VIII. Submodels

DATA: $(Y_i, z_i^T)^T, \ i = 1, \dots, n \qquad z_i \in \mathcal{Z} \subseteq \mathbb{R}^p$

AIM: Find a suitable linear model for $E(\mathbb{Y}|\mathbb{Z})$

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbb{Z} = \begin{pmatrix} z_1^T \\ \vdots \\ z_n^T \end{pmatrix}.$$

TWO CANDIDATES:

$$\mathbb{X}^0 = \begin{pmatrix} X_1^{0\,T} \\ \vdots \\ X_n^{0\,T} \end{pmatrix} = t_0(\mathbb{Z}), \qquad \mathbb{X} = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix} = t(\mathbb{Z})$$

Notation: $r_0 = \text{rank}(\mathbb{X}^0) \qquad \mathbb{X}^0_{n \times k_0}$

ya.s!

$r = \text{rank}(\mathbb{X}), \qquad \mathbb{X}_{n \times k}$

ASSUMPTION: $0 < r_0 \le k_0 < n$

$0 < r \le k < n, \quad r_0 < r$

MODELS: $M_0: \ \mathbb{Y}|\mathbb{Z} \sim (\mathbb{X}^0 \beta^0, \sigma^2 I_n)$

$M: \ \mathbb{Y}|\mathbb{Z} \sim (\mathbb{X}\beta, \sigma^2 I_n).$

8.1 Submodel

1

Def 8.1 Submodel

2

We say that the model M_0 is the submodel (ie the nested model) of the model M if $\mathcal{M}(X^0) \subset \mathcal{M}(X)$ with $r_0 < r$.

Notation: We will write $M_0 \subset M$.

$M_0 \equiv$ more parsimonious expression of $\mathbb{E}(Y/Z)$.

- M_0 holds $\equiv$ $\mathbb{E}(Y/Z) \in \mathcal{M}(X^0) \subset \mathcal{M}(X)$

$$\equiv \exists \beta \in \mathbb{R}^k \ \exists \beta^0 \in \mathbb{R}^{k_0} \quad \mathbb{E}(Y/Z) = X\beta = X^0\beta^0$$

- M_0 does not hold, M does:

$$\mathbb{E}(Y/Z) \in \mathcal{M}(X) \setminus \mathcal{M}(X^0)$$

$\equiv$ no $\beta^0 \in \mathbb{R}^{k_0}$ exist such that $\mathbb{E}(Y/Z) = X^0\beta^0$

8.1.1 Projection Considerations

$\mathcal{M}(X^0) \subsetneq \mathcal{M}(X) \subseteq \mathbb{R}^n$

$$\rightarrow P := (p_1, \dots, p_n) = (\overset{r_0}{Q^0}, \underbrace{\overset{r-r_0}{Q^1}, \overset{n-r}{N}}_{N^0})\ n$$

orthonormal basis of $\mathbb{R}^n$

basis constructed such that

$$(\underbrace{\overbrace{Q^0}, \overbrace{Q^1}^{\mathcal{M}(X)}, \overset{\mathcal{M}(X)^\perp}{N}}_{\mathcal{M}(X^0)^\perp})$$

$\mathcal{M}(X^0)$ $\quad$ $\mathcal{M}(X^0)^\perp$

$Q^0 \equiv \mathcal{M}(X^0)$

$N^0 := (Q^1, N) \equiv \mathcal{M}(X^0)^\perp$

$Q := (Q^0, Q^1) \equiv \mathcal{M}(X)$

$N \equiv \mathcal{M}(X)^\perp$

In the same way as in Chapter 2:

