Linear Regression

Fitting models

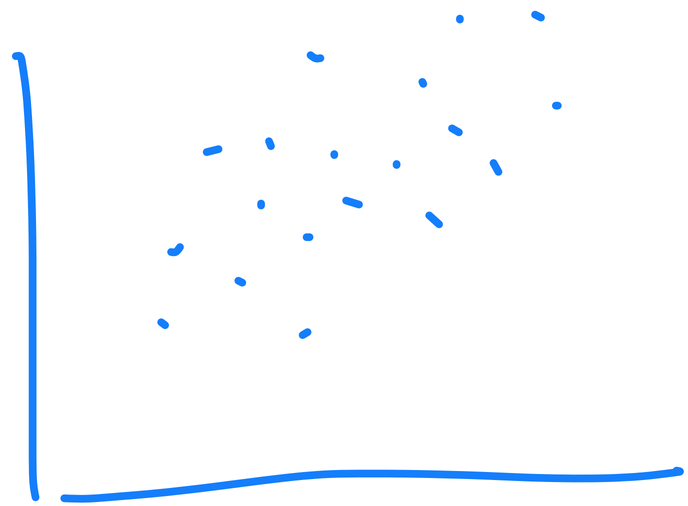
$y_i = \alpha + \beta x_i + \varepsilon_i$

