<u>Recall</u>: . $(\underset{\sim}{Y}, X\underline{\beta}, \sigma^2 I_{n\times n})$ - linear model

. $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \varepsilon_i \qquad 1 \le i \le n$

ε_i = uncorrelated r.v. with mean 0 and variance $\sigma^2 I_{n\times n}$.

<u>Least square Estimate</u>:

A (random) vector $\hat{\underline{\beta}} \in \mathbb{R}^m$ is called a least square estimate of β if

$$(*) \quad \|\underline{y} - X\hat{\underline{\beta}}\|_2 = \min_{\beta \in \mathbb{R}^m} \|\underline{Y} - X\underline{\beta}\|_2$$

<u>Solved for $\hat{\beta}$</u>

least square estimate $\hat{\beta}$ of β satisfies

$$(X^T X)\hat{\underline{\beta}} = X^T \underline{y} \quad \cdots \quad (1)$$

(normal equations)

Example:

Given the data

x	y
2	1
3/2	2
4	1

linear model

$$y = \beta_0 + \beta_1 x + \underset{\uparrow \text{ ERROR}}{\varepsilon}$$

$$X = \begin{pmatrix} 1 & 2 \\ 1 & 3/2 \\ 1 & 4 \end{pmatrix} \qquad \underset{\sim}{y} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \qquad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Find the least square estimate $\hat{\beta}$

Solve: $\quad X^T X \beta = X^T y$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3/2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 3/2 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 7.5 \\ 7.5 & \frac{89}{4} \end{pmatrix}$$

$$X^T \underset{\sim}{y} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3/2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 9 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 7.5 \\ 7.5 & \frac{89}{4} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 4 \\ 9 \end{pmatrix}$$

Ex: (linear algebra)

$$\hat{\beta}_0 = \frac{43}{21} \qquad \hat{\beta}_1 = -\frac{2}{7}$$

Understanding the estimates $\hat{\beta}_0$ & $\hat{\beta}_1$

Simple linear regression:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad 1 \le i \le n$$

ε_i = uncorrelated mean 0 & variance σ^2

Least square estimates: $\hat\beta_0$, $\hat\beta_1$

they solve $\quad X^T X \beta = X^T \underline{y} \quad$ — ②

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad \underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & \cdots & & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

$$X^T \underline{y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$ (Ex.)

Solution to ②

$$\left.\begin{aligned} \hat\beta_0 &= \bar y - \hat\beta_1 \bar x \\ \hat\beta_1 &= \frac{\sum_{i=1}^n (x_i - \bar x)(y_i - \bar y)}{\sum_{i=1}^n (x_i - \bar x)^2} \end{aligned}\right\}$$ explicit Solution.

$$\bar x = \frac{1}{n}\sum_{i=1}^n x_i \qquad \bar y = \frac{1}{n}\sum_{i=1}^n y_i$$

RSS ≡ Residual sum of squares

$$(y_1 - \hat\beta_0 - \hat\beta_1 x_1)^2 + (y_2 - \hat\beta_0 - \hat\beta_1 x_2)^2 + \cdots + (y_n - \beta_0 - \beta_1 x_n)^2$$

Q: How to assess the accuracy of this?

Recall: (Day 1) In R - we drew 100 lines and then tried to find the "best line".

Lecture 2:

Ans: Y - some characteristic of the population

μ - mean of Y. , variance of $Y = \sigma^2$

Given:- $y_1, y_2, \dots, y_n$ $n \geq 1$ n - sample points

Estimate μ from data:

$$\hat{\mu} := \frac{1}{n} \sum_{i=1}^{n} y_i \equiv \text{estimate of } \mu$$

Q: Is $\hat{\mu}$ a good estimator of μ?