$$I_n = P^TP = PP^T = Q^0Q^{0T} + Q^1Q^{1T} + NN^T$$

$$= QQ^T + NN^T$$

$$= Q^0Q^{0T} + N^0N^{0T}$$

$H^0 := Q^0Q^{0T}$ = projections to $\mathcal{M}(X^0)$

$M^0 := N^0N^{0T} = Q^1Q^{1T} + NN^T$ ≡ projections to $\mathcal{M}(X^0)^\perp$

$H = QQ^T = H^0 + Q^1Q^{1T}$ ≡ projections to $\mathcal{M}(X)$

$M = NN^T = M^0 - Q^1Q^{1T}$ ≡ projections to $\mathcal{M}(X)^\perp$

Q^1Q^{1T} ≡ projections to $\mathcal{M}(Q^1)$

$$P = (Q^0, Q^1, N)$$

$$\underbrace{Q^0}_{\mathcal{M}(X^0)} \quad \underbrace{Q^1, N}_{\mathcal{M}(X^0)^\perp}$$

$$\underbrace{Q^0, Q^1}_{\mathcal{M}(X)} \quad \mathcal{M}(X)^\perp$$

Take $y \in \mathbb{R}^n$

$$y = Iy = (Q^0 Q^{0T} + Q^1 Q^{1T} + NN^T)y$$

$$= \underbrace{(Q^0 Q^{0T} + Q^1 Q^{1T})y}_{= Hy = \hat{y}} + \underbrace{NN^T y}_{= My = u}$$

$$= \underbrace{Q^0 Q^{0T} y}_{= H^0 y =: \hat{y}^0} + \underbrace{(Q^1 Q^{1T} + NN^T)y}_{= M^0 y =: u^0}$$

$$\left.\begin{array}{ll} \hat{y} \in \mathcal{M}(X), & \hat{y}^0 \in \mathcal{M}(X^0) \\ u \in \mathcal{M}(X)^\perp, & u^0 \in \mathcal{M}(X^0)^\perp \end{array}\right\} \text{all are projections of } y \text{ into respective spaces}$$

$$d := Q^1 Q^{1T} y = \hat{y} - \hat{y}^0 = u^0 - u$$

$= \text{projection of } y \text{ into } \mathcal{M}(Q^1)$

8.1.2 Properties of submodel related quantities

4

$\mathsf{M}_0 : Y \mid Z \sim (X^0 \beta^0, \sigma^2 I_n)$, $\mathsf{M} : Y \mid Z \sim (X\beta, \sigma^2 I_n)$

Notation:

$$\hat{Y}^0 := H^0 Y = Q^0 Q^{0T} Y$$

$$U^0 := Y - \hat{Y}^0 = M^0 Y = (Q^1 Q^{1T} + NN^T) Y$$

$$SS_e^0 = \| U^0 \|^2$$

$$\nu_e^0 = n - r_0$$

$$MS_e^0 = \frac{SS_e^0}{\nu_e^0}$$

"standard" quantities based on submodel

↑

$$\mathbb{D} := Q^1 Q^{1T} Y = \hat{Y} - \hat{Y}^0 = U^0 - U$$ = projection of Y into $\mathcal{M}(Q^1)$

Theorem 8.1 On a submodel

5

Consider two linear models $\mathsf{M} : Y \mid Z \sim (X\beta, \sigma^2 I_n)$ and $\mathsf{M}_0 : Y \mid Z \sim (X^0\beta^0, \sigma^2 I_n)$ such that $\mathsf{M}_0 \subset \mathsf{M}$. Let the submodel holds, i.e., let $\mathsf{E}(Y \mid Z) \in \mathcal{M}(X^0)$.

Then

(i) $\hat{Y}^0$ is BLUE of a vector parameter $\mu^0 = X^0\beta^0 = \mathsf{E}(Y \mid Z)$.

→ Theorem 2.4 (Gauss-Markov).

(ii) The submodel residual mean square MS_e^0 is the unbiased estimator of the residual variance σ^2.

→ Lemma 2.7 (Moments of residual sum of squares).

(iii) Statistics ~~are~~ $\hat{Y}^0$ and U^0 are conditionally, given Z, uncorrelated.

