LECTURE 3

Simple Linear Regression

$(x_1, y_1) \ldots (x_n, y_n)$ $\quad y_i \in \mathbb{R}$, $x_i \in \mathbb{R}$

$$y_i = \underbrace{\boxed{\beta_0} + \boxed{\beta_1}}_{\text{parameters}} x_i + \underbrace{\varepsilon_i}_{\text{error}}$$

Statistical Inference: Frequentist / Bayesian

Frequentist Inference

parameters: θ in general (β_0, β_1) $\quad$ data: $(x_1, y_1) \ldots (x_n, y_n)$

① Construct some estimate of θ.

(e.g least squares estimators)

Maximum Likelihood Estimate.

$\varepsilon_1, \ldots \varepsilon_n \underset{\sim}{iid} N(0, \sigma^2)$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$y_i \overset{\text{independent}}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2) \leftarrow$$

$$\text{Likelihood} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right]$$

Optimization (Maximization)

$$\text{log-likelihood} = \sum_{i=1}^{n} \left[\log\left(\frac{1}{\sqrt{2\pi}}\right) - \log\sigma - \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right]$$

$$= n \log \frac{1}{\sqrt{2\pi}} - n \log \sigma - \frac{\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}$$

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\boxed{n \log \frac{1}{\sqrt{2\pi}} - n \log \sigma - \frac{S(\beta_0, \beta_1)}{2\sigma^2}}$$

Check: $\underbrace{\hat{\beta}_0, \hat{\beta}_1}_{\text{MLE}}$ coincide with the least squares estimators

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Check: $\hat{\sigma}_{MLE} = \sqrt{\frac{S(\hat{\beta}_0, \hat{\beta}_1)}{n}}$

② Characterize the distribution of the estimates.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$\sum (y_i - \bar{y})(x_i - \bar{x})$
$= \sum y_i (x_i - \bar{x})$
$- \sum \bar{y} \underbrace{(x_i - \bar{x})}$

$\boxed{y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)}$, Assume x_i s are fixed.

$$= \frac{\sum_{i=1}^{n} y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\sim N\left(\frac{\sum_{i=1}^{n} (\beta_0 + \beta_1 x_i)(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} , \frac{\sigma^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\left[\sum (x_i - \bar{x})^2\right]^2} \right)$$

$$= N\left(\beta_1 , \frac{\sigma^2}{\sum (x_i - \bar{x})^2}\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_1 , \frac{\sigma^2}{\sum (x_i - \bar{x})^2}\right)$$

$$\hat{\sigma}^2_{MLE} = \frac{S(\hat{\beta}_0, \hat{\beta}_1)}{n}$$

$$\boxed{n \frac{\hat{\sigma}^2_{MLE}}{\sigma^2} \sim \chi^2_{n-2}}$$

chi-squared distribution with $n-2$ degrees of freedom

Mean of $\chi^2_{n-2} = n-2$

$$\boxed{\mathbb{E}\, \hat{\sigma}^2_{MLE} = \sigma^2 \frac{n-2}{n}}$$

$\hat{\sigma}^2_{MLE}$: biased estimator

$$\boxed{\hat{\sigma}^2_{unbiased} = \frac{n}{n-2} \hat{\sigma}^2_{MLE} = \frac{1}{n-2} S(\hat{\beta}_0, \hat{\beta}_1)}$$

If we do Auto Regression:

$$\hat{\beta}_1 = \frac{\sum \boxed{(y_i - \bar{y})(y_{i-1} - \bar{y})}}{\boxed{\sum (y_{i-1} - \bar{y})^2}}$$

③ $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum (x_i - \bar{x})^2}\right)$

C.I for β_1: $\left[\hat{\beta}_1 - \frac{\sigma}{\sqrt{\sum (x_i - \bar{x})^2}} z_{\frac{\alpha}{2}}, \hat{\beta}_1 + \frac{\sigma}{\sum (x_i - \bar{x})^2} z_{\frac{\alpha}{2}}\right]$

