

$$\text{Model: } \vec{Y} = X\vec{\beta} + \vec{\varepsilon} \quad E(\vec{Y}) = X\vec{\beta} \quad \text{cov}(\vec{Y}) = \sigma^2 I$$

$$\vec{Y} \sim \text{MVN}_n(X\vec{\beta}, \sigma^2 I) \quad *$$

Unbiased estimator

$$\hat{\beta} = (X^T X)^{-1} X^T \vec{Y} \quad \rightarrow \quad \hat{Y} = E(\hat{\vec{Y}}) = X\hat{\beta} = H\vec{Y}$$

$$\hat{\beta} \sim \text{MVN}_{k+1}(\vec{\beta}, \sigma^2 (X^T X)^{-1}) \quad *$$

7.9 Model misspecification (Pt. 1)

Pt. 2
about
 $\sigma^2 I$ later.

$$X = [X_1 \quad X_2]$$

$$(*) \quad \vec{Y} = X_1 \vec{\beta}_1 + X_2 \vec{\beta}_2 + \vec{\varepsilon} \quad \text{true, full model}$$

← includes intercept

$$(**) \quad \vec{Y} = X_1 \vec{\beta}_1^* + \vec{\varepsilon} \quad \text{reduced, 'wrong', model.}$$

Consequences of assuming (**):

$$\hat{\beta}_1^* = (X_1^T X_1)^{-1} X_1^T \vec{Y}$$

Thm 7.9a (bias)

$$\begin{aligned} E(\hat{\beta}_1^*) &= (X_1^T X_1)^{-1} X_1^T (X_1 \vec{\beta}_1 + X_2 \vec{\beta}_2) \\ &= \vec{\beta}_1 + \underbrace{(X_1^T X_1)^{-1} X_1^T X_2}_{\text{bias depends on } X_1, X_2, \vec{\beta}_2} \vec{\beta}_2 \end{aligned}$$

bias depends on $X_1, X_2, \vec{\beta}_2$.

Thm 7.9c (i) (variance)

$$\hat{\beta} = (X^T X)^{-1} X^T \vec{y} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \text{ 'true' model}$$

For β_j in $\vec{\beta}_1$, $\text{Var}(\hat{\beta}_j) > \text{Var}(\hat{\beta}_j^*)$

"Bias-variance tradeoff"

Underfitting (**) $\rightarrow$ high bias

Overfitting (*) $\rightarrow$ high variance

In terms of bias we should add as many variables to our model as possible, but the cost is variance $\rightarrow$ greater uncertainty.

Thm 7.10

If $X_1^T X_2 = 0$, then $\hat{\beta}_1^* = \hat{\beta}_1$ (and $E(\hat{\beta}_1^*) = \vec{\beta}_1$)

Proof:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^T X)^{-1} X^T \vec{y}$$

$$X^T X = \begin{pmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{pmatrix} \stackrel{X_1^T X_2 = 0}{=} \begin{pmatrix} X_1^T X_1 & 0 \\ 0 & X_2^T X_2 \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} (X_1^T X_1)^{-1} & 0 \\ 0 & (X_2^T X_2)^{-1} \end{pmatrix} \begin{pmatrix} X_1^T \\ X_2^T \end{pmatrix} \vec{y} = \begin{pmatrix} (X_1^T X_1)^{-1} X_1^T \vec{y} \\ (X_2^T X_2)^{-1} X_2^T \vec{y} \end{pmatrix}.$$

→ For designing experiments,
we may prefer to choose x -s so
that they are orthogonal.

7.7 R^2 'coefficient of determination'

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 = \vec{y}^T (I - \frac{1}{n} J) \vec{y}$$

variation
around the
average,
a very simple
model

↑
total

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

variation around 'the line',
more complex model

↑
error

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

"explained
variation"
"small" if $\hat{y}_i \approx \bar{y}$,
"large" if $\hat{y}_i \approx y_i$.

$$SST = SSR + SSE$$

$$SSR = SST - SSE$$

Explained variance is total variance
minus 'model' variance.

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

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

↑
model
explains
very little
 $SSE \approx SST$

↑
model
explains
a lot,
 $SSR \approx SST$

R example
groceries

NB: $R_a^2 = \left(R^2 - \frac{k}{n-1} \right) \frac{n-1}{n-k-1}$
adjusted

- Ensures $E(R_a^2) = 0$ if all $\beta_1, \dots, \beta_k = 0$.
while $E(R^2) = k/n-1$
- Punishes addition of many x-s (large k)
whereas R^2 always increases (no decrease)

R example

↑ Note final line of model summary.

Chap. 8 Tests + confidence intervals

8.1 Test of overall regression

$$\vec{Y} = X\vec{\beta} + \vec{\varepsilon} = \beta_0 + X_1\vec{\beta}_1 + \vec{\varepsilon} \quad \varepsilon \sim MVN_n(\vec{0}, \sigma^2 I)$$

$$H_0: \vec{\beta}_1 = \vec{0}$$

$$H_1: \vec{\beta}_1 \neq \vec{0}$$

at least one
coef. in $\vec{\beta}_1$ different
from 0.

