# Linear Regression in 2D

$$x = \{ x_i \} \qquad i = 1, \ldots, n$$

$$y = \{ y_i \} \qquad i = 1, \ldots, n$$

$$y_{fit} = b_0 + b_1 x$$

$$y_{fit} = \{ b_0 + b_1 x_i \}$$

$$\chi^2 = \sum_{i=1}^{n} \left( y_{fit} - y_i \right)^2$$

$$\chi^2 = \sum_{i=1}^{n} \left( b_0 + b_1 x_i - y_i \right)^2$$

$$\frac{\partial \chi^2}{\partial b} = 2 \sum_{i=1}^{n} \left( b_0 + b_1 x_i - y_i \right) = ($$

$\partial b_0$

$$\frac{\partial \chi^2}{\partial b_1} = 2\sum_{i=1}^{n}\left(b_0 + b_1 x_i - y_i\right)x_i$$

$$= 2\sum_{i=1}^{n}\left(b_0 x_i + b_1 x_i^2 - x_i \right.$$

$$= 0$$

$$b_0 \overset{n}{\boxed{\sum_{i=1}^{n} 1}} + b_1 \overset{\text{sum x}}{\boxed{\sum_{i=1}^{n} x_i}} - \overset{\text{sum y}}{\boxed{\sum_{i=1}^{n} y_i}} = 0$$

$$b_0 \underset{\text{sum x}}{\boxed{\sum_{i=1}^{n} x_i}} - b_1 \underset{\text{sum x2}}{\boxed{\sum_{i=1}^{n} x_i^2}} - \underset{\text{sum xy}}{\boxed{\sum_{i=1}^{n} x_i y_i}} =$$

$$\begin{pmatrix} n & \text{sumx} \\ \text{sumx} & \text{sumx2} \end{pmatrix}\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \text{sumy} \\ \text{sumxy} \end{pmatrix}$$

$$\begin{pmatrix} b_0 \\ \end{pmatrix} \quad \begin{pmatrix} n & \text{sumx} \\ \end{pmatrix}^{-1}\begin{pmatrix} \text{sumy} \\ \end{pmatrix}$$

$$\begin{pmatrix} b_1 \end{pmatrix} = \begin{pmatrix} sumx & sumx2 \end{pmatrix} \begin{pmatrix} sumxy \end{pmatrix}$$

$$\downarrow$$

$$= \frac{1}{\underbrace{n \cdot sumx2 - (sumx)^2}_{\Delta}} \begin{pmatrix} sumx2 & -sumx \\ -sumx & n \end{pmatrix} \times \begin{pmatrix} sumy \\ sumxy \end{pmatrix}$$

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} sumx2 \cdot sumy - sumx\ sum \\ n \cdot sumxy - sumx \cdot sum \end{pmatrix}$$

---

linear_regression(sumx, sumy, sumxy, sumx, n

→ returns $b\phi$, $b1$

---

2

What is "$\sigma$" is this instance?

If there are uncertainties associated with each data point, then:

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_{fit} - y_i)^2}{\sigma_i^2}$$

inverse weighting.

Suppose $\sigma_i = \sigma$ for all data pts.

$$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_{fit} - y_i)^2$$

Reduced $\chi^2 \equiv \chi^2 / \text{degree of freedom.}$

$$\chi^2_{reduced} = \frac{1}{\sigma^2 \nu_{error}} \sum_{i=1}^{n} (y_{fit} - y_i)^2$$

$\boxed{\text{Our best estimate of } \sigma \text{ is that value that makes } \chi^2_{\text{reduced}} = 1}$

$$\therefore \quad \sigma^2 = \frac{1}{\nu_{error}} \cdot \sum_{i=1}^{n} \left(y_{fit} - y_i\right)^2$$

$$\boxed{\sigma = \sqrt{\frac{1}{\nu_{error}} \cdot \sum_{i=1}^{n} \left(y_{fit} - y_i\right)^2}}$$

## Question 5

**Given** :

$$n, \ \overset{\text{sumx}}{\sum_i x_i}, \ \overset{\text{sumy}}{\sum_i y_i}, \ \overset{\text{sumx2}}{\sum_i x_i^2},$$

$$\underset{\text{sumxy}}{\sum_i x_i y_i}, \ \underset{\text{sumy2}}{\sum_i y_i^2}$$

u slope = linear_regression

$y_{int}$, slope

(sumx, sumy, sumxy, sumx2, n)

That part is easy enough!

$$\boxed{\text{What if they had asked us to Calculate a best estimate of } \sigma \text{ for this question?}}$$

Since we do not have the raw data, we cannot use the expression for $\sigma$ above 

But, we do have the summary statistics, and we also have the slope and y-intercept of the best-fit line ( $b_0$, $b_1$)

$$\sigma^2 = \frac{1}{\nu_{error}} \sum_{i=1}^{n} (y_{fit} - y_i)^2$$

$$= \frac{1}{\nu_{error}} \left( \sum_{i=1}^{n} \left( y_{fit}^2 - 2 y_{fit} \cdot y_i + y_i^2 \right) \right)$$

$$= \frac{1}{\nu_{error}} \left[ \sum_{i=1}^{n} \left( (b_0 + b_1 x_i)^2 - 2 (b_0 + b_1 x_i) \times y_i + y_i^2 \right) \right]$$

$$= \frac{1}{\nu_{error}} \left[ \sum_{i=1}^{n} \left[ b_0^2 + 2 b_0 b_1 x_i + b_1^2 x_i^2 - 2 b_0 y_i - 2 b_1 x_i y_i + y_i^2 \right] \right]$$

$$\sigma^2 = \frac{1}{\nu_{error}} \left[ b_0^2 n + 2 b_0 b_1 \sum_i x_i + b_1^2 \sum_i x_i^2 - 2 b_0 \sum_i y_i \right.$$

$$\left. - 2 b_1 \sum x_i y_i \right]$$

$$- 2 y_i \quad \ldots \quad + \sum_i y_i^2 \Big]$$

$$\boxed{\sigma = \sqrt{\frac{1}{\nu_{error}} \Big[ b_0^2 n + 2 b_0 b_1 \cdot \mathrm{sumx} + b_1^2\, \mathrm{sumx2} - 2 b_0\, \mathrm{sumy} - 2 b_1 \cdot \mathrm{sumxy} - \mathrm{sumy2} \Big]}}$$



?

$140 + a$

$a + b = 1448.6 - 1330$

$(140 + a)^2 + (140+b)^2$

$- 140^2 - 140^2 = 147414.45$

$- 140700$

$a + b = 118.6$

$(140+a)^2 + (140+b)^2 = 45914.45$

$-149.3121$

$a =$

$b = 294.9121$

$-18.3121$

$425.9121$

$$-1303$$

$$a + b = 1448.6$$

$$(131+a)^2 + (131+b)^2 = 147414.5$$

$$-133383$$

$$+131^2 + 131^2$$

---

$$\left[\frac{\;}{y_0+a}, \frac{\;}{y_1+b}, \frac{\;}{y_2+c}\right.$$

$$a+b+c = 10.57 - 10.53766036$$

$$(y_0+a)^2 + (y_1+b)^2 + (y_2+c)^2$$

$$- y_0^2 - y_1^2 + y_2^2 = 7.8653 -$$

$$6.8426369$$

$$x_0(y_0+a) + x_1(y_1+b) + x_2(y_2+c)$$

$$- x_0y_0 - x_1y_1 - x_2y_2 = 987.536$$

$$-984.24874$$

$\sigma = 0.25548$