dependent - y_i

ε_i - random noise

$\varepsilon_i \sim U(\ )$

$E[\varepsilon_i] = 0$



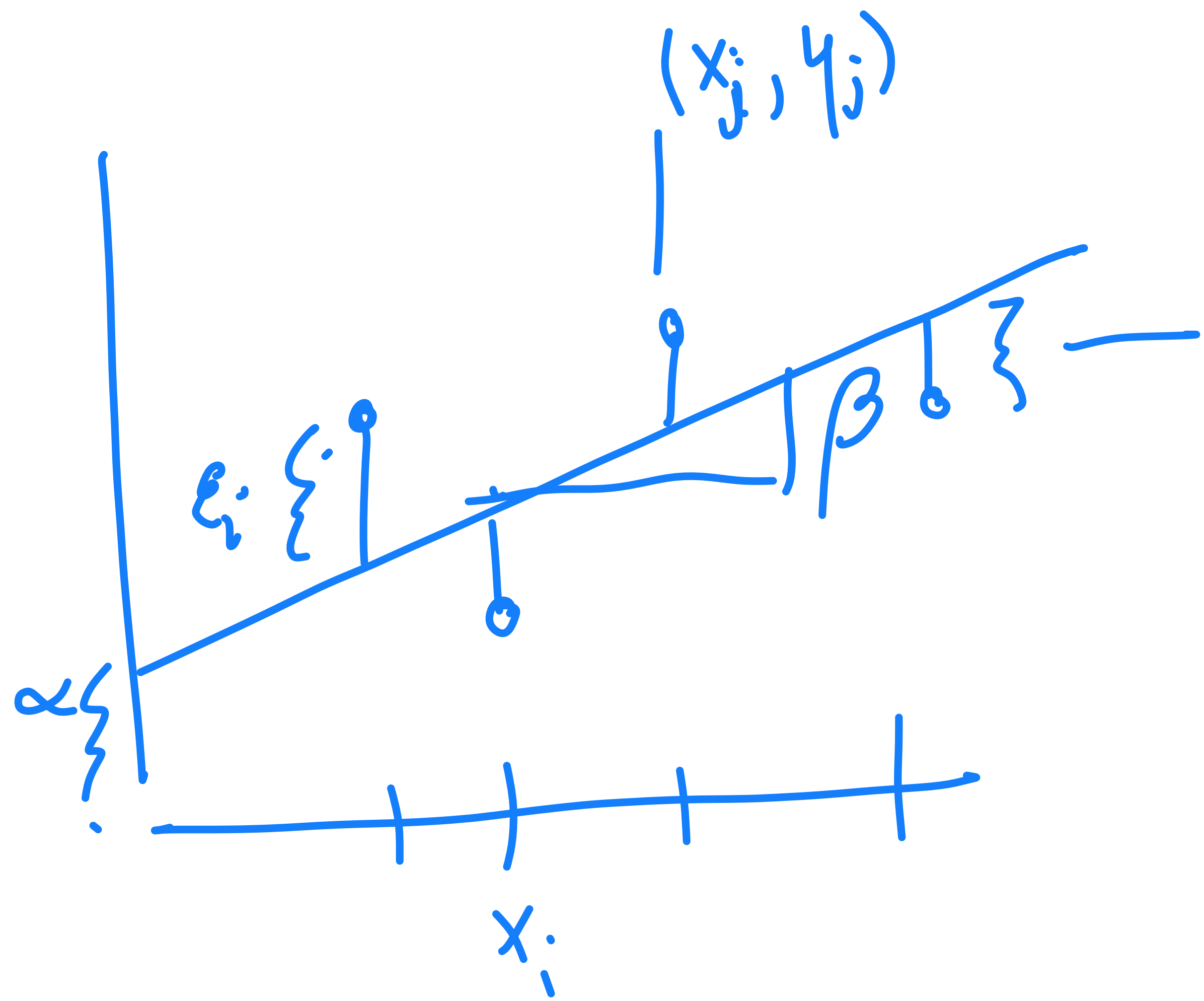
x_i - independent

Models.

Generative model.

Find parameters

α, β $\qquad$ ξ - parameters of dist



(x_j, y_j)
ϵ_i
β
α
x_i

Method of least squares

Objective function / penalty.

$$S(\alpha, \beta) = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

Central Limit Theorem

$\bar{X} = \sum X_i$

independent ?

$$Z_n = \frac{\bar{X}/n - \mu}{\sigma} \sim N(0,1)$$

$\lim_{n \to \infty}$

n is "large" n

Binomial

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

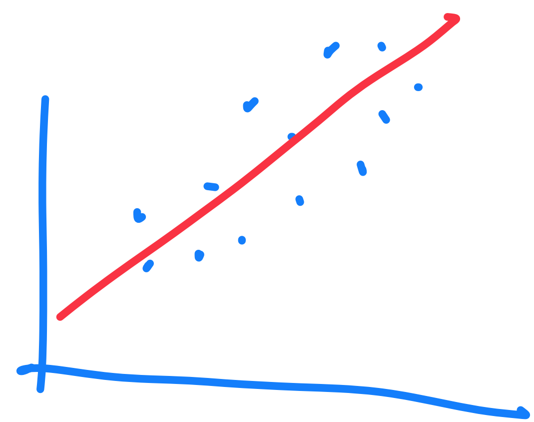
$$E[\varepsilon_i] = 0$$

zero mean noise

Q: what is α, β?

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

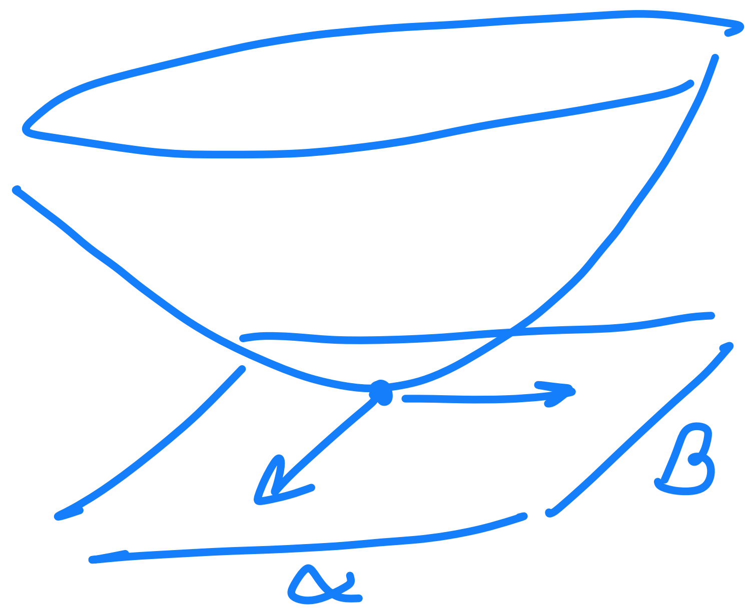
Estimator for $\alpha, \beta \rightarrow \hat{\alpha}, \hat{\beta}$

