$X_1, \dots, X_n \sim N(\mu, \sigma)$

μ unknown

$(1-\alpha)100\%$ confidence interval μ

case 1: σ known

CI: $\bar{X} \pm Z_{\alpha/2} \cdot \sigma/\sqrt{n}$

$\bar{X} = 1$ $\quad \alpha = .05$

$\sigma = 6$

$n = 36$

$Z_{\alpha/2} = Z_{.025} = 1.96$

CI: $\left(1 - 1.96\frac{6}{6},\ 1 + 1.96\frac{6}{6}\right)$

$= (-0.96, 2.96)$

case 2: σ unknown

have to estimate σ from the data

$$S = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

S

CI: $\bar{X} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$

$t_{\alpha/2, n-1} \rightarrow$ scipy.stats.t.ppf(1 - α/2, n-1)

$X_1, \dots, X_{15}$ calculate $\hat{\mu}$, $\hat{\sigma}$, 90% CI for μ

$\hat{\mu}$ = np.mean(X)

$\hat{\sigma}$ = np.std(X)

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

s = np.std(X, ddof=1)

np.sum((X - np.mean(X))**2) / (n-1)

$H_0: \mu = 0$

$H_A: \mu \neq 0$

$\rightarrow$ case 1: $\sigma = .3$

case 2: σ unknown

rejection region in case 1:

$$|\bar{X}| > Z_{\alpha/2} \cdot \frac{3}{\sqrt{n}}$$

case 2:

$$|\bar{X}| > t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}$$

Final exam

April 20th 7:10-10:00 pm
(Tues)

$(X_1, Y_1), \ldots, (X_n, Y_n)$

X's = heights
Y's = weights

$$Y_i = \alpha + \underline{\underline{\beta}} X_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma)$$

$$Y_i \sim N(\alpha + \beta X_i, \underline{\underline{\sigma}})$$

α, β, σ

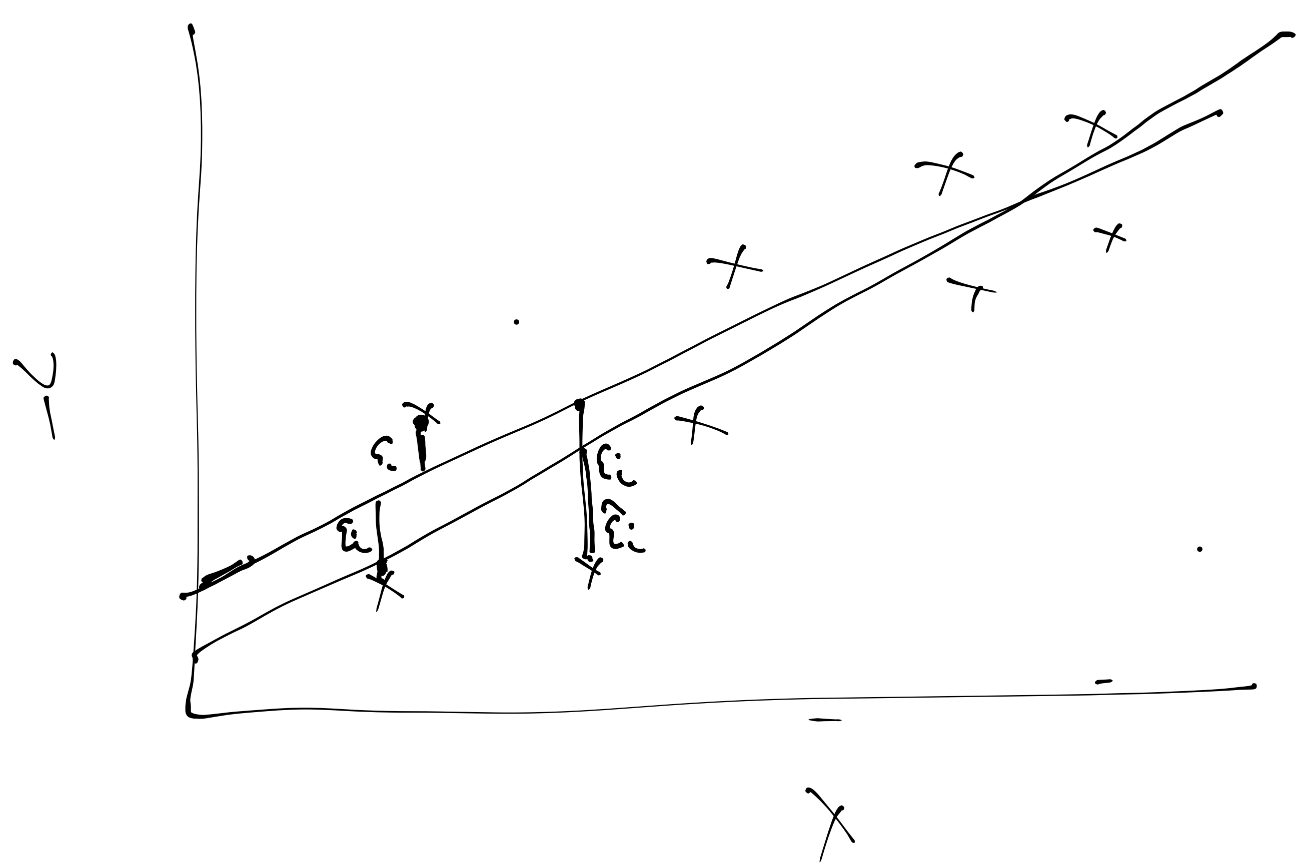
$\Longrightarrow$ fit it using maximum likelihood
and we find that maximizing the
likelihood for α, β

