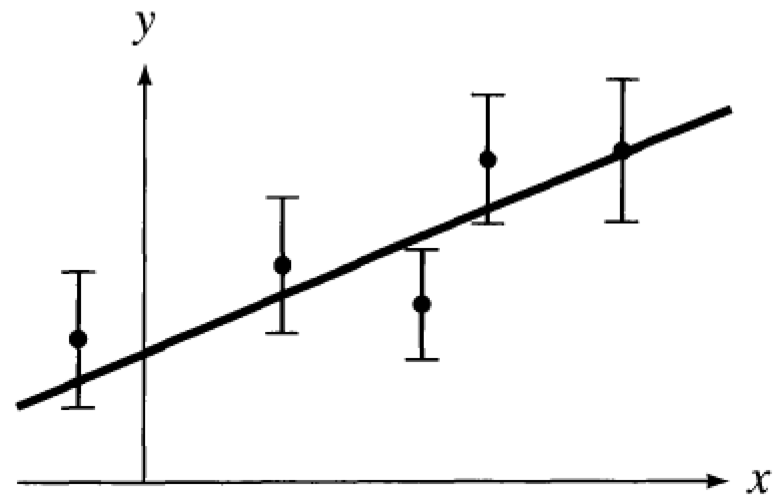
Regresión lineal

$$(x_1,\ y_1), \ldots, (x_N,\ y_N).$$

$$y = A + Bx$$



- Asumimos que los x_i no tienen incertidumbre
- Los parámetros A, B son los que minimizan la distancia sobre el eje *y* entre los puntos medidos y la recta

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)^2}{\sigma_{y_i}^2}.$$

Cuadrados mínimos

Si $$\sigma_{y_1} = \sigma_{y_2} = \ldots\ldots = \sigma_{yN} = \sigma_y$$

$$\frac{\partial \chi^2}{\partial A} = \frac{-2}{\sigma_y{}^2} \sum_{i=1}^{N} (y_i - A - Bx_i) = 0$$

$$\frac{\partial \chi^2}{\partial B} = \frac{-2}{\sigma_y{}^2} \sum_{i=1}^{N} x_i(y_i - A - Bx_i) = 0.$$

$$A = \frac{\sum x^2 \sum y - \sum x \sum xy}{\Delta}$$

$$B = \frac{N \sum xy - \sum x \sum y}{\Delta}$$

$$\Delta = N \sum x^2 - (\sum x)^2$$

$$\sigma_A{}^2 = \sum \left(\frac{\partial A}{\partial y_i} \sigma_y \right)^2$$

$$\sigma_B{}^2 = \sum \left(\frac{\partial B}{\partial y_i} \sigma_y \right)^2$$

$$\sigma_A = \sigma_y \sqrt{\frac{\sum x^2}{\Delta}}$$

$$\sigma_B = \sigma_y \sqrt{\frac{N}{\Delta}}$$

- Estimación del σ_y

Asumimos que todos lo y_i tienen la misma función distribución normal con el mismo σ_y

La dispersión de los datos alrededor de la recta estimada tendrá la misma desviación estándar

$$\sigma_y = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - A - Bx_i)^2},$$

Coeficiente de correlación

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}.$$

En el caso que $y_i = A + Bx_i$ $\bar{y} = A + B\bar{x}.$

$$r = \frac{B \sum (x_i - \bar{x})^2}{\sqrt{\sum (x_i - \bar{x})^2 B^2 \sum (x_i - \bar{x})^2}} = \frac{B}{|B|} = \pm 1.$$

Si no están correlacionados $\sum (x_i - \bar{x})(y_i - \bar{y}) = 0$

$$r = 0$$

Probabilidad $Prob_N(|r| \geq r_o)$ que N mediciones de dos variables x e y no correlacionadas puedan dar por resultado un coeficiente de correlación con $|r| \geq r_o$. Los valores corresponden a porcentaje de probabilidad y los casilleros en blanco corresponden a valores de probabilidad inferior a 0.05%

	r_o										
N	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
3	100	94	87	81	74	67	59	51	41	29	0
6	100	85	70	56	43	31	21	12	6	1	0
10	100	78	58	40	25	14	7	2	0.5		0
20	100	67	40	20	8	2	0.5	0.1			0
50	100	49	16	3	0.4						0

Ajuste de cuadrados mínimos pesado con la incertidumbre

Si $\sigma_{yi} = \sigma_i$ son distintos definimos $w_i = \dfrac{1}{{\sigma_i}^2}$

$$\chi^2 = \sum_i^N w_i (y_i - A - Bx_i)^2$$

$$A = \frac{\sum wx^2 \sum wy - \sum wx \sum wxy}{\Delta} \qquad B = \frac{\sum w \sum wxy - \sum wx \sum wy}{\Delta}$$

$$\Delta = \sum w \sum wx^2 - (\sum wx)^2$$

$$\sigma_A = \sqrt{\frac{\sum wx^2}{\Delta}} \qquad \sigma_B = \sqrt{\frac{\sum w}{\Delta}}$$

Justificación del principio de cuadrados mínimos

$$Prob_{A,B}(y_i) \;\propto\; \frac{1}{\sigma_y} e^{-(y_i - A - Bx_i)^2/2\sigma_y{}^2}.$$

$$\begin{aligned} Prob_{A,B}(y_1, \ldots, y_N) &= Prob_{A,B}(y_1) \cdots Prob_{A,B}(y_N) \\ &\propto \frac{1}{\sigma_y{}^N} e^{-\chi^2/2}, \end{aligned}$$

$$\chi^2 \;=\; \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)^2}{\sigma_y{}^2}.$$

El máximo de probabilidad corresponde al mínimo de χ^2