Test statistic:

$$F = \frac{SSR/k}{SSE/(n-k-1)} = \frac{(SST - SSE)/k}{SSE/(n-k-1)}$$

very simple
↓ model
 H_0 ← full model
 H_1

3 min: values of F under H_0, H_1 .

H_0 : small H_1 : large

Thm 8.1 d

(i) if $H_0: \vec{\beta}_1 = 0$ false, $F \sim F(k, n-k-1, \lambda)$

$$\lambda = \frac{1}{2\sigma^2} \vec{\beta}^T \underbrace{X^T (I - \frac{1}{n} J)}_{X^T X_c} X^T \vec{\beta}$$

$X^T X_c$ 'centered' X -S.

(ii) if H_0 true,

$$F \sim F(k, n-k-1)$$

Proof (def F)

$$\text{Use } F = \frac{\frac{1}{\sigma^2} SSR/k}{\frac{1}{\sigma^2} SSE/(n-k-1)}$$

$$\text{We know: } (n-k-1) \frac{s^2}{\sigma^2} = \frac{SSE}{\sigma^2} \sim \chi^2(n-k-1) \checkmark$$

Need $\frac{1}{\sigma^2} SSR \sim \chi^2(k, \lambda)$ and SSR, SSE independent.

$$SSR = \vec{y}^T \underbrace{(H - \frac{1}{n} \mathbf{1}\mathbf{1}^T)}_{\text{idempotent? symmetric } V} \vec{y}$$

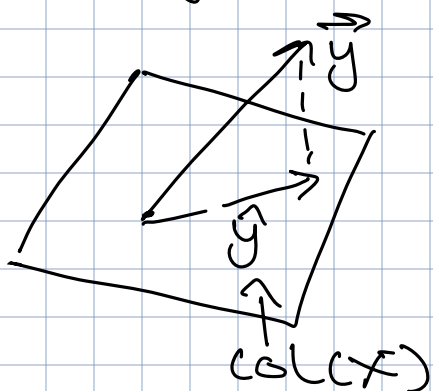
$$\begin{aligned} (H - \frac{1}{n} \mathbf{1}\mathbf{1}^T)(H - \frac{1}{n} \mathbf{1}\mathbf{1}^T) &= H \cdot H - \frac{1}{n} \mathbf{1}\mathbf{1}^T H - \frac{1}{n} H \mathbf{1}\mathbf{1}^T + (\frac{1}{n} \mathbf{1}\mathbf{1}^T)(\frac{1}{n} \mathbf{1}\mathbf{1}^T) \\ &= H - \frac{2}{n} H \mathbf{1}\mathbf{1}^T + \frac{1}{n} \mathbf{1}\mathbf{1}^T \end{aligned}$$

Note $H \mathbf{1} = \mathbf{1} \rightarrow H - \frac{1}{n} \mathbf{1}\mathbf{1}^T$ idempotent.

Proof: Use projections.

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

- An observation $\vec{y}$ of $\vec{Y}$ is not in $\text{col}(X)$, not a lin. comb. of columns of X . (deterministic model, no random noise)
- $\hat{\vec{y}} = X \hat{\beta}$ is in $\text{col}(X)$.
- H projects onto $\text{col}(X)$. (orthogonal proj.)



$\hat{\vec{y}}$ as close to $\vec{y}$ as possible, in $\text{col}(X)$.

cannot solve (for $\vec{\beta}$)

$$\begin{aligned} \vec{y} &= X \vec{\beta} \\ \downarrow \\ \text{least squares solves} \\ H \vec{y} &= X \vec{\beta}. \end{aligned}$$

$$\Rightarrow H \vec{j} = ?$$

Try to project vector of ones onto column space of X .

$$= \vec{j} \quad (\text{since } \vec{j} \text{ first column of } X).$$

$$\Rightarrow H\vec{j} = \vec{j}.$$

Back to SSR.

$$H - \frac{1}{n}J \quad \text{idem., sym., rank } k+1-1 = k.$$

$$\text{Then } \vec{y}^T (H - \frac{1}{n}J) \vec{y} \sim \chi^2(k, \lambda)$$

$$\begin{aligned} \lambda &= \vec{\mu}^T (H - \frac{1}{n}J) \vec{\mu} \quad \vec{\mu} = X\vec{\beta} \\ &= \vec{\beta}^T X^T (H - \frac{1}{n}J) X \vec{\beta} \end{aligned}$$

$$\begin{aligned} &\downarrow \text{NB: } HX = H \quad (\text{again by projection}) \\ &= \vec{\beta}^T X^T (I - \frac{1}{n}J) X \vec{\beta} \end{aligned}$$

$$\text{If Ho true, } X\vec{\beta} = \beta_0 \vec{j}, \lambda = \beta_0^2 \vec{j}^T (I - \frac{1}{n}J) \vec{j} = 0.$$

$$\text{Independence SSE, SSR: } (I - H)(H - \frac{1}{n}J)$$

$$= H - \frac{1}{n}J - H \cdot H + \frac{1}{n}HJ = H - \frac{1}{n}J - H + \frac{1}{n}J = 0.$$