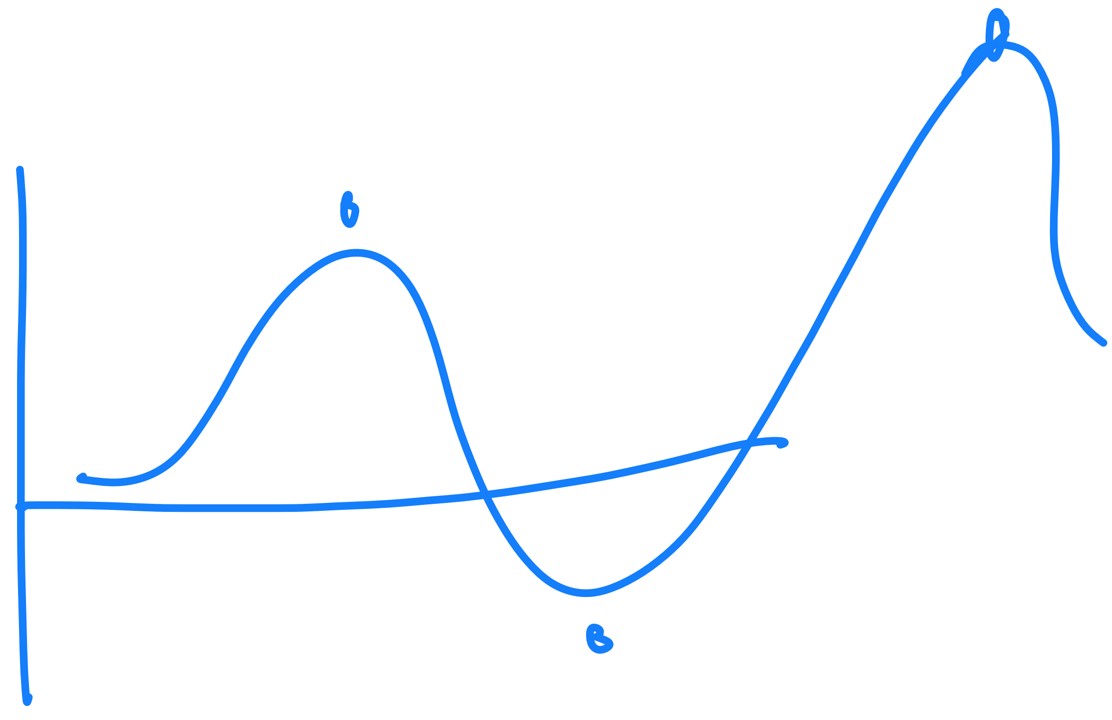


$$S(\alpha,\beta) = \sum_{i=1}^{n} \left(\alpha + \beta x_i - y_i\right)^2$$

$$\hat{\alpha}, \hat{\beta} = \underset{\alpha,\beta}{\operatorname{argmin}}\ S(\alpha,\beta).$$

$$\frac{\partial S}{\partial \alpha} = 0 \qquad \frac{\partial S}{\partial \beta} = 0$$





$$\frac{\partial S}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (\alpha + \beta x_i - y_i)^2$$

$$= 2\left[n\alpha + \beta \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \right] = 0$$

$$\frac{\partial S}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{n} (\alpha + \beta x_i - y_i)^2$$

$$= 2 \sum_{i=1}^{n} (y_i - \alpha - \beta x_i) x_i = 0$$

$$n\alpha + \beta \sum x_i = -\sum y_i$$

$$\alpha \sum x_i + \beta \sum x_i^2 = \sum x_i y_i$$

$\Big\}$ linear eqs.

Solve for α, β

$$\hat{\beta} = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\hat{\alpha} = \frac{1}{n} \sum y_i - \hat{\beta} \frac{1}{n} \sum x_i = \bar{y}_n \underline{- \beta \bar{x}_n}$$

