

Rep:

Model $\vec{y} = X\vec{\beta} + \vec{\varepsilon}$

Assume 1) $E(\vec{\varepsilon}) = \vec{0}$

2,3) $\text{Cov}(\vec{\varepsilon}) = \sigma^2 I$

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

$\text{Cov}(y_i, y_j) = 0$

4)* $\vec{\varepsilon} \sim \text{MVN}_n(\vec{0}, \sigma^2 I)$.

Require X $n \times (k+1)$, $k+1 < n$, full rank
 $\uparrow$ k covariates $\uparrow$ column of ones.

Estimator for $\vec{\beta}$: $\hat{\beta} = (X^T X)^{-1} X^T \vec{y}$

minimizes $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \text{SSE}$

$\hat{y}_i = \vec{x}_i^T \hat{\beta}$

$\uparrow$ row i of X .

7.3.2 Properties of $\hat{\beta}$ (A1,2,3)

Thm 7.3b $\hat{\beta}$ unbiased

Proof $E(\hat{\beta}) = E((X^T X)^{-1} X^T \vec{y})$

" $A\vec{y}$ "

$= (X^T X)^{-1} X^T E(\vec{y})$

$= (X^T X)^{-1} X^T X \vec{\beta} = \vec{\beta}$

Thm 7.3C $\text{Cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$

Proof: $\text{Cov}(\hat{\beta}) = \text{Cov}(X\vec{y}) = A \Sigma A^T$

$$= (X^T X)^{-1} X^T \overset{\sigma^2 I}{\Sigma} X (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1}$$

R: π example

$$\hat{\beta}_0 = -0.42 \quad \hat{\beta}_1 = 3.14$$

$$\begin{aligned} \text{SD}(\hat{\beta}_1) &\approx 0.0002 \\ \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= -1.7 \cdot 10^{-3} \end{aligned}$$

$$\text{Var}(\hat{\beta}_0) = 2.26 \quad \text{Var}(\hat{\beta}_1) = 3.1 \times 10^{-6}$$

7.3.3 Estimator for σ^2

Requires estimate of σ^2 .

Note $\sigma^2 = \text{Var}(Y_i)$ for all i

$$\stackrel{\text{def}}{=} E[(Y_i - E(Y_i))^2]$$

↑ can estimate expected values by sample average, we have n realizations.

$$E(Y_i) = \vec{X}_i^T \vec{\beta} \quad \text{unknown mean} \quad \rightarrow \quad \hat{E}(Y_i) = \vec{X}_i^T \hat{\beta} = \hat{Y}_i$$

↑ just notation.

Try $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$

Recall: $\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T \vec{y} = H \vec{y}$

$H = X(X^T X)^{-1} X^T$ symmetric, idempotent, rank $k+1$

Proof: $H^T = (X(X^T X)^{-1} X^T)^T$

$= X \underbrace{(X^T X)^{-1}}_{(X^T)^T} X^T$

$= X(X^T X)^{-1} X^T = H$ symmetric

$H H = X(X^T X)^{-1} X^T \underbrace{X(X^T X)^{-1} X^T}_I$

$= X(X^T X)^{-1} X^T$ idempotent

$\text{rank}(H) = \text{trace}(H)$ by thm 2.13 d)

Use $\text{trace}(AB) = \text{trace}(BA)$

$\text{trace}(H) = \text{trace}(X^T X (X^T X)^{-1})$

$= \text{trace}(I_{k+1}) = k+1$ rank

Back to $\hat{\sigma}^2$:

$\hat{\sigma}^2 = \frac{1}{n} (\vec{y} - \hat{y})^T (\vec{y} - \hat{y})$

$$\vec{y} - \hat{y} = (I - H) \vec{y}$$

↓

$$(\vec{y} - \hat{y})^T (\vec{y} - \hat{y}) =$$

$$((I - H) \vec{y})^T (I - H) \vec{y} = \vec{y}^T (I - H) \vec{y}$$

"quadratic form"

$$E(\vec{y}^T (I - H) \vec{y}) = \text{tr}((I - H) \overset{\sigma^2 I}{\Sigma})$$

'trace' formula

$$+ E(\vec{y})^T (I - H) E(\vec{y})$$

$$\text{tr}(I - H) = n - (k + 1)$$

$$\begin{aligned} (I - H) E(\vec{y}) &= (I - H) X \vec{\beta} = X \vec{\beta} - X(X^T X)^+ X^T X \vec{\beta} \\ &= X \vec{\beta} - X \vec{\beta} = 0 \end{aligned}$$

$$E(\vec{y}^T (I - H) \vec{y}) = \sigma^2 (n - (k + 1))$$

$$E(\hat{\sigma}^2) = \sigma^2 \frac{n - (k + 1)}{n} \quad \text{biased.}$$

$$s^2 = \frac{1}{n - (k + 1)} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n - (k + 1)} \vec{y}^T (I - H) \vec{y}$$

unbiased estimator for σ^2 .

