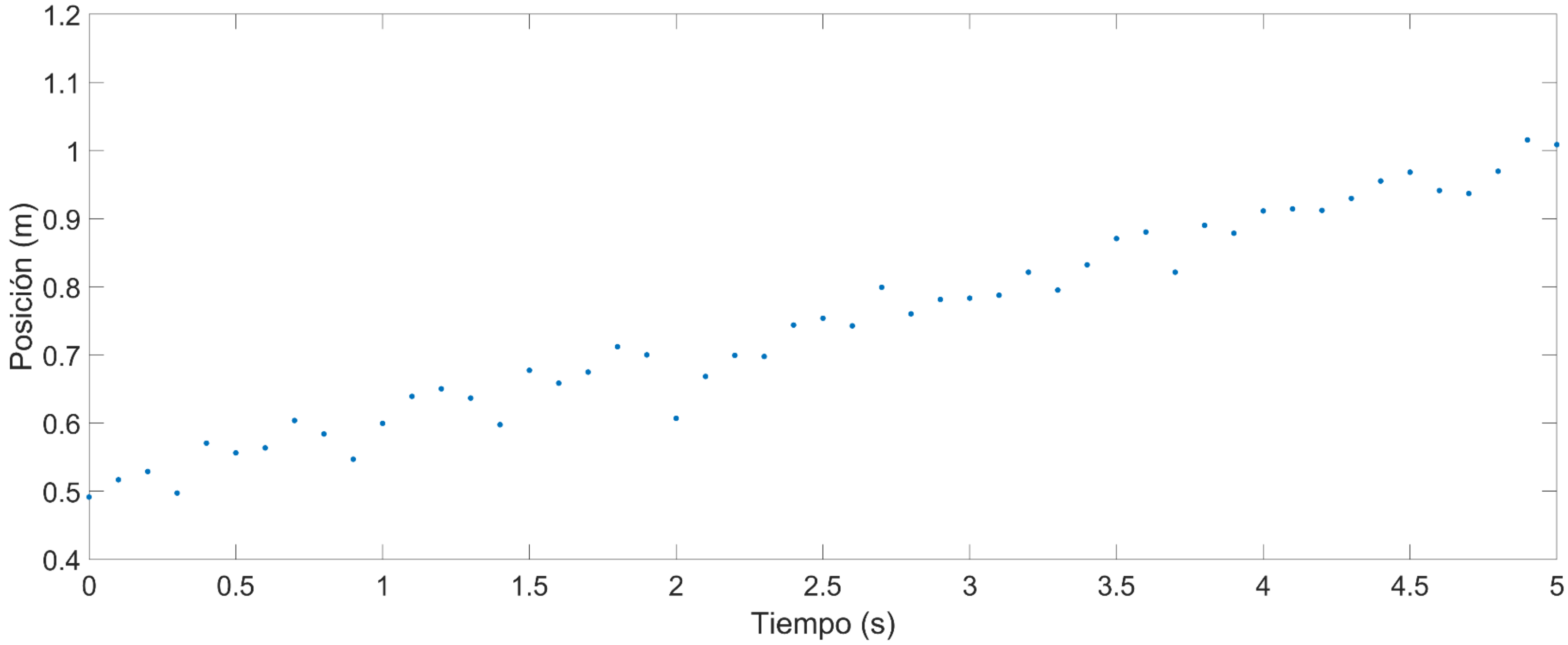
- En forma **indirecta**, midiendo la posición a dos tiempos distintos, y luego haciendo el cociente entre la distancia recorrida y el tiempo.

$$v = \frac{x_2 - x_1}{t_2 - t_1} \qquad \Delta v = \sqrt{\sum_i \left(\frac{\partial v}{\partial q_i}\right) (\Delta q_i)^2}$$

En ambos casos, podríamos medir *N* veces, y reportar la media de todas las velocidades obtenidas. El error sería entonces:

