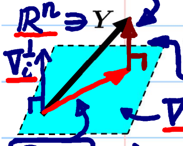
$P_{V_i}Y = (I-P_{V_i})Y$



LNp.31

Since $V_i = \text{span}\{A_i\}$, $P_{V_i} = A_i(A_i^TA_i)^{-1}A_i^T$ and $P_{V_i}Y = A_i(A_i^TA_i)^{-1}A_i^TY$

* For the model space V_i, ← i.e. consider the model: $Y \sim X_0\beta_0 + X_1\beta_1 + \cdots + X_i\beta_i$

$$RSS_{V_i} = \|Y\|^2 - \|P_{V_i}Y\|^2 = Y^TY - (P_{V_i}Y)^TP_{V_i}Y = Y^TY - Y^TP_{V_i}^TP_{V_i}Y$$

$$= Y^TY - Y^TP_{V_i}Y = Y^T(I-P_{V_i})Y = \|(I-P_{V_i})Y\|^2$$

$= P_{V_i}P_{V_i} = P_{V_i}^2 = P_{V_i}$

$P_{V_i^\perp}^T P_{V_i^\perp} = P_{V_i^\perp} =$

$\mathbb{R}^n = V_i \oplus V_i^\perp$

$V_i = V_{i-1} \oplus W_i$

3/16 (exercise, prove that this is an orthogonal projection matrix.

Since $W_i = V_i \cap V_{i-1}^\perp$ and $V_{i-1} \subset V_i$,

Hint. $P_{V_{i-1}}P_{V_i} = P_{V_i}P_{V_{i-1}} = P_{V_{i-1}}$)

For a vector space V, $Y^TP_VY = \|P_VY\|^2$ is a quadratic form.

$$P_{W_i} = (I - P_{V_{i-1}})P_{V_i} = P_{V_i} - P_{V_{i-1}}P_{V_i} = P_{V_i} - P_{V_{i-1}},$$

and $P_{W_i}Y = (I-P_{V_{i-1}})P_{V_i}Y = (P_{V_i} - P_{V_{i-1}})Y = P_{V_i}Y - P_{V_{i-1}}Y$.

check the graphs in LNp.32

* Since $\mathbb{R}^n = W_0 \oplus W_1 \oplus \cdots \oplus W_k \oplus V_k^\perp$ and $W_0 \perp W_1 \perp \cdots \perp W_k \perp V_k^\perp$,

- $Y = P_{W_0}Y + P_{W_1}Y + \cdots + P_{W_k}Y + P_{V_k^\perp}Y$
- $P_{W_0}Y \perp P_{W_1}Y \perp \cdots \perp P_{W_k}Y \perp P_{V_k^\perp}Y$
- $\|Y\|^2 = \|P_{W_0}Y\|^2 + \|P_{W_1}Y\|^2 + \cdots + \|P_{W_k}Y\|^2 + \|P_{V_k^\perp}Y\|^2$ cf.

Partition of sums of squares in Type I ANOVA

cf. Type III ANOVA

check ★ in LNp.32

* When $X_0, X_1, \ldots, X_k$ are mutually orthogonal, $W_i = \text{span}\{X_i\}$

$$A_i(A_i^TA_i)^{-1}A_i^T - A_{i-1}(A_{i-1}^TA_{i-1})^{-1}A_{i-1}^T = P_{W_i} = P_{V_i} - P_{V_{i-1}} = X_i(X_i^TX_i)^{-1}X_i^T.$$

$A_i = [A_{i-1}\ X_i]$

- Consider the sequential ANOVA: for $i = 1, \ldots, k$, Recall. ANOVA in LNp.26

Under the linear model in LNp.31 $H_0^{(i)}: \beta_i = 0$ $(\omega_i = V_{i-1})$ vs. $H_A^{(i)}: \beta_i \neq 0$ $(\Omega_i = V_i)$

Type III ANOVA cf.

(P1) $RSS_{\omega_i} - RSS_{\Omega_i} = RSS_{V_{i-1}} - RSS_{V_i} = Y^T(I-P_{V_{i-1}})Y - Y^T(I-P_{V_i})Y$

Sum of squares

$$= Y^T(P_{V_i} - P_{V_{i-1}})Y = Y^TP_{W_i}Y = Y^TP_{W_i}^TP_{W_i}Y = \|P_{W_i}Y\|^2$$

check LNp.25,26 and $df_{\omega_i} - df_{\Omega_i} = \dim(W_i) = r_i$.

