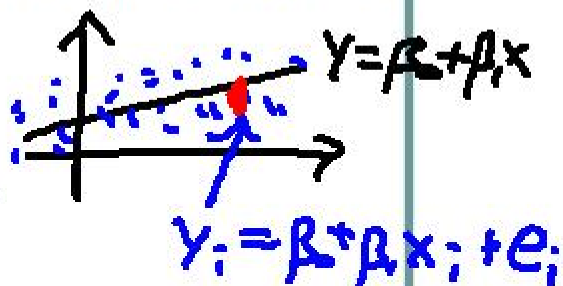
Simple Linear Regression

- Instead of expressing E(Y|X) and Var(Y|X) separately, we may say
  - $y_i = \beta_0 + \beta_1 x_i + e_i$ ← ith observation $i = 1, 2 \cdots n$
  - $E(e_i) = 0$, $Var(e_i) = \sigma^2$, e_i's are i.i.d.



variance function

- Compare with
  - Mean function: $E(Y|X=x) = \beta_0 + \beta_1 x$
  - Variance function: $Var(Y|X=x) = \sigma^2$
- e_i : statistical error
  - Vertical distance between y and the "truth" $E(Y|X=x_i)$
  - $e_i = y_i - \beta_0 - \beta_1 x_i$ = Random error component
- We can assume x is known and e, y are random

Least Squares estimator

- Hat is important!!!

Known from data

unknown True: $y_i = \beta_0 + \beta_1 x_i + e_i$

$$\hat{y}_i = \hat{\mathrm{E}}(Y|X = x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

y_i, x_i $\beta_0, \beta_1, \sigma^2$

$\hat{\beta}_1, \hat{\beta}_0, \hat{\sigma}^2$

- Be familiar with the terms: observed data, parameter, parameter estimates, error, residual, fitted value…

e_i $\hat{e}_i$ $\hat{y}_i$

- With hat = Known estimated value
  = Involve data only
- No hat = Observed/Unknown true value

Least Squares estimators

- Minimize Residual sum of squares

$$\boxed{RSS(\beta_0, \beta_1) = \sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i)]^2}$$

$$\frac{\partial RSS}{\partial \beta_0} = 0$$
$$\frac{\partial RSS}{\partial \beta_1} = 0$$

- Step 1. Differentiate RSS w.r.t. β_0 & β_1
- Step 2. Solving 2 equations 2 unknowns
- Answer:

$\frac{1}{n}\Sigma x_i$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

(see textbook P.273)

Least Squares estimator

- Notations used in the book

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

Sample mean

$$\bar{x} = \frac{1}{n}\sum_i x_i, \qquad \bar{y} = \frac{1}{n}\sum_i y_i,$$

Sum of squares

$$SXX = \sum_i (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2,$$

$$SYY = \sum_i (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2,$$

$$SXY = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$$

Estimators

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}, \quad \hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{SXY}{SXX}$$

Fitted value for case i

$$\hat{y}_i = \hat{\mathrm{E}}(Y|X = x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

residual for case i

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

e_i
↑ error

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

$Var(X) = E(X^2) - (EX)^2$

How to estimate σ^2 ?

- There are 3 parameters
  - $y_i = \beta_0 + \beta_1 x_i + e_i$
  - $E(e_i) = 0$, $Var(e_i) = \sigma^2$, e_i's are i.i.d.
- Note that $\sigma^2 = Var(e_i) = E(e_i^2)$,
  - Looking at $\sum \hat{e}_i^2$ helps estimate σ^2
  - Fact: $E(\sum \hat{e}_i^2) = (n-2)\sigma^2$
  - We estimate σ^2 by

$$\boxed{\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-2}}$$

$Y_i = \beta_0 + \beta_1 X_i + e_i$

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

$\hat{e}_i = Y_i - \hat{Y}_i \approx e_i$

Sample variance : $\dfrac{\sum(X_i - \bar{X})^2}{n-1}$ $\quad E(\sum(X_i - \bar{X})^2) = (n-1)\sigma^2$

Computations

- $$\underline{\sum \hat{e}_i^2} \stackrel{\text{def}}{=} \sum_{i=1}^{n} [\underbrace{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}_{\hat{e}_i}]^2 \stackrel{\text{def}}{=} RSS(\hat{\beta}_0, \hat{\beta}_1) \stackrel{\text{def}}{=} RSS$$