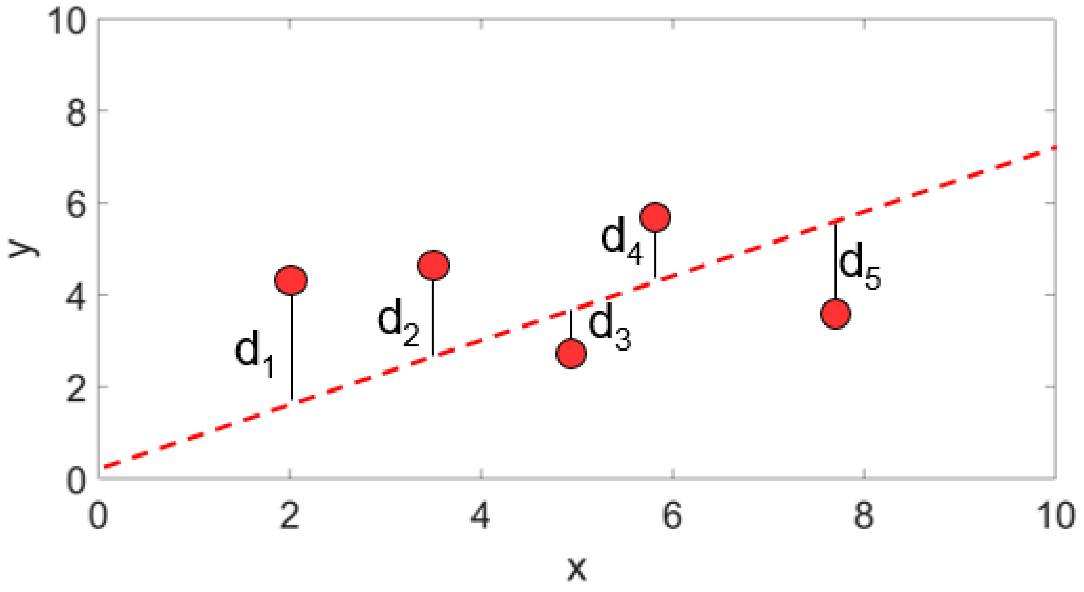
$$\Delta v_C = \sqrt{(\Delta v_A)^2 + (\Delta v_B)^2} = \sqrt{\sum_{i=1}^{N} \frac{(v_i - \overline{v})^2}{n(n-1)} + (\Delta v_B)^2}$$

- Si medimos la posición en N tiempos, típicamente vamos a ver algo así:



En un MRU: $$x(t) = x_o + v(t - t_o)$$

- ¿Cómo podemos encontrar el x y v óptimos?
- Esto es un caso de un problema general: dados puntos (x_i, y_i) ¿cuál es la recta que está "más cerca" de los puntos?



$$y = f(x) = a + bx$$

Buscamos a y b, tales que $\sum d_i^2$ sea mínima

- Queremos minimizar:

$$\sum d_i^2 = \chi^2 = \sum[y_i - (a + bx_i)]^2$$

- Que es una función de a y b (los (x_i, y_i) son las mediciones!)

$$\frac{\partial\chi^2}{\partial a} = 0 \qquad \frac{\partial\chi^2}{\partial b} = 0$$

$$\chi^2 = \sum[y_i - (a + bx_i)]^2 = \sum(y_i - a - bx_i)^2 =$$

$$\sum\left[y_i^2 + a^2 + (bx_i)^2 - 2y_i a - 2y_i bx_i + 2abx_i\right] =$$

$$\sum y_i^2 + Na^2 + b^2\sum x_i^2 - 2a\sum y_i - 2b\sum x_i y_i + 2ab\sum x_i$$

$$\chi^2 = \sum y_i^2 + Na^2 + b^2\sum x_i^2 - 2a\sum y_i - 2b\sum x_i y_i + 2ab\sum x_i$$

$$\frac{\partial \chi^2}{\partial a} = 2Na - 2\sum y_i + 2b\sum x_i = 0$$

$$\frac{\partial \chi^2}{\partial b} = 2b\sum x_i^2 - 2\sum x_i y_i + 2a\sum x_i = 0$$

Resolviendo

$$a = \frac{1}{\Delta}\left(N\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i\right)$$

