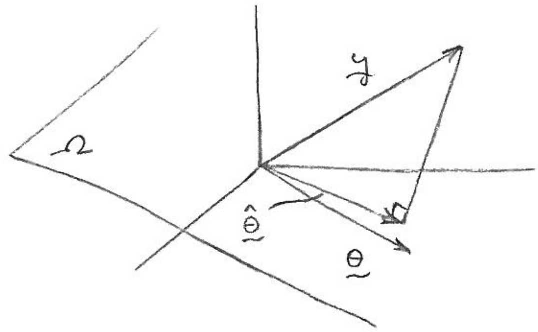
A linear model

$$\boldsymbol{y} = \boldsymbol{\theta} + \boldsymbol{u} \quad , \quad E\boldsymbol{u} = \boldsymbol{0} \quad , \quad \boldsymbol{\theta} \in \Omega$$



Let $\hat{\boldsymbol{\theta}} = P_\Omega \boldsymbol{y}$ (orthogonal projection)

$$E\hat{\boldsymbol{\theta}} = E[P_\Omega \boldsymbol{y}] = P_\Omega E\boldsymbol{y} = P_\Omega(\boldsymbol{\theta} + \boldsymbol{0}) = P_\Omega \boldsymbol{\theta} = \boldsymbol{\theta}$$

hence $\hat{\boldsymbol{\theta}}$ is an unbiased estimate for $\boldsymbol{\theta}$

Least squares principle

$$\|\boldsymbol{u}\|^2 = \|\boldsymbol{y} - \boldsymbol{\theta}\|^2 = (\boldsymbol{y} - \boldsymbol{\theta})^T(\boldsymbol{y} - \boldsymbol{\theta})$$
$$= (\boldsymbol{y} - \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\boldsymbol{y} - \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$
$$= (\boldsymbol{y} - \hat{\boldsymbol{\theta}})^T(\boldsymbol{y} - \hat{\boldsymbol{\theta}}) + 0 + 0 + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$
$$\geq (\boldsymbol{y} - \hat{\boldsymbol{\theta}})^T(\boldsymbol{y} - \hat{\boldsymbol{\theta}}) \quad , \quad \text{equality for } \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$$

The residual vector

$$\hat{\boldsymbol{u}} = \boldsymbol{y} - \hat{\boldsymbol{\theta}} = \boldsymbol{y} - P_\Omega \boldsymbol{y} = (