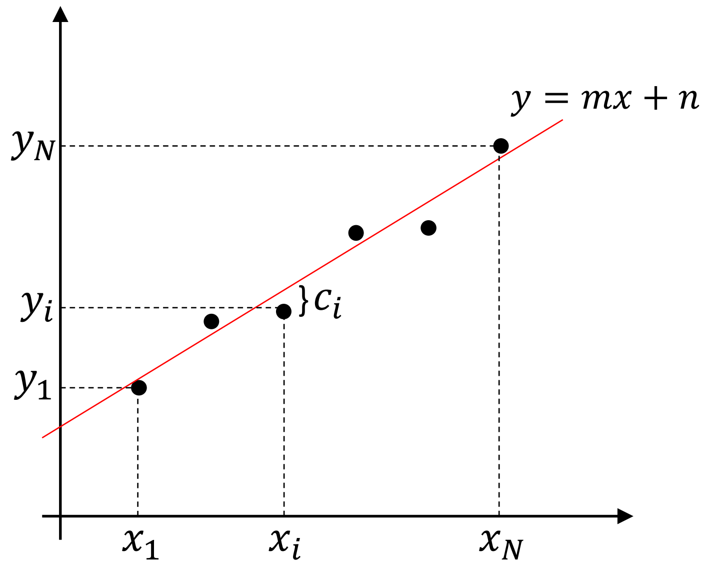
Ajuste lineal: método de mínimos cuadrados

Ajuste lineal por mínimos cuadrados

Dados N pares de medidas: $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$

y la recta: $y = mx + n$

buscamos encontrar los coficientes m y n que mejor aproximan la recta a la serie de datos.



Los coeficientes c_i (con i de 1 a N) representan la diferencia entre y_i y el valor que la recta toma en x_i. Es decir,

$$c_i = y_i - mx_i - n$$

Consideramos que la mejor aproximación está dada por los valores de m y n que minimizan la suma cuadrática de los c_i, esto es, mimimizar la función:

$$f(m, n) = \sum_{i=1}^{N} c_i^2 = \sum_{i=1}^{N} (y_i - mx_i - n)^2$$

Ajuste lineal por mínimos cuadrados

Para minimizar $f(m,n)$, igualamos sus derivadas a 0. Como es una función de dos variable, consideramos sus derivadas parciales. Esto es, derivamos f dos veces; primero considerando m como única variable y segundo, considerando n como única variable:

$$\frac{\partial f}{\partial m} = \frac{\partial \sum_{i=1}^{N}(y_i - mx_i - n)^2}{\partial m} = \sum_{i=1}^{N} 2(y_i - mx_i - n)(-x_i) = 0$$

$$\frac{\partial f}{\partial n} = \frac{\partial \sum_{i=1}^{N}(y_i - mx_i - n)^2}{\partial n} = \sum_{i=1}^{N} 2(y_i - mx_i - n)(-1) = 0$$

Operando cuidadosamente se obtiene:

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

$$\boxed{n = \bar{y} - m\bar{x}}$$

donde $\bar{x}$ y $\bar{y}$ son los promedios calculados con las series de datos.

Incertidumbre en los coeficientes

Para estimar las incertidumbres en los coeficientes m y n, asumimos lo siguiente:

- Los x_i no tienen incertidumbre;
- Los y_i tienen todos la misma incertidumbre que estimamos como la desviación estándar σ_c de los c_i.

Las incertidumbres en los parámetros las calculamos con la ecuación cuadrática de propagación de incertidumbres:

$$\delta m = \sqrt{\sum_{i=1}^{N}\left(\frac{\partial m}{\partial y_i}\delta y_i\right)^2} = \sigma_c\sqrt{\sum_{i=1}^{N}\left(\frac{\partial m}{\partial y_i}\right)^2} \longrightarrow \boxed{\delta m = \frac{\sigma_c}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}}$$

$$\delta n = \sqrt{\sum_{i=1}^{N}\left(\frac{\partial n}{\partial y_i}\delta y_i\right)^2} = \sigma_c\sqrt{\sum_{i=1}^{N}\left(\frac{\partial n}{\partial y_i}\right)^2} \longrightarrow \boxed{\delta n = \frac{\sigma_c}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}\sqrt{\frac{\sum_{i=1}^{N}{x_i}^2}{N}}}$$

Coeficiente de correlación lineal

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

¿Qué tan bueno es el ajuste?

- $R^2 = 1$ Correlación total.
- $R^2 = 0$ No hay correlación.