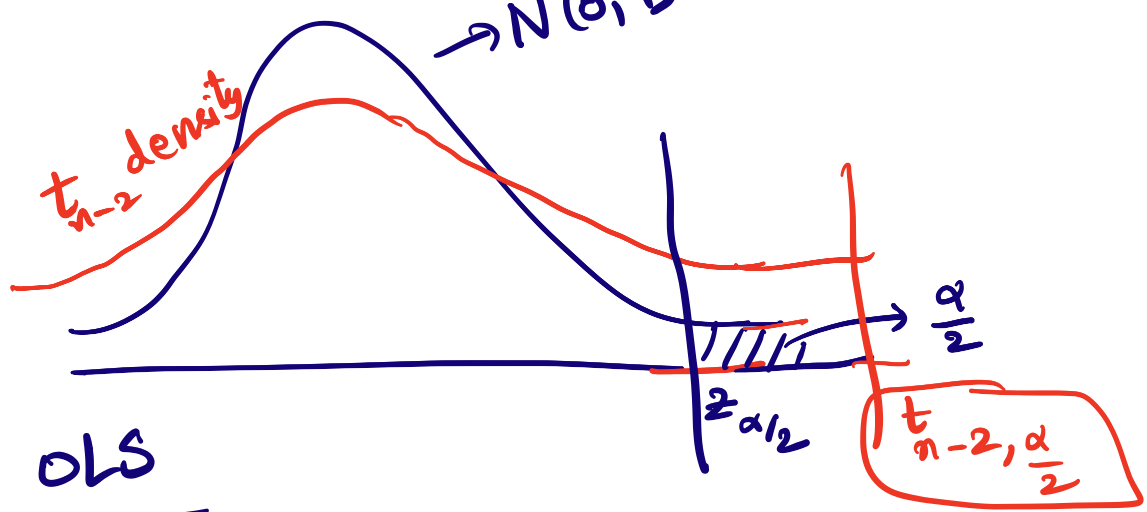
Because σ is unknown, you replace it by $\hat{\sigma}_{unbiased}$

$$\boxed{\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum (x_i - \bar{x})^2}\right)}$$

$$\boxed{\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{unbiased}} \sqrt{\sum (x_i - \bar{x})^2}} \sim t_{n-2}$$

Studentized $\hat{\beta}_1$

Student t-distn with n-2 degrees of freedom



sm. OLS

coef	s.e	t	[ ]
*	*	*	*

$\sigma \longrightarrow \hat{\sigma}_{\text{unbiased}}$

$z_{\frac{\alpha}{2}} \longrightarrow t_{n-2,\frac{\alpha}{2}}$

$\boxed{Y_1 \sim N(Y_2, 1)}$

$Y_1 - Y_2 \sim N(0,1)$

Bayesian Inference

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$y_i \overset{\text{independent}}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$

β_0, β_1, σ

① Select a probability distribution for the parameters (PRIOR)

— Captures your information about the parameters before you see the data

② Write likelihood & treat the likelihood as the conditional density of the data given the parameters.

③ Use the rules of probability (Bayes Rule) to compute the conditional distribution of parameters given observed data. → POSTERIOR DISTRIBUTION

④ Use the posterior distribution to infer the parameters.