A: • $\hat{\mu}$ is an unbiased estimate of μ $(E[\hat{\mu}] = \mu)$

can this be carried over to $\hat{\beta}_0, \hat{\beta}_1$? ✓

Ex: True

• $\mathrm{Var}(\hat{\mu}) = \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} y_i\right)$

$$= \frac{\sigma^2}{n}$$

(n large $\Rightarrow$ reduction in variance of $\hat{\mu}$)

can this be carried over to $\hat{\beta}_0, \hat{\beta}_1$? ✓

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

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

• $\equiv$ orange are deterministic

• $E\, y_i = \beta_0 + \beta_1 x_i$

$\mathrm{Var}(y_i) = \sigma^2$

Ex:-

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

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

Observation: - more x_i are spread out

$\Rightarrow$ smaller is the variance $(\hat{\beta}_1)$

Interval estimate for β_0, β_1:-

$$\left(\hat{\beta}_1 - 2\sqrt{\text{var}(\hat{\beta}_1)},\ \hat{\beta}_1 + 2\sqrt{\text{var}(\hat{\beta}_1)} \right)$$

is a 95% confidence interval for β_1

$$\left(\hat{\beta}_0 - 2\sqrt{\text{var}(\hat{\beta}_0)},\ \hat{\beta}_0 + 2\sqrt{\text{var}(\hat{\beta}_0)} \right)$$

is a 95% confidence interval for β_0

Recall: a similar procedure yields a confidence interval for μ.

Hypothesis testing: $H_0: \mu = 0$ $\quad H_A = \mu \neq 0$

(Device test statistic to see if we have evidence to reject the null hypothesis)

can this be carried over to $\hat{\beta}_0, \hat{\beta}_1$? ✓

$H_0: \beta_1 = 0$ $\quad\quad$ $H_A = \beta_1 \neq 0$

"$y_i = \beta_0 + \varepsilon_i$" (X is not associated with y)

- Standard error $(\hat{\beta}_1)$ <<< small
  - even small values of $\hat{\beta}_1 \underset{\text{evidence}}{\rightsquigarrow} \beta_1 \neq 0$

- standard error $(\hat{\beta}_1)$ >>> large
  - only large values of $\hat{\beta}_1 \underset{\text{evidence}}{\rightsquigarrow} \beta_1 \neq 0$.

In practice one does a t- test

$$t = \frac{\hat{\beta}_1 - 0}{\sqrt{var(\hat{\beta}_1)}} \quad — (*)$$

Under $H_0: \beta_1 = 0$ (there is no relationship between x and y)

(*) has t_{n-1} - distribution

lecture 3: To test $\beta_1 = 0$

$\mathbb{P}(|t_{n-1}| > |t|)$ .. p value

a small p-value indicates that it is unlikely to observe a substantial between y and x

Assess the model :-

The quality of a linear regression is assessed by two quantities

- residual standard error (RSE)
- R^2 statistic.

RSE : – provides an absolute measure of "lack of fit"

RSS ≡ Residual sum of squares

$$(y_1 - \hat\beta_0 - \hat\beta_1 x_1)^2 + (y_2 - \hat\beta_0 - \hat\beta_1 x_2)^2 + \cdots + (y_n - \beta_0 - \beta_1 x_n)^2$$

$$RSE := \sqrt{\frac{1}{n-2} RSS}$$ ("estimate of σ")

$$= \sqrt{\frac{1}{n-2}\left(\sum_{i=1}^{n} (y_i - \hat\beta_0 - \hat\beta_1 x_i)^2\right)}$$ – measured in units of y

"$y_i \simeq \hat\beta_0 - \hat\beta_1 x_i$" ≡ RSE small

✓ ≡ "model fits the data

< "$y_i > \hat\beta_0 - \hat\beta_1 x_i$" ≡ RSE large

✗ ≡ "model fits the data

R^2 statistic :

$$TSS = \sum_{i=1}^{n} (y_i - \bar y)^2$$

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

- TSS $\equiv$ measures the variance of $\underline{y}$

  RSS $\equiv$ measures the variability that is not explained post fit from the model

  TSS − RSS $\equiv$ measures the amount of variability in the data $\underline{y}$ that is explained by performing least squares.

  R^2 = measures the proportion of variability in $\underline{y}$ that can be explained by using $\underline{x}$

$R^2 \simeq 1$ (close to 1) large proportion of variability is explained by the model

$R^2 \simeq 0$ (close to 0) linear model is perhaps wrong