R^2

Π example

$$s^2 = \frac{1}{7-2} \sum_{i=1}^7 (y_i - \hat{y}_i)^2 = 9.35 \quad s \approx 3.058 \dots$$

model-summary - -

So far: $\hat{\beta} = (X^T X)^{-1} X^T \vec{y}$ and $s^2 = \frac{1}{n-k+1} \vec{y}^T (I-H) \vec{y}$

unbiased estimators for $\vec{\beta}$ and σ^2 .

Now: Add assumption 4, $\varepsilon \sim \text{MVN}_n(\vec{0}, \sigma^2 I)$.

chap. 7.6 Normal model

Thm 7.6a If $\vec{y} \sim \text{MVN}_n(X\vec{\beta}, \sigma^2 I)$ then

$\hat{\beta}$ and $\hat{\sigma}^2$ are maximum likelihood estimators.

"Proof:" Use $L(\vec{\beta}, \sigma^2) = f(\vec{y})$

$$= \frac{1}{\sqrt{(2\pi)^n |\sigma^2 I|}} e^{-\frac{1}{2} (\vec{y} - X\vec{\beta})^T (\sigma^2 I)^{-1} (\vec{y} - X\vec{\beta})}$$

NB: MLE $\hat{\sigma}^2$ biased $\rightarrow$ often use s^2

Thm 7.6 b Properties

$$(i) \hat{\beta} \sim MVN_{k+1}(\vec{\beta}, \sigma^2 (X^T X)^{-1})$$

$$(ii) \frac{n\hat{\sigma}^2}{\sigma^2} = \frac{n-(k+1)s^2}{\sigma^2} \sim \chi^2(n-k-1)$$

(iii) $\hat{\beta}$ and $\hat{\sigma}^2$ (s^2) independent.

Proof:

$$(i) \hat{\beta} = A\vec{Y} \quad A = (X^T X)^{-1} X^T \quad \text{thm } \underline{\hspace{2cm}}$$

$$(ii) \frac{1}{\sigma^2} \vec{Y}^T (I-H) \vec{Y} \quad \text{Corollary 1 to thm 5.5.}$$

(iii) independence of linear and quadratic forms:

$$\text{Cov}(\underbrace{A\vec{Y}}_{(X^T X)^{-1} X^T}, \underbrace{B\vec{Y}}_{(I-H)}) = A \Sigma B^T = 0 \quad \Sigma = \sigma^2 I$$

$$(X^T X)^{-1} X^T (I-H)$$

$$X^T - X^T X (X^T X)^{-1} X^T = X^T - X^T = 0.$$

7.7 R^2 ← how good is the model?

Sums of Squares

← Total sum of squares.

$$\begin{aligned} SST &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \vec{y}^T (I - \frac{1}{n} \mathbf{1} \mathbf{1}^T) \vec{y} \end{aligned}$$

variation in y , no model

$$\begin{aligned} SSE &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \vec{y}^T (I - H) \vec{y} \end{aligned}$$

variation around the line (model)

$$\begin{aligned} SST &= \vec{y}^T (I - \frac{1}{n} \mathbf{1} \mathbf{1}^T + H - H) \vec{y} \\ &= \vec{y}^T (H - \frac{1}{n} \mathbf{1} \mathbf{1}^T) \vec{y} + \vec{y}^T (I - H) \vec{y} \\ &= SSR + SSE \end{aligned}$$

SSR = regression sum of squares

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \vec{y}^T (H - \frac{1}{n} \mathbf{1} \mathbf{1}^T) \vec{y}$$

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$$

"coefficient of determination"

$$0 \leq R^2 \leq 1$$

↑
if $SSE \approx SST$,
model did not
explain more
variation in y .

↑ if $SSE \ll SST$, model
'removed' most of
the variation in y .

NB * adding a new covariate x^* to model increases (cannot decrease) R^2 even if x^* is not meaningful.

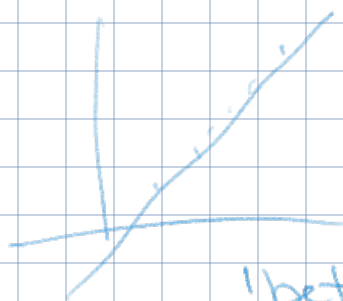
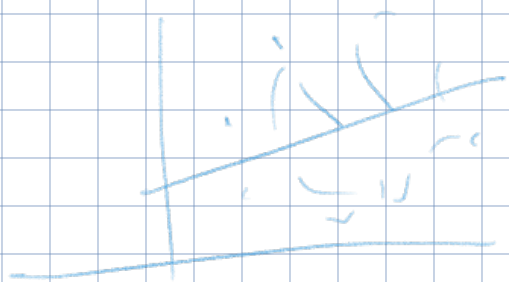
* If all β -s = 0, $E(R^2) = \frac{k}{n-1}$
(not β_0)

corrections $\rightarrow E(R_a^2) = 0$ if all β -s = 0 (1)

+ $R_a^2 = 1$ if $R^2 = 1$ (2)

$$R_a^2 = \underbrace{\left(R^2 - \frac{k}{n-1}\right)}_{(1)} \underbrace{\frac{(n-1)}{n-k-1}}_{(2)} = \frac{(n-1)R^2 - k}{n-k-1}$$

consider



'better' model.