- $$RSS = \sum_{i=1}^{n} [\underbrace{y_i - \bar{y}}_{a} + \underbrace{\bar{y} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}_{b}]^2$$

1

$(a+b)^2 = a^2 + 2ab + b^2$

$-2\hat{\beta}_1 SXY = -2\hat{\beta}_1^2 SXX$

$\hat{\beta}_1^2 SXX$

$$= \sum_{i=1}^{n} [y_i - \bar{y}]^2 + 2\sum_{i=1}^{n} [\bar{y} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)][y_i - \bar{y}] + \sum_{i=1}^{n} [\bar{y} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

SYY

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

2 $\boxed{\hat{\beta}_1 [\bar{x} - x_i]}$

$$= SYY - \hat{\beta}_1^2 SXX = SYY - \frac{SXY^2}{SXX}$$

- Therefore

difficult easy

$$\boxed{\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-2} = \frac{SYY - \frac{SXY^2}{SXX}}{n-2}}$$

$SYY = \sum (Y_i - \bar{y})^2$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Substitute $\bar{Y}$

$$\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

$$= \hat{\beta}_0 + \hat{\beta}_1 \bar{X} - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

$$= \hat{\beta}_1 (\bar{X} - X_i)$$

$$RSS = \Sigma(Y_i - \beta_0 - \beta_1 X_i)^2$$

$$\hookrightarrow \min_{\beta_0, \beta_1} RSS \rightarrow \hat{\beta}_1 = \frac{SXY}{SXX} = \frac{\Sigma(X_i - \bar{x})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{x})^2}$$

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

$$\sigma^2 = Var(e_i) = E(e_i^2)$$

$$E(\Sigma \hat{e}_i^2) = (n-2)\sigma^2 \Rightarrow \frac{\Sigma \hat{e}_i^2}{n-2} = \hat{\sigma}^2 \quad (E(\hat{\sigma}^2) = \sigma^2)$$

Quantify how good is the estimator

Want to show:

$$E(\hat{\beta}_0) = \beta_0$$

$$E(\hat{\beta}_1) = \beta_1$$

$$Var(\hat{\beta}_0) \longrightarrow 0 \quad (n \to \infty)$$

$$Var(\hat{\beta}_1) \longrightarrow 0 \quad (n \to \infty)$$

Properties of least squares estimators

X-non random
Y-random

1. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ can be written as a linear combination of y_i
   - $\widehat{\beta}_1 = \sum \underbrace{(\frac{x_i - \bar{x}}{SXX})}_{x} \underbrace{y_i}_{y}$ $\hat{\beta}_0 = \sum \underbrace{[\frac{1}{n} - \bar{x}\left(\frac{x_i - \bar{x}}{SXX}\right)]}_{x} \underbrace{y_i}_{y}$
2. The fitted line passes through $(\bar{x},\bar{y})$
   - Check the derivative of RSS w.r.t. β_0
3. The sum of residuals=0 (not errors!)
   - Check the derivative of RSS w.r.t. β_0
4. The estimators are unbiased
   (Expectation=true value)
   - $\mathrm{E}(\widehat{\beta}_0|X) = \beta_0, \quad \mathrm{E}(\widehat{\beta}_1|X) = \beta_1, \quad \mathrm{E}(\widehat{\sigma}^2|X) = \sigma^2$

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

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

$\sum e_i \neq 0$

$\sum \hat{e}_i = 0$
$= \sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$
$= n\bar{y} - n\hat{\beta}_0 - \hat{\beta}_1\bar{x}\cdot n$
$= n(\bar{y} - \hat{\beta}_0 - \hat{\beta}_1\bar{x})$

Goal $\boxed{\hat{\beta}_1 = \sum \binom{\text{related}}{\text{to } x} Y_i}$

$Y_i = \beta_0 + \beta_1 X_i + e_i$

$E(Y_i) = \beta_0 + \beta_1 X_i$

$Var(Y_i) = Var(e_i) = \sigma^2$

If $\hat{\beta}_1 = \sum a_i Y_i$ ← non-random

then $E(\hat{\beta}_1) = \sum a_i E(Y_i) = \cdots$

$Var(\hat{\beta}_1) = \sum a_i^2 Var(Y_i) = \cdots$

$$\hat{\beta}_1 = \frac{\sum (Y_i - \bar{Y})(X_i - \bar{X})}{SXX}$$

$$= \sum \frac{(X_i - \bar{X})}{SXX} Y_i - \frac{\sum \bar{Y}(X_i - \bar{X})}{SXX}$$