$$b = \frac{1}{\Delta}\left(N\sum x_i y_i - \sum x_i \sum y_i\right)$$

$$\Delta = N\sum x_i^2 - \left(\sum x_i\right)^2$$

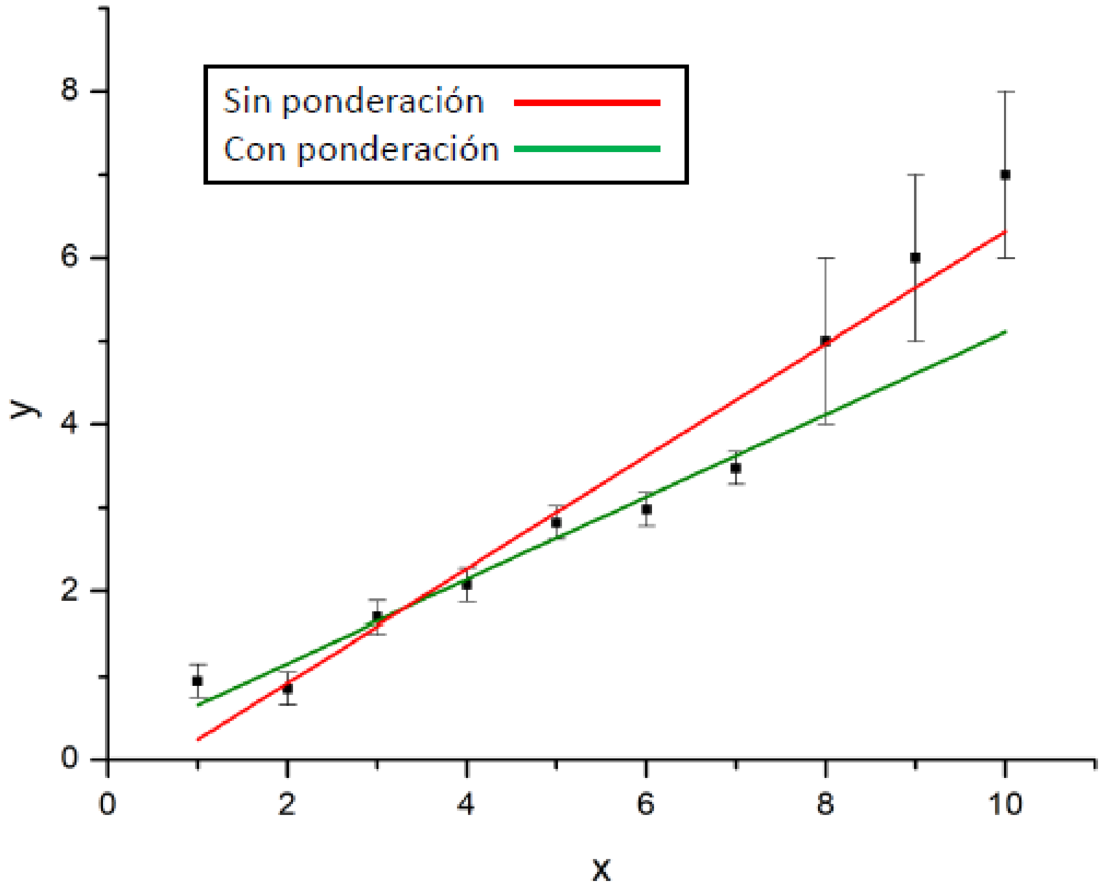
- En un caso un poco más general, consideramos errores en las mediciones de y
- Nos interesa que la recta ajuste mejor a los puntos medidos con mayor precisión.
- Esto se logra definiendo:

$$\chi^2 = \sum \left[\frac{y_i - (a + bx_i)}{\sigma_i} \right]^2$$

$$a = \frac{1}{\Delta'} \left(\sum \frac{x_i^2}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} \right) \qquad \Delta' = \sum \frac{1}{\sigma_i} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2} \right)^2$$

$$b = \frac{1}{\Delta'} \left(\sum \frac{1}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} \right)$$

- La ponderación del error mitiga la distorsión que pueden introducir las mediciones poco precisas.
- En un ejemplo donde solo consideramos el error en la variable y



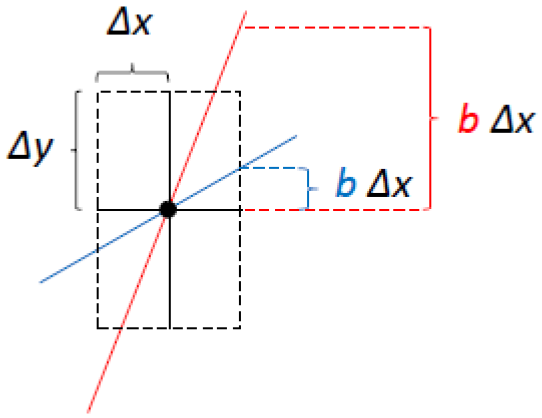
- ¿Y si el error en x no es despreciable? Hay métodos generales para errores en ambas variables, pero son más complicados.
- Lo habitual es elegir como x la variable con menor error, y despreciarlo. ¿Qué criterio se puede tomar?

$$\Delta x \leq \Delta y$$

¿Y si x e y no son la misma magnitud?

$$\frac{\Delta x}{x} \leq \frac{\Delta y}{y}$$

Mejor ¿pero no depende de más nada?