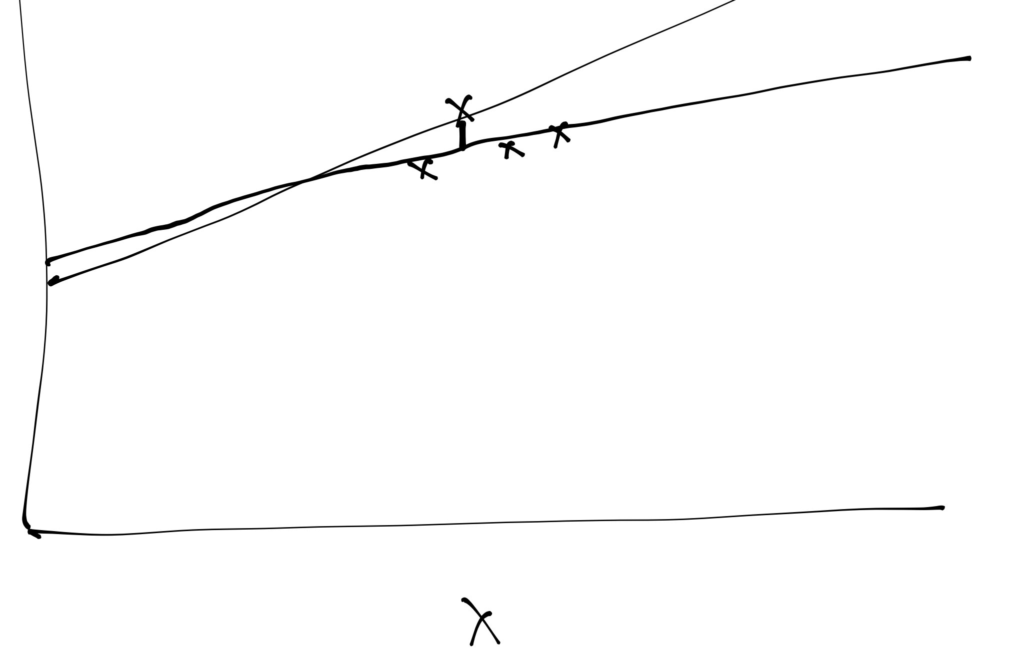
minimize $\sum \left(\underline{y_i} - \underline{\alpha - \beta x_i} \right)^2$