$\rightarrow \bar{Y} \underbrace{\sum (X_i - \bar{X})}_{0}$

$\sum (X_i - \bar{X})(X_i - \bar{X})$

$$E(\hat{\beta}_1) = \sum \frac{X_i - \bar{X}}{SXX} E(Y_i) = \sum \frac{(X_i - \bar{X})}{SXX} (\beta_0 + \beta_1 X_i) = 0 + \beta_1 \frac{\sum (X_i - \bar{X}) X_i}{SXX}$$

$$\cdots\cdots = \beta_1 \checkmark$$

$$Var(\hat{\beta}_1) = \sum \left(\frac{X_i - \bar{X}}{SXX}\right)^2 Var(Y_i) = \sigma^2 \frac{\sum (X_i - \bar{X})^2}{SXX^2} = \frac{\sigma^2}{SXX} \xrightarrow[n \to \infty]{?} 0$$

Properties of least squares estimators

1. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ can be written as a linear combination of y_i
   - $\widehat{\beta}_1 = \sum \left(\frac{x_i - \bar{x}}{SXX}\right) y_i \quad \hat{\beta}_0 = \sum \left[\frac{1}{n} - \bar{x}\left(\frac{x_i - \bar{x}}{SXX}\right)\right] y_i$
2. The fitted line passes through $(\bar{x}, \bar{y})$
   - Check the derivative of RSS w.r.t. β_0
3. The sum of residuals=0 (not errors!)
   - Check the derivative of RSS w.r.t. β_0
4. The estimators are unbiased (Expectation=true value)

unbiase → $\mathrm{E}(\widehat{\beta}_0|X) = \beta_0, \quad \mathrm{E}(\widehat{\beta}_1|X) = \beta_1, \quad \mathrm{E}(\widehat{\sigma}^2|X) = \sigma^2$

consistent → $\hat{\beta}_0 \longrightarrow \beta_0$

Variance of least squares estimators

Unknown True Variance	Estimators of Variance
done $\mathrm{Var}(\widehat{\beta}_1 \mid X) = \sigma^2 \frac{1}{SXX}$	$\widehat{\mathrm{Var}}(\widehat{\beta}_1 \mid X) = \widehat{\sigma}^2 \frac{1}{SXX}$
DIY $\mathrm{Var}(\widehat{\beta}_0 \mid X) = \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{SXX})$	$\widehat{\mathrm{Var}}(\widehat{\beta}_0 \mid X) = \widehat{\sigma}^2 (\frac{1}{n} + \frac{\bar{x}^2}{SXX})$
DIY $\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1 \mid X) = -\sigma^2 \frac{\bar{x}}{SXX}$	$\widehat{\mathrm{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1 \mid X) = -\widehat{\sigma}^2 \frac{\bar{x}}{SXX}$
$\rho(\widehat{\beta}_0, \widehat{\beta}_1) = \frac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$	$\widehat{\rho}(\widehat{\beta}_0, \widehat{\beta}_1) = \frac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$
$\frac{\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1)}{\sqrt{\mathrm{Var}(\widehat{\beta}_0)\ \mathrm{Var}(\widehat{\beta}_1)}}$	

See textbook p.273-275

Variance of least squares estimators

unknown

known
estimated
from
data

Unknown True Variance	Estimators of Variance
$\mathrm{Var}(\widehat{\beta}_1 \mid X) = \sigma^2 \dfrac{1}{SXX}$	$\widehat{\mathrm{Var}}(\widehat{\beta}_1 \mid X) = \widehat{\sigma}^2 \dfrac{1}{SXX}$
$\mathrm{Var}(\widehat{\beta}_0 \mid X) = \sigma^2 \left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{SXX}\right)$	$\widehat{\mathrm{Var}}(\widehat{\beta}_0 \mid X) = \widehat{\sigma}^2 \left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{SXX}\right)$
$\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1 \mid X) = -\sigma^2 \dfrac{\bar{x}}{SXX}$	$\widehat{\mathrm{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1 \mid X) = -\widehat{\sigma}^2 \dfrac{\bar{x}}{SXX}$
$\rho(\widehat{\beta}_0, \widehat{\beta}_1) = \dfrac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$	$\widehat{\rho}(\widehat{\beta}_0, \widehat{\beta}_1) = \dfrac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$