Lo mejor es elegir x e y, aplicar cuadrados mínimos para estimar b, y chequear que se satisfaga $b\,\Delta x \leq \Delta y$

Si no se cumple, podemos invertir $x = c + m\,y$

Y darlo vuelta de nuevo $y = a + b\,x \quad \Rightarrow \quad b = \frac{1}{m} \quad a = -\frac{c}{m}$

¿Qué errores tienen a y b ? En su determinación están involucradas las N mediciones de (x_i, y_i), así que se estima **propagando** el error de todas ellas.

Suponiendo que las distintas mediciones no están correlacionadas entre sí:

$$\sigma_z^2 = \sum_{i=1}^{N} \left[\sigma_i^2 \left(\frac{\partial z}{\partial y_i} \right)^2 \right] \qquad z = a, b$$

$$\sigma_a^2 = \frac{1}{\Delta'} \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2}$$

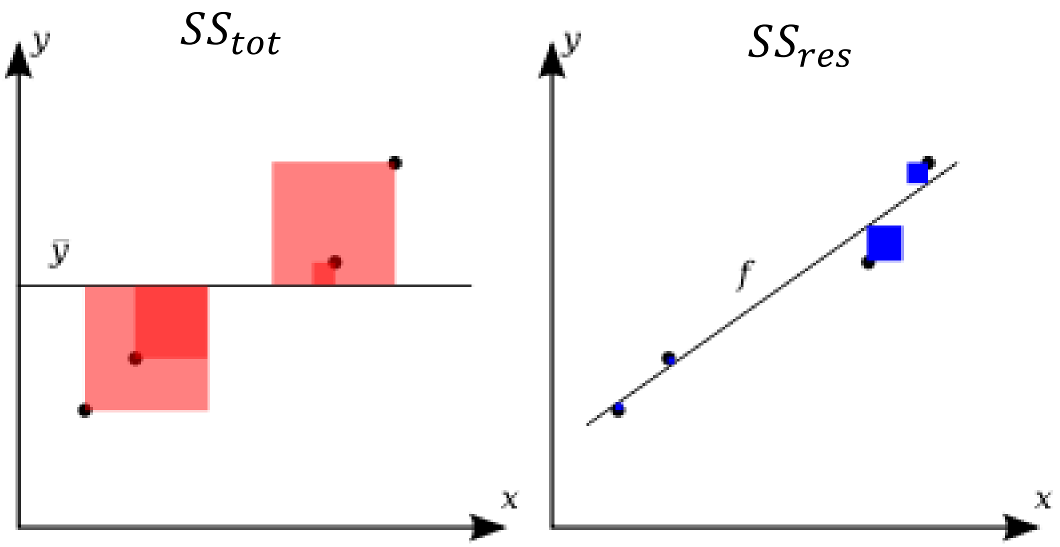
$$\Delta = \sum \frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2} \right)^2$$