$P_{W_i}u_i$ may be ≈ 0 even if u_i is not (e.g., $\text{cor}(u_i, u_{i-1}) \approx 1$) $\text{cor}(u_i, u_k) \approx 1$

may not be 0 even when $H_0^{(i)}: \beta_i = 0$ ($\Rightarrow u_i = 0$) is true. (e.g., $\text{cor}(u_i, u_k) \approx 1$)

(P2) $P_{W_i}Y = P_{W_i}X\beta + P_{W_i}\varepsilon$, where

$Y = X\beta + \varepsilon$

– $P_{W_i}X\beta = (P_{V_i} - P_{V_{i-1}})(X_0\beta_0 + \cdots + X_{i-1}\beta_{i-1} + X_i\beta_i + \cdots + X_k\beta_k)$

$\because V_{i-1} = \text{span}\{X_0, \ldots, X_{i-1}\}$ and $W_i \subset V_{i-1}^\perp$, $\therefore P_{W_i}X_t = 0$ for $t = 0, 1, \ldots, i-1$

$$= (P_{V_i} - P_{V_{i-1}})(X_i\beta_i + \cdots + X_k\beta_k) = P_{W_i}(u_i + \cdots + u_k)$$

* If $X_0, X_1, \ldots, X_k$ are mutually orthogonal, $\Rightarrow u_i \in \text{span}\{X_i\} = W_i$ cf.

$$P_{W_i}X\beta = P_{W_i}u_i = (X_i(X_i^TX_i)^{-1}X_i^T)X_i\beta_i = X_i\beta_i = u_i.$$

$\text{rank}(P_{W_i}) = \dim(W_i) = r_i$

By (N1) in LNp.30 – $P_{W_i}\varepsilon \sim N(0, \sigma^2 P_{W_i}IP_{W_i}^T)$, where $P_{W_i}IP_{W_i}^T = P_{W_i}$ and $P_{W_i} = P_{W_i}^-$.

$\varepsilon \sim N(0, \sigma^2 I)$ – Thus, $P_{W_i}Y \sim N(P_{W_i}(u_i + \cdots + u_k), \sigma^2 P_{W_i})$ $(\sigma^2 P_{W_i})(\frac{1}{\sigma^2}P_{W_i})(\sigma^2 P_{W_i}) = \sigma^2 P_{W_i}$ $\quad \underbrace{\frac{1}{\sigma^2}P_{W_i}}_{(\sigma^2 P_{W_i})^-}$

(P3) $\|P_{W_i}Y\|^2 = Y^TP_{W_i}Y = (X\beta + \varepsilon)^TP_{W_i}(X\beta + \varepsilon)$

random variable

$$= (X\beta)^TP_{W_i}(X\beta) + 2(X\beta)^TP_{W_i}\varepsilon + \varepsilon^TP_{W_i}\varepsilon$$

function of parameters β

$$= \|P_{W_i}X\beta\|^2 + 2(X\beta)^TP_{W_i}\varepsilon + \varepsilon^TP_{W_i}\varepsilon,$$

FYI, $\|P_{W_i}Y\|^2/\sigma^2 \sim$ noncentral $\chi^2_{r_i}(\delta/\sigma^2)$

By (N6) in LNp.30

where $\varepsilon^TP_{W_i}\varepsilon = \varepsilon^TP_{W_i}^TP_{W_i}P_{W_i}\varepsilon = \sigma^2(P_{W_i}\varepsilon)^T(P_{W_i}/\sigma^2)(P_{W_i}\varepsilon) \sim \sigma^2\chi^2_{r_i}$

(P4) $E(RSS_{V_{i-1}} - RSS_{V_i}) = E(\boldsymbol{Y}^T \boldsymbol{P}_{W_i} \boldsymbol{Y}) = E(\|\boldsymbol{P}_{W_i}\boldsymbol{Y}\|^2)$

$= \|\boldsymbol{P}_{W_i}\boldsymbol{X}\boldsymbol{\beta}\|^2 + 2(\boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{P}_{W_i} E(\boldsymbol{\varepsilon}) + E(\boldsymbol{\varepsilon}^T \boldsymbol{P}_{W_i} \boldsymbol{\varepsilon})$

$= \|\boldsymbol{P}_{W_i}(\boldsymbol{u}_i + \cdots + \boldsymbol{u}_k)\|^2 + r_i\sigma^2$

E(sum of squares)

規

What if orthogonality exists?