See textbook p.273-275

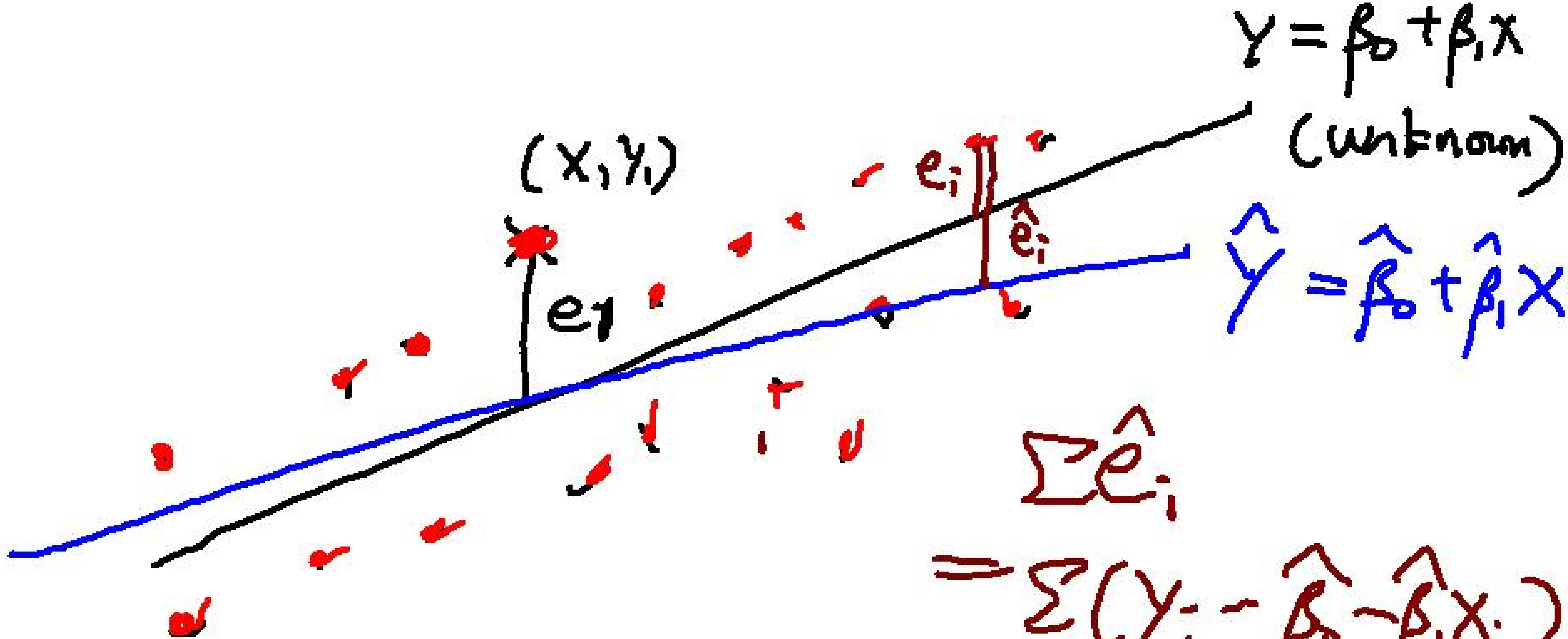
Variance of least squares estimators

Unknown True Variance	Estimators of Variance
$\mathrm{Var}(\hat{\beta}_1 \mid X) = \sigma^2 \frac{1}{SXX}$	$\widehat{\mathrm{Var}}(\hat{\beta}_1 \mid X) = \hat{\sigma}^2 \frac{1}{SXX}$
$\mathrm{Var}(\hat{\beta}_0 \mid X) = \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{SXX})$	$\widehat{\mathrm{Var}}(\hat{\beta}_0 \mid X) = \hat{\sigma}^2 (\frac{1}{n} + \frac{\bar{x}^2}{SXX})$
$\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1 \mid X) = -\sigma^2 \frac{\bar{x}}{SXX}$	$\widehat{\mathrm{Cov}}(\hat{\beta}_0, \hat{\beta}_1 \mid X) = -\hat{\sigma}^2 \frac{\bar{x}}{SXX}$
$\rho(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$	$\hat{\rho}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$

$\rho(\hat{\beta}_0, \hat{\beta}_1) = \hat{\rho}(\hat{\beta}_0, \hat{\beta}_1)$

See textbook p.273-275

God



$$\sum \hat{e}_i$$
$$= \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$
$$= n(\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X})$$
$$= 0$$