$$\sigma_b^2 = \frac{1}{\Delta'} \sum_{i=1}^{N} \frac{1}{\sigma_i^2}$$

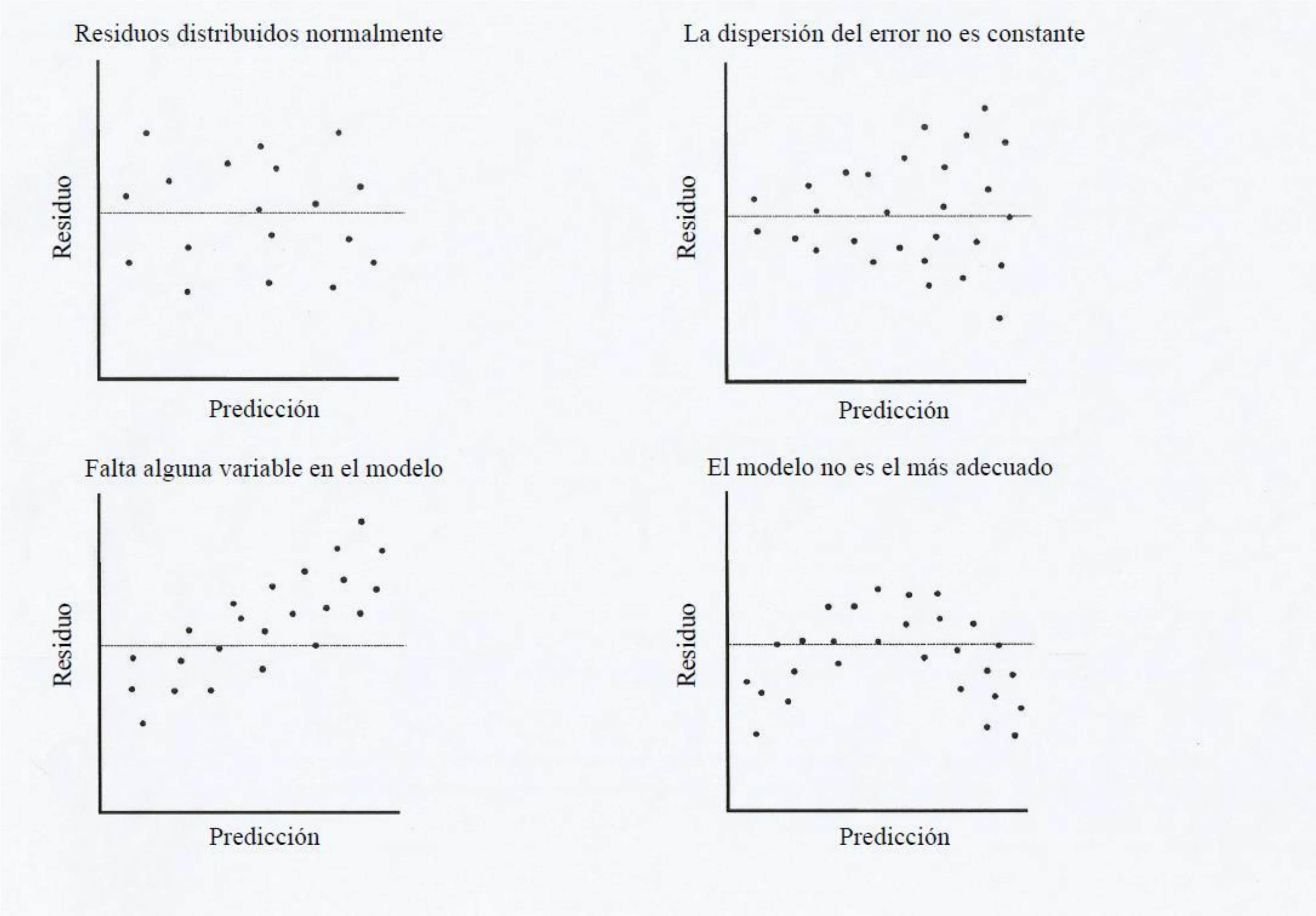
Importante! Las estimaciones de a y b no son independientes! Una fluctuación en a implica una en b, y viceversa. Esto es fundamental a la hora de propagar los errores de a y b en cálculos que los involucran.

- ¿Cómo evaluamos el ajuste?

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left[y_i - f(x_i)\right]^2}{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2} = 1 - \frac{SS_{res}}{SS_{tot}}$$



- Análisis de residuos



Diferencia entre variable observada y predicha a partir de la regresión:

$$y_i - \hat{y}_i$$

Para regresión lineal, se asume que tiene una **distribución normal**

El concepto de cuadrados mínimos se puede aplicar a todo tipo de funciones, no solamente a rectas. Hay una gran diferencia según si la función es lineal o no lineal en los parámetros que se quieren determinar.

Lineales $\longrightarrow$ $a + b\,x + c\,x^2, \quad a\,\mathrm{sen}(x), \quad a\exp(x), \quad a\log(x)$

Todo lo que vimos sigue valiendo, las ecuaciones normales son diferentes pero siempre tienen solución, y es única.

El concepto de cuadrados mínimos se puede aplicar a todo tipo de funciones, no solamente a rectas. Hay una gran diferencia según si la función es lineal o no lineal en los parámetros que se quieren determinar.

No lineales $\longrightarrow$ $a + a^2\, x, \quad \mathrm{sen}(a\, x), \quad \exp(a\, x), \quad \log(a\, x)$

Las ecuaciones normales no siempre tienen solución única, y en general no tienen solución cerrada. La suma de los cuadrados se minimiza **numéricamente** usando algoritmos de optimización.

