Simple Linear Regression

Wednesday, September 1, 2021 6:23 PM

- Overview
  - Model

$$Y \approx \beta_0 + \beta_1 X$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- Estimation
  - Training data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
  - $RSS = e_1^2 + e_2^2 + \ldots e_n^2 \qquad e_i = y_i - \hat{y}_i$

Residual sum of squares

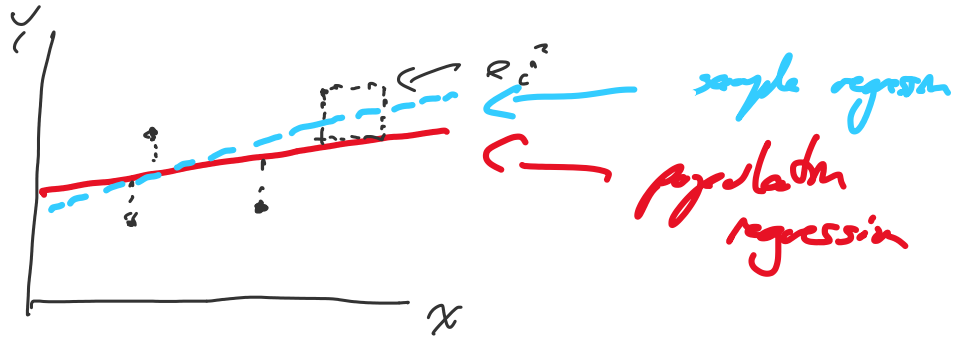
$$\min_{\hat{\beta}_0, \hat{\beta}_1} RSS$$

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

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

- Analysis
  - True relationship:

$$Y = \underbrace{\beta_0 + \beta_1 X}_{f(X)} + \varepsilon$$



* Bias: Sample statistic vs Population value

e.g. mean $\Rightarrow \hat{\mu}$ vs. μ

as $n \to \infty$ in my sample, does $\hat{\mu} \to \mu$?

- Variance of my estimate:

e.g. $Var(\hat{\mu}) = SE(\hat{\mu})^2 = \frac{\sigma^2}{n}$ ← population variance

$$SE(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]$$

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

$\sigma^2 = Var(\varepsilon)$

* Confidence Intervals:
  - X% confidence interval → "true" value is within a defined range with X percent probability

$\hat{\beta}_1 \pm 2\,SE(\hat{\beta}_1)$ the 95% prob that

actual β_1 is within $[\hat{\beta}_1 - 2SE(\hat{\beta}_1), \hat{\beta}_1 + 2SE(\hat{\beta}_1)]$

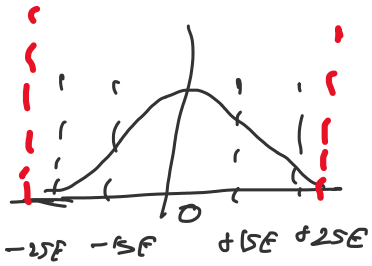
* Assumption that ε is Gaussian (Normal)
  ↳ with Low n, Student-t

↳ Hypothesis Tests
- H_0: Null Hypothesis, No Relationship
- H_a: Alternative, Is Relationship

$H_0: \beta_1 = 0$
$H_a: \beta_1 \neq 0$

* T-Stat:

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$$



t-dist with $n-2$ df → probability of observing $\hat{\beta}_1 \neq 0$

p-value → < 0.05 $> 95\%$

↳ Accuracy
- RSE: Estimate of $sd(\varepsilon)$

$$RSE = \sqrt{\frac{1}{n-2} RSS} = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

adjust df

* measure of "lack of fit"

- R^2 ⇒ Proportion of variance explained

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

$TSS = \sum (y - \bar{y})^2$ ← total sum of squares

* $R^2 = r^2 \rightarrow r = cor(X, Y)$

⇒ single var case!