M339G: September 10th, 2025.

Simple Linear Regression.

The Model.

In general: $f(X)$ ... f is the ideal fit, i.e.,

$Y = f(X) + \varepsilon$ w/ ε is independent of X and $\varepsilon \sim N(0, \sigma^2)$ for all predictors

$$\boxed{f(x) = \mathbb{E}[Y \mid X = x]}$$

Solve the optimization problem

$$\mathbb{E}\left[(Y - g(X))^2 \mid X = x\right] \xrightarrow{g} \min$$

$\hat{f}$ ... the fitted function/fit in a family of candidates g

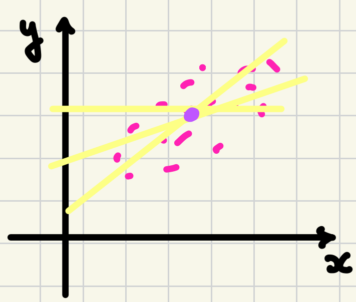
Def'n. In simple linear regression the model is

$$Y = \boxed{\beta_0 + \beta_1 \cdot X} + \varepsilon$$

w/ ε and X independent and $\varepsilon \sim N(0, \sigma^2)$

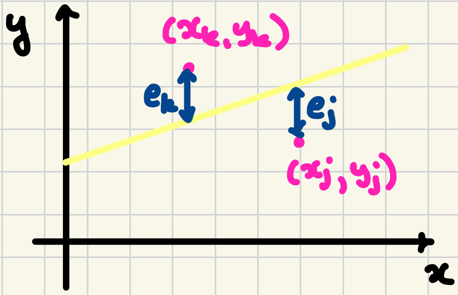
Method: Find ESTIMATORS $\hat{\beta}_0$ and $\hat{\beta}_1$ for β_0 and β_1

Set up:



data set:

$(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$



$\boxed{\text{Residuals}}$

Every line of fit would have the form

$$\hat{y} = \hat{b}_0 + \hat{b}_1 \cdot x$$

w/ $\hat{b}_0$ and $\hat{b}_1$ "candidate" coefficients

$\Rightarrow$ **Residuals**

$$SSE = RSS = \sum_{j=1}^{n} (y_j - \hat{y}_j)^2 \xrightarrow{\hat{b}_0, \hat{b}_1} \min$$

$$\sum_{j=1}^{n} (y_j - \hat{b}_0 - \hat{b}_1 \cdot x_j)^2 \xrightarrow{\hat{b}_0, \hat{b}_1} \min$$

Additionally: For Unbiasedness: $\boxed{\sum e_i = 0}$ $\leftarrow$

Differentiate w/ respect to $\hat{b}_0$ and $\hat{b}_1$

$$\frac{\partial RSS}{\partial \hat{b}_0} = \underbrace{-2\sum_{j=1}^{n} (y_j - \hat{b}_0 - \hat{b}_1 x_j)}_{=0} = 0$$

$$\left(\sum_{j=1}^{n} y_j = \sum_{j=1}^{n} (\hat{b}_0 + \hat{b}_1 x_j) = \right.$$

$$\left. = n\hat{b}_0 + \hat{b}_1 \sum_{j=1}^{n} x_j \right) // :n$$

$$\boxed{\bar{y} = \hat{b}_0 + \hat{b}_1 \cdot \bar{x}}$$ normal equation

$$\frac{\partial RSS}{\partial \hat{b}_1} = \underbrace{-2\sum_{j=1}^{n}\left((y_j - \hat{b}_0 - \hat{b}_1 \cdot x_j)\cdot x_j\right)}_{=0}$$

$$\sum_{j=1}^{n} x_j y_j - \hat{b}_0 \sum_{j=1}^{n} x_j - b_1 \sum_{j=1}^{n} x_j^2 = 0$$

$$\boxed{\hat{b}_1 = \frac{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2} = \frac{n\sum_{j=1}^{n} x_j y_j - \sum_{j=1}^{n} x_j \cdot \sum_{j=1}^{n} y_j}{n \cdot \sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2} = \frac{\sum_{j=1}^{n} x_j y_j - \frac{1}{n}\sum_{j=1}^{n} x_j \cdot \sum_{j=1}^{n} y_j}{\sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)^2}}$$

Q: $\boxed{\sum_{j=1}^{n} e_j x_j = \cancel{?}\ 0}$

Our estimators:

$$\hat{\beta}_1 = \text{"}\frac{Cov[X,Y]}{Var[X]}\text{"} = \frac{S_{XY}}{S_X^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \cdot \bar{X}$$