A veces pueden convertirse en un problema lineal haciendo un cambio de variable

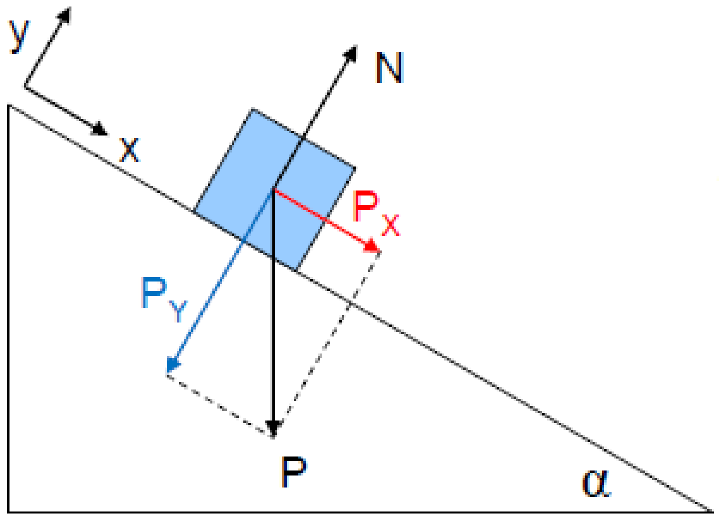
$$y = \exp(a\, x), \quad u = \ln(y) \quad \Rightarrow \quad u = a\, x$$

Experiencia

- Vamos a estudiar el movimiento de un carrito en un plano inclinado
- Medimos posición en función del tiempo
- Determinamos la aceleración para distintas inclinaciones del riel
- A partir de los datos, calculamos la aceleración de la gravedad

Ecuación horaria

- Vamos a suponer que no hay rozamiento (riel / aire)
- Suponemos movimiento 1D

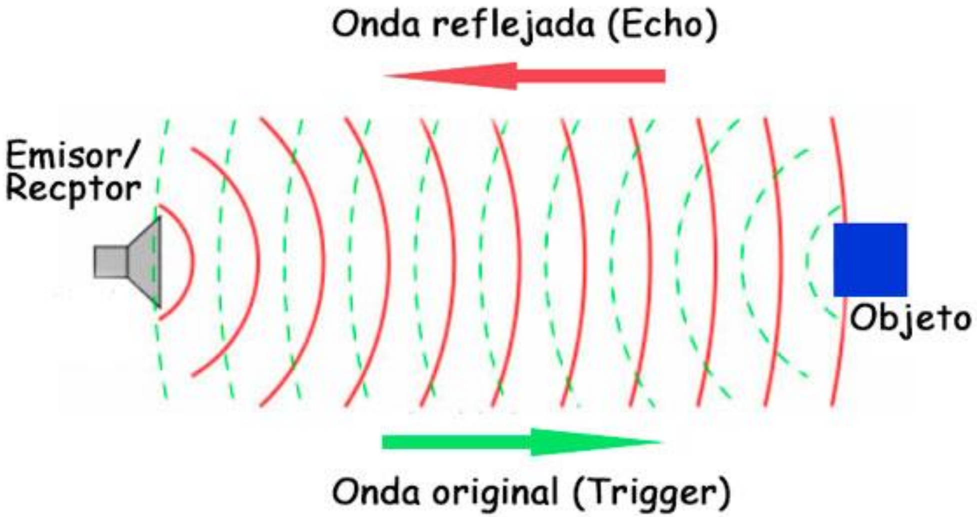


$$\hat{x})\ m\ddot{x} = P_x = mgsen(\alpha)$$

$$\hat{y})\ m\ddot{y} = N - P_y = 0$$

$$\ddot{x} = gsen(\alpha) \rightarrow x(t) = x_o + v_o(t - t_o) + \frac{1}{2}a(t - t_o)^2$$