→ Theorem 6.2 (iv) ⇐ calculations inside the proof

$\boxed{5}$

(iv) a random vector $D = \hat{Y} - \hat{Y}^0 = U^0 - U$ satisfies $\|D\|^2 = SS_e^0 - SS_e$.

Proof:

$$P = \Big(\underbrace{\overset{\mathcal{M}(X^0)}{Q^0},\ \overbrace{Q^1}^{\mathcal{M}(X^0)^\perp}}_{\mathcal{M}(X)}\ ,\ \underset{\mathcal{M}(X)^\perp}{N}\Big)$$

D is projection of Y

$$D = \hat{Y} - \hat{Y}^0 = U^0 - U$$

$$U^0 = D + U = \underbrace{Q^1 Q^{1T} Y}_{D \in \mathcal{M}(Q^1)} + \underbrace{N N^T Y}_{U \in \mathcal{M}(N)}$$

$D \perp U$ (projections into orthogonal subspaces of R^n)

$\Rightarrow$ Pythagoras $\underbrace{\|U^0\|^2}_{SS_e^0} = \|D\|^2 + \underbrace{\|U\|^2}_{SS_e}$

i.e. $\|D\|^2 = SS_e^0 - SS_e$.

∎

(v) If additionally a normal linear model is assumed, i.e., if $Y \mid Z \sim \mathcal{N}_n(X^0\beta^0, \sigma^2 I_n)$ then the statistics $\hat{Y}^0$ and U^0 are conditionally, given Z, independent and

$$F_0 = \frac{\dfrac{SSe^0 - SSe}{r - r_0}}{\dfrac{SSe}{n-r}} = \frac{\dfrac{SSe^0 - SSe}{\nu_e^0 - \nu_e}}{\dfrac{SSe}{\nu_e}} \sim F_{r-r_0,\, n-r} \equiv F_{\nu_e^0 - \nu_e,\, \nu_e}$$

Comments before proof:

$$F_0 = \frac{\dfrac{\|D\|^2}{r - r_0}}{\dfrac{SSe}{n-r}} \qquad D = U^0 - U = \hat{Y} - \hat{Y}^0$$

IDEA: if M_0 holds then $\hat{Y}^0 \simeq \hat{Y}$, $D \simeq 0$

$SSe^0 \simeq SSe$

$F_0 \equiv$ reasonable quantification to evaluate whether M_0 holds

---

- independence of $\hat{Y}^0$ and U^0:
  Theorem 6.2 (iv) (LSE under normality)

- $$F_0 = \frac{\dfrac{\|D\|^2}{\sigma^2(r - r_0)} \sim \chi^2_{r-r_0} \ ?}{\dfrac{SSe}{\sigma^2(n-r)} \sim \chi^2_{n-r} \ ?} \qquad \underline{\perp\!\!\!\perp} \ ?$$

(a) $\dfrac{SSe}{\sigma^2} \sim \chi^2_{n-r}$ by Theorem 6.2 (vii) ← both conditionally given Z and unconditionally

M_0 holds (it is assumed)

$\Longrightarrow$ M holds as well and hence $\dfrac{SSe}{\sigma^2} \sim \chi^2_{n-r}$

(b) $\|D\|^2 = \|\underbrace{Q^1 Q^{1T} Y}_{\text{projection of } Y}\|^2 = Y^T Q^1 \underbrace{Q^{1T} Q^1}_{I} Q^{1T} Y =$

$$= Y^T Q^1 Q^{1T} Y = \|Q^{1T} Y\|^2$$

$$Q^{1T} Y \,|\, Z \sim N(\underbrace{Q^{1T} X^0 \beta^0}_{0},\ \sigma^2 \underbrace{Q^{1T} Q^1}_{I_{r-r_0}})$$