$$E[\hat{\alpha}] = E[\bar{Y}_n] - \bar{X}_n E[\hat{\beta}]$$

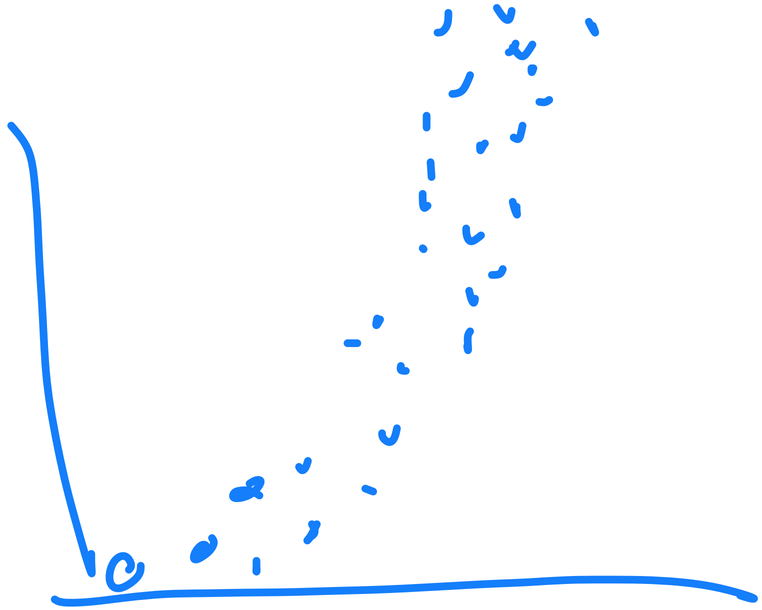
assume $\hat{\beta}$ is unbiased

$$= \frac{1}{n}\sum_{i=1}^{n} E[Y_i] - \bar{X}_n \beta$$

$$= \frac{1}{n}\sum E[\alpha + \beta X_i + \varepsilon_i] - \bar{X}_n \beta$$

$$= \frac{1}{n}\sum_{i=1}^{n} [\alpha + \cancel{\beta \bar{X}_n}] - \cancel{\bar{X}_n \beta} = \alpha$$

Next Steps.



$$y_i = \alpha + \beta x_i + \gamma x_i^2 + \varepsilon_i$$

$$S(\alpha, \beta, \gamma) =$$

$$\sum (\alpha + \beta x_i + \gamma x_i^2 - y_i)^2$$

$$\alpha^2, \beta^2, \gamma^2, \alpha, \beta, \gamma$$

$$\frac{\partial S}{\partial \alpha}, \frac{\partial S}{\partial \beta}, \frac{\partial S}{\partial \gamma} = 0, 0, 0$$

$$y_i = \alpha e^{\gamma x_i} + \varepsilon_i$$

$$+\varepsilon_i$$

$$\ln y_i = \ln \alpha + \gamma x_i$$

Significance?

$y_i = \alpha + \beta x_i + \varepsilon_i$

① $H_0 = \beta = 0$

$H_1 = y$ depends on x — $\beta \neq 0$, $\beta < 0$, $\beta > 0$

② Error rate $= 0.05$

③ Assume H_0

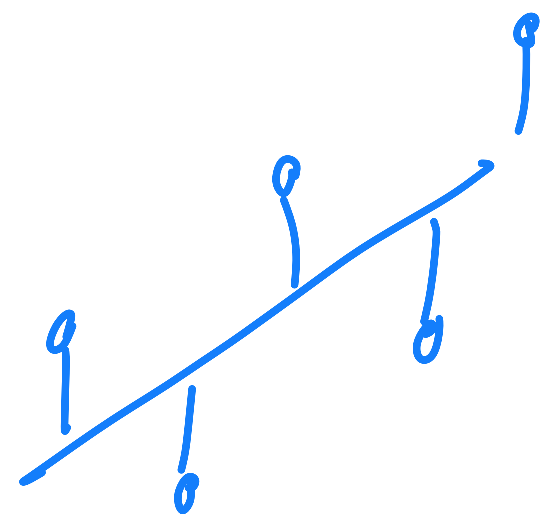
④ test statistic

$$t = \frac{\hat{\beta}}{s_b} = \frac{\hat{\beta}}{s_e / \sqrt{\sum_i (x_i - \bar{x})^2}}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} \qquad SSE = \sum_{i=1}^{n} \varepsilon_i^2$$

$$\varepsilon_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)$$

$$t \sim s\text{-}t\,(df = n-2).$$



$$\frac{\hat{\beta}}{s_b} = \frac{\hat{\beta}}{s_e \big/ \sqrt{\sum_i (X_i - \bar{X})^2}} \sim t$$