隨

$=\underset{\sim}{0}$ $\sigma^2\chi^2_{r_i}$

Z^TMZ
$=(Z-\mu)^TM(Z-\mu)$
$+2\mu^TMZ-\mu^TM\mu$

– **Note**. If $\boldsymbol{M}$: a symmetric matrix, and $\boldsymbol{Z} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

Z: a random variable
$E(Z^2)=[E(Z)]^2+Var(Z)$

cf. $E(\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}) = \boldsymbol{\mu}^T\boldsymbol{M}\boldsymbol{\mu} + \text{trace}(\boldsymbol{M}\boldsymbol{\Sigma})$.

$E(Z^TMZ) = 2\mu^TM\mu - \mu^TM\mu + E[(Z-\mu)^TM(Z-\mu)]$
$= E\{tr[(Z-\mu)^TM(Z-\mu)]\}$
$= E\{tr[M(Z-\mu)(Z-\mu)^T]\}$
$= tr\{M\, \underbrace{E[(Z-\mu)(Z-\mu)^T]}_{=\Sigma}\}$

(P5) For the residual space $V_k^\perp$,

– $RSS_{V_k} = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P}_{V_k})\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{P}_{V_k^\perp}\boldsymbol{Y} = \|\boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}\|^2$

$P_{W_i}Y$ cf.

– $\boldsymbol{P}_{V_k^\perp}\boldsymbol{Y} = \underbrace{\boldsymbol{P}_{V_k^\perp}\boldsymbol{X}\boldsymbol{\beta}}_{=\underset{\sim}{0}} + \boldsymbol{P}_{V_k^\perp}\boldsymbol{\varepsilon} = \boldsymbol{P}_{V_k^\perp}\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2\boldsymbol{P}_{V_k^\perp})$

$\|P_{W_i}Y\|^2$ cf.

– $\|\boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}\|^2 = \boldsymbol{\varepsilon}^T\boldsymbol{P}_{V_k^\perp}\boldsymbol{\varepsilon} \sim \sigma^2\chi^2_{n-r}$ $dim(V_k^\perp)$

$E(RSS_{V_{i-1}} - RSS_{V_i})$ cf.

– $E(RSS_{V_k}) = E(\|\boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}\|^2) = (n-r)\sigma^2$ 隨

ε $\cancel{\in}$ span{X}

Type III ANOVA cf.

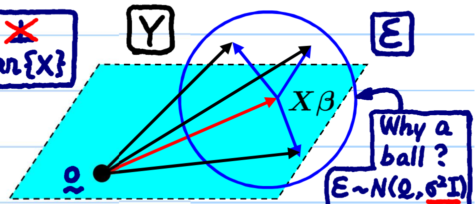
(P6) Since $W_0 \perp W_1 \perp \cdots \perp W_k \perp V_k^\perp$,

By (N4) & (N5) in LNp.30

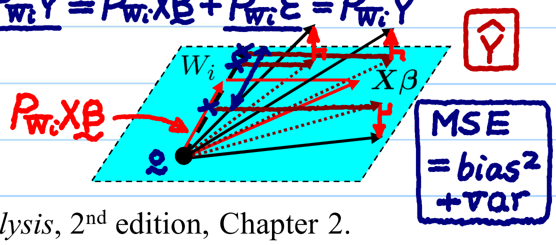
$\Rightarrow$ $\boldsymbol{P}_{W_0}\boldsymbol{Y}, \ldots, \boldsymbol{P}_{W_k}\boldsymbol{Y}, \boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}$ are *independent* random vectors

$P_{W_i}P_{W_j}^T=\underset{\sim}{0}$ for $i \neq j$

$\Rightarrow$ $\|\boldsymbol{P}_{W_0}\boldsymbol{Y}\|^2, \ldots, \|\boldsymbol{P}_{W_k}\boldsymbol{Y}\|^2, \|\boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}\|^2$ are *independent* random variables



$P_{W_i}Y = P_{W_i}X\beta + P_{W_i}\varepsilon = P_{W_i}\hat{Y}$



❖ **Further reading**: Seber and Lee (2003), *Linear Regression Analysis*, 2nd edition, Chapter 2.