Remember: $P = (\overset{r_0}{\underset{M(X^0)}{Q^0}}, \underbrace{\overset{r-r_0}{Q^1}, \overset{n-r}{N}}_{M(X^0)^\perp})$ = orthonormal basis of $\mathbb{R}^n$

$$\Rightarrow \frac{1}{\sigma} Q^{1T} Y \,|\, Z \sim N(0, I_{r-r_0})$$

$$\Rightarrow \left\|\frac{1}{\sigma} Q^{1T} Y\right\|^2 = \frac{1}{\sigma^2}\|D\|^2 \sim \chi^2_{r-r_0}$$

(conditionally given Z)

conditional distribution does not depend on condition, so also unconditionally

---

(c) numerator = $\|D\|^2$, denominator = $SS_e = \|U\|^2$

$$\begin{array}{l} \text{numer.} \\ \text{denom.} \end{array} \begin{pmatrix} D \\ U \end{pmatrix} = \begin{pmatrix} Q^1 Q^{1T} \\ N N^T \end{pmatrix} \underset{\sim N \text{ (given } Z)}{Y} \Rightarrow \begin{pmatrix} D \\ U \end{pmatrix} \Big| Z \sim \underset{\text{(jointly)}}{N}$$

$$\operatorname{cov}(D, U \,|\, Z) = \operatorname{cov}(Q^1 Q^{1T} Y, N N^T Y \,|\, Z) =$$

$$= Q^1 Q^{1T} \underbrace{\operatorname{var}(Y|Z)}_{\sigma^2 I_n} N N^T = \sigma^2 Q^1 \underbrace{Q^{1T} N}_{0} N^T = 0$$

$$\Rightarrow \text{given } Z:\ D \perp\!\!\!\perp U \Rightarrow \|D\|^2 \perp\!\!\!\perp \|U\|^2$$

That is $F_0 = \dfrac{\boxed{\dfrac{\|D\|^2}{\sigma^2(r-r_0)}} \sim \chi^2_{r-r_0} \text{ (given } Z)}{\boxed{\dfrac{\|U\|^2}{\sigma^2(n-r)}} \sim \chi^2_{n-r}} \ \perp\!\!\!\perp \ \sim F_{r-r_0,\, n-r}$ (given Z)

Conditional distrib. does not depend on condition

$\Rightarrow$ also unconditionally $F_0 \sim F_{r-r_0,\, n-r}$ $\square$

8.1.3 Series of submodels

$\boxed{7}$

Let us consider <u>three</u> candidate models

M_0: $Y/Z \sim (X^0\beta^0, \sigma^2 I_n)$, $X^0 = t_0(Z)$

M_1: $\sim (X^1\beta^1, \sigma^2 I_n)$, $X^1 = t_1(Z)$

M : $\sim (X\beta, \sigma^2 I_n)$, $X = t(Z)$

$$\mathcal{M}(X^0) \subsetneqq \mathcal{M}(X^1) \subsetneqq \mathcal{M}(X)$$

ranks: $r_0 < r_1 < r < n$

$$M_0 \subset M_1 \subset M$$

$\boxed{8}$

<u>Notation</u>:

- $\hat{Y}^0, U^0, SS_e^0, \nu_e^0 = n - r_0, MS_e^0$
  - quantities based on model M_0
- $\hat{Y}^1, U^1, SS_e^1, \nu_e^1 = n - r_1, MS_e^1$
  - quantities based on model M_1
- $\hat{Y}, U, SS_e, \nu_e = n - r, MS_e$
  - quantities based on model M

<u>Theorem 8.2</u> <u>On submodels</u>

$\boxed{9}$

Consider three <u>normal</u> linear models $M: Y/Z \sim N_n(X\beta, \sigma^2 I_n)$, $M_1: Y/Z \sim N_n(X^1\beta^1, \sigma^2 I_n)$, $M_0: Y/Z \sim N_n(X^0\beta^0, \sigma^2 I_n)$ such that $M_0 \subset M_1 \subset M$. Let the (smallest) submodel M_0 hold, i.e., let $E(Y/Z) \in \mathcal{M}(X^0)$. Then

$$9] \quad F_{0,1} = \frac{\frac{SS_e^0 - SS_e^1}{r_1 - r_0}}{SS_e/(n-r)} = \frac{\frac{SS_e^0 - SS_e^1}{\nu_e^0 - \nu_e^1}}{\frac{SS_e}{\nu_e}} \sim F_{r_1 - r_0,\, n-r}.$$