Back to linear regression

$$\underbrace{\beta_0, \beta_1, \sigma}_{} \sim$$

$\beta_0, \beta_1, \log\sigma \overset{iid}{\sim} \text{Unif}[-C, C]$ for a large C

$$\log\sigma \sim \text{Unif}[-C, C]$$

$$\Rightarrow f_{\log\sigma}(x) = \frac{I\{-C < x < C\}}{2C}$$

$$\Rightarrow f_{\sigma}(u) = f_{\log\sigma}(\log u) \times \frac{1}{u}$$

$$= \frac{I\{-C < \log u < C\}}{2Cu}$$

$$= \frac{I\{e^{-C} < u < e^{C}\}}{2Cu}$$

$$f_\sigma(\sigma) \propto \frac{1}{\sigma}$$

$$f_{2\sigma}(t) = f_\sigma\left(\frac{t}{2}\right) \times \frac{1}{2}$$

$$= \frac{2}{t} \times \frac{1}{2} = \frac{1}{t}$$

$$\boxed{\beta_0, \beta_1, \log\sigma \ \underset{\sim}{\text{iid}} \ \text{Unif}(-C, C)}$$

$$f_{\beta_0,\beta_1,\sigma}(\beta_0,\beta_1,\sigma) = \frac{I\{-C < \beta_0, \beta_1, \log\sigma < C\}}{2C \times 2C \times 2C \times \sigma}$$

$$\propto \frac{1}{\sigma} I\{-C < \beta_0, \beta_1, \log\sigma < C\}$$

Likelihood:

$$f_{\text{data}|\beta_0,\beta_1,\sigma}(\text{data}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)$$

$$\propto \prod_{i=1}^{n} \left[\frac{1}{\sigma} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)\right]$$

$$= \sigma^{-n} \exp\left[-\frac{S(\beta_0, \beta_1)}{2\sigma^2}\right]$$

$$\underbrace{f_{\beta_0,\beta_1,\sigma}}_{\text{PRIOR}}, \quad \underbrace{f_{\text{data}|\beta_0,\beta_1,\sigma}}_{\text{Likelihood}} \longrightarrow f_{\beta_0,\beta_1,\sigma|\text{data}}$$

<u>BAYES RULE</u>: How to go from $\mathbb{P}(A|B)$ to $\mathbb{P}(B|A)$?

e.g A : someone died

B : they were hanged

$$\mathbb{P}(A|B) = 1$$

$$\mathbb{P}(B|A) \approx 0$$

$\mathbb{P}(B|A)$

$\mathbb{P}(B^c|A)$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A)}$$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c)}$$

$$f_{\underset{B}{\text{parameters}}|\underset{A}{\text{data}}}(\text{parameters}) = \frac{f_{\text{data}|\text{parameters}}(\text{data})\, f_{\text{parameters}}(\text{parameters})}{f_{\text{data}}(\text{data})}$$

f_{data} & $f_{\text{data}|\text{parameters}}$

$$= \frac{f_{\text{data}|\text{parameters}}(\text{data})\, f_{\text{parameters}}(\text{parameters})}{\int f_{\text{data}|\text{parameters}}(\text{data})\, f_{\text{parameters}}(\text{parameters})\, d(\text{parameters})}$$

Simplified Bayes Rule:

$$\underbrace{f_{\text{parameters}|\text{data}}(\text{parameters})}_{\text{Posterior}} \propto \underbrace{f_{\text{data}|\text{parameters}}(\text{data})}_{\text{Likelihood}}\, \underbrace{f_{\text{parameters}}(\text{parameters})}_{\text{Prior}}$$

$$f_{\beta_0, \beta_1, \sigma \mid data}(\beta_0, \beta_1, \sigma) \propto \frac{1}{\sigma^n} \exp\left(-\frac{S(\beta_0, \beta_1)}{2\sigma^2}\right) \frac{1}{\sigma} I\{-C < \beta_0, \beta_1, \log\sigma < C\}$$

$$\text{posterior} \propto \sigma^{-n-1} \exp\left(-\frac{S(\beta_0, \beta_1)}{2\sigma^2}\right) I\{-C < \beta_0, \beta_1, \log\sigma < C\}$$

$$\boxed{\beta_0, \beta_1}$$

X_1, X_2, X_3 $\quad f_{X_1, X_2, X_3}(x_1, x_2, x_3) \rightarrow$ INTEGRATION

Integration: $f_{\beta_0, \beta_1 \mid data}$