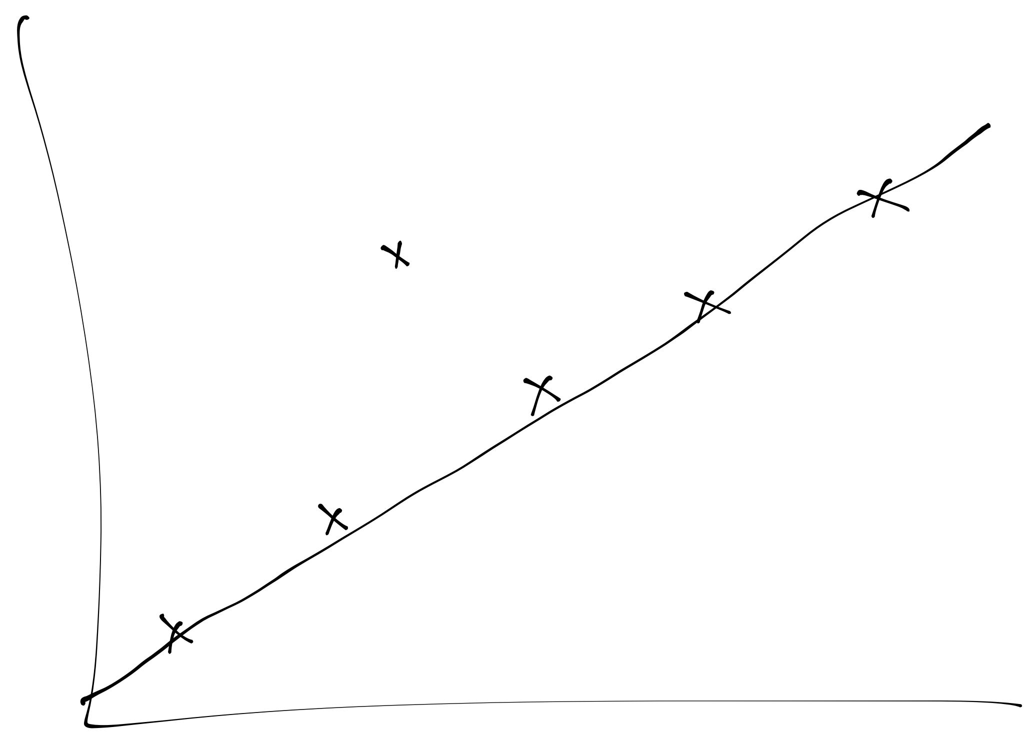
$$\hat{\beta} = \frac{\sum_{i=1} (x_i - \bar{x}) y_i}{\sum x_i^2 - n\bar{x}^2} \qquad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\sum x_i^2 - n\bar{x}^2 = SS_x$$

$$\hat{\beta} \sim N\left(\beta, \sqrt{\frac{\sigma^2}{\sum x_i^2 - n\bar{x}^2}}\right)$$







residuals : ith residual

$$\hat{\varepsilon}_i = (Y_i - \hat{\alpha} - \hat{\beta} x_i)$$

$$SS_R = \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

calculate confidence intervals for α, β

↳ bootstrap to calculate confidence intervals for α, β

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

↳ resample pairs and make m bootstrap samples

calculate $\hat{\alpha}, \hat{\beta}$ for each bootstrap sample

$(0,3), (1,2), (2,4)$

↳ $(0,3), (0,3), (2,4)$

$(2,4), (0,3), (2,4)$

never going to get a $(0,4)$ $(1,3)$

↳ fitted values $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$

expectation of $y \mid x_0$ → some point x, may not be in the data

$\hat{\alpha} + \hat{\beta} x_0$

→ confidence interval for $\hat{\alpha} + \hat{\beta} x_0$

→ prediction interval for $\hat{\alpha} + \hat{\beta} x_0$

↳ includes the variance from a residual term

Confidence interval for $\hat{\alpha} + \hat{\beta} x_0$

given m:

step 1: generating m bootstrap samples of pairs $(\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_n, \tilde{Y}_n)$

step 2: for each sample, calculate $\hat{\alpha}_j, \hat{\beta}_j$ (sample j)

$$\hat{y}_j = \hat{\alpha}_j + \hat{\beta}_j x_0$$

step 3: calculate the appropriate percentiles for $\hat{y}_j$

Prediction Intervals

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ say I know μ and σ

predict X_{n+1}

95% prediction interval: (L, U) s/t $X_{n+1} \in (L, U)$ 95%

$$\mu \pm Z_{\alpha/2} \cdot \sigma$$

suppose we dont know μ

↳ then we have to account for error in estimating μ

$$\bar{X} \pm Z_{\alpha/2} \cdot \sigma \sqrt{1 + \frac{1}{n}}$$

$\Rightarrow$ how do we use the bootstrap to generate prediction for regression?

prediction interval for $\hat{\alpha} + \hat{\beta} x_0$

given m:

step 1: generating m bootstrap samples of pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$

step 2: for each sample (sample j)

$\hat{\alpha}_j, \hat{\beta}_j$

(X_{ij}, Y_{ij}) = ith pair in bootstrap sample

ε_{ij} = ith error term in sample j

$\hat{y}_j = \hat{\alpha}_j + \hat{\beta}_j x_0 + \varepsilon_{ij}$

i is randomly from the error terms in jth bootstrap sample

step 3: get percentile $\hat{y}_j$