Proof: Let us construct an orthonormal vector basis of $\mathbb{R}^n$ as follows:

$$P = (\underbrace{\overset{r_0}{Q^0}}_{\mathcal{M}(X^0)}, \underbrace{\overset{r_1-r_0}{Q^1}, \overset{r-r_1}{Q^2}, \overset{n-r}{N}}_{\mathcal{M}(X^0)^\perp})$$

$$\underbrace{Q^0, Q^1}_{\mathcal{M}(X^1)} \quad \underbrace{Q^2, N}_{\mathcal{M}(X^1)^\perp}$$

$$\underbrace{Q^0, Q^1, Q^2}_{\mathcal{M}(X)} \quad \underbrace{N}_{\mathcal{M}(X)^\perp}$$

$$SSe^0 = \|U^0\|^2 = \|(Q^1Q^{1T} + Q^2Q^{2T} + NN^T)Y\|^2 =$$

$$\overset{\text{Pythagoras}}{=} \|Q^1Q^{1T}Y\|^2 + \|Q^2Q^{2T}Y\|^2 + \|NN^TY\|^2 =$$

$$\overset{\text{as previously}}{=} \|Q^{1T}Y\|^2 + \|Q^{2T}Y\|^2 + \|N^TY\|^2$$

$$SSe^1 = \|U^1\|^2 = \|(Q^2Q^{2T} + NN^T)Y\|^2 = \dots$$

$$= \|Q^{2T}Y\|^2 + \|N^TY\|^2$$

$$\Rightarrow SSe^0 - SSe^1 = \|Q^{1T}Y\|^2 \quad (= \|\underbrace{Q^1Q^{1T}}_{\text{projection to } \mathcal{M}(Q^1)}Y\|^2)$$

$$SSe = \|NN^TY\|^2 = \|N^TY\|^2$$

as previously: given Z: $N^TY \perp\!\!\!\perp Q^{1T}Y$

$$\frac{1}{\sigma^2}\|Q^{1T}Y\|^2 \sim \chi^2_{r_1-r_0}$$

$$\frac{1}{\sigma^2}\|N^TY\|^2 \sim \chi^2_{n-r}$$

$$\Rightarrow F_{0,1} = \frac{\frac{\|Q^{1T}Y\|^2}{\sigma^2(r_1-r_0)}}{\frac{\|N^TY\|^2}{\sigma^2(n-r)}} \sim F_{r_1-r_0,\,n-r} \quad (\text{given } Z, \text{ for almost all values of } Z)$$

$\sim F_{r_1-r_0,\,n-r}$ also unconditionally

Let us return to only <u>two</u> models

$M_0 \subset M$

$\mathcal{M}(X^0) \subsetneq \mathcal{M}(X)$

rank: $r_0 < r$

$$P = (\underbrace{\overset{r_0}{Q^0}}_{\mathcal{M}(X^0)}, \underbrace{\overset{r-r_0}{Q^1}, \overset{n-r}{N}}_{\mathcal{M}(X^0)^\perp})$$

$$\underbrace{\qquad\qquad\qquad}_{\mathcal{M}(X)} \quad \underbrace{\qquad}_{\mathcal{M}(X)^\perp}$$

$$D = U^0 - U = \hat{Y} - \hat{Y}^0 = Q^1 Q^{1T} Y =: D(1|0)$$

$$SS_e^0 - SS_e = \|D\|^2 = \|\underbrace{Q^1 Q^{1T}}_{D(1|0)} Y\|^2$$

↑ projection of Y into $\mathcal{M}(X) \setminus \mathcal{M}(X^0) = \mathcal{M}(Q^1)$