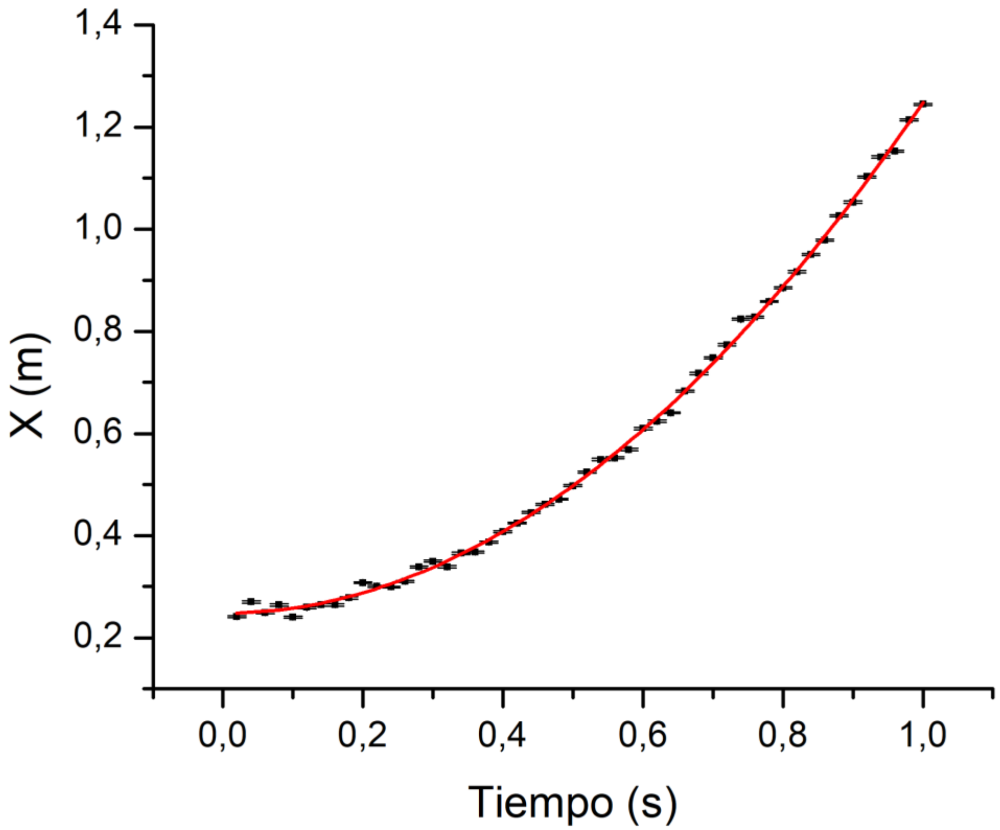
¿Y cómo medimos la posición?



- Mide distancia según el tiempo de viaje de un pulso de ultrasonido
- Cuidado con el rango, no funciona bien a distancia corta
- Requiere calibración

https://www.vernier.com/product/motion-detector/

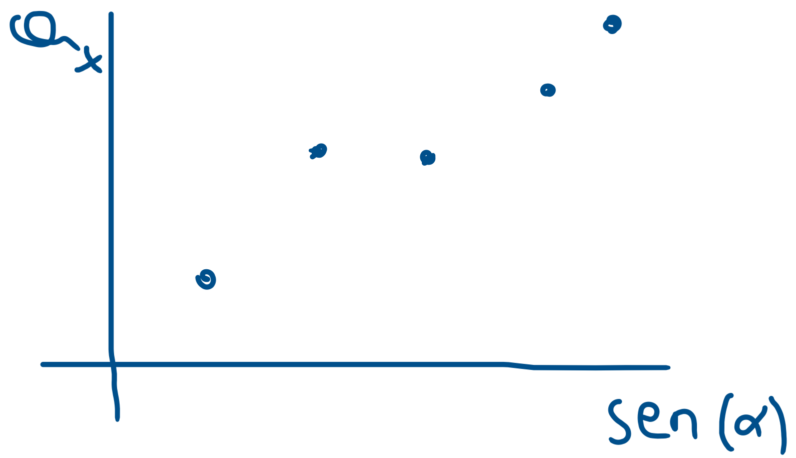
¿Cómo se va a ver una medición?



- A tener en cuenta:
  - Estimar errores
  - Hay que medir el ángulo de inclinación de la rampa, ¿ideas?
  - ¿Cómo obtenemos la aceleración?

Medición de g

- Vamos a usar que $a_x = g\,\mathrm{sen}(\alpha)$
- Podemos hacer el experiment para varias inclinaciones, calcular a_x para cada una y luego hacer un ajuste para encontrar g



Atentos con los errores!

Agenda

Para 5 inclinaciones

1) Medir ángulo (con su error!)

2) Hacer 5 tiradas del carrito desde el reposo y obtener su aceleración

3) Promediar estos valores para tener una medida para cada ángulo (con su error!)

4) Obtener g ajustando la aceleración contra sen del ángulo