<u>Notation</u> (Differences when dealing with a submodel)

$M_A \subset M_B$

$$D(M_B|M_A) = D(B|A) := \hat{Y}^B - \hat{Y}^A = U^A - U^B$$

$$SS(M_B|M_A) = SS(B|A) := \|D(B|A)\|^2$$

$$= SS_e^A - SS_e^B$$

= numerator of the F-statistic

8.1.4 Statistical test to compare nested models

Two normal linear models considered

$M_0: \; Y|Z \sim N_n(X^0 \beta^0, \sigma^2 I_n)$

$M: \; Y|Z \sim N_n(X\beta, \sigma^2 I_n)$, $M_0 \subset M$

$H_0: \mathbb{E}(Y|Z) \in \mathcal{M}(X^0)$

$H_1: \mathbb{E}(Y|Z) \in \mathcal{M}(X) \setminus \mathcal{M}(X^0)$

- Is model M significantly better than a (simpler) model M_0?
- Does larger regression space $\mathcal{M}(X)$ provide a significantly better expression for $\mathbb{E}(Y|Z)$ than $\mathcal{M}(X^0)$?

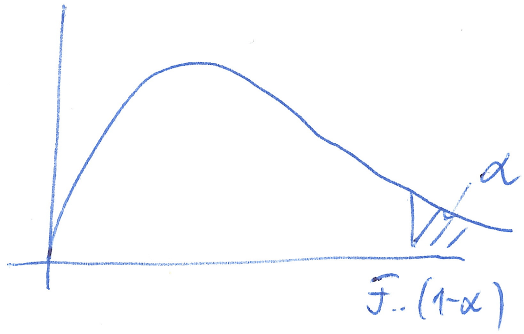
$$F_0 = \frac{\frac{SSe^0 - SSe}{r - r_0}}{\frac{SSe}{n-r}} = \frac{\frac{\|\hat{Y} - \hat{Y}^0\|^2}{r - r_0}}{\frac{SSe}{n-r}} = \frac{\frac{\|D(M|M_0)\|^2}{r - r_0}}{\frac{SSe}{n-r}} \overset{H_0}{\sim} F_{r-r_0, n-r}$$

(NORMALITY!)

but asymptotics work here...

IDEA: If M_0 holds, $D \simeq 0$

$\Rightarrow$ reasonable to reject H_0 if F_0 large



Possible critical region

$C(\alpha) = [F_{r-r_0, n-r}(1-\alpha), \infty)$

$\Rightarrow$ p-value $= 1 - CDF_{F, r-r_0, n-r}(f_0)$,

f_0 = value of F_0 obtained with given data

Three (or more) normal linear models

$M_0: \; Y|Z \sim N_n(X^0\beta^0, \sigma^2 I_n)$

$M_1: \; Y|Z \sim N_n(X^1\beta^1, \sigma^2 I_n)$

$M: \; Y|Z \sim N_n(X\beta, \sigma^2 I_n), \quad M_0 \subset M_1 \subset M$

$H_0: \; \mathbb{E}(Y|Z) \in \mathcal{M}(X^0)$

$H_1: \; \mathbb{E}(Y|Z) \in \mathcal{M}(X^1) \setminus \mathcal{M}(X^0)$

Possible test statistic

$$F_{0,1} = \frac{\dfrac{SS_e^0 - SS_e^1}{r_1 - r_0}}{\dfrac{SS_e}{n-r}} \overset{H_0}{\sim} \mathcal{F}_{\underset{\text{rank}(X^1)}{r_1} - \underset{\text{rank}(X^0)}{r_0},\; n - \overset{\text{rank}(X)}{r}}$$

To standardize $\|D(1|0)\|^2$, any larger model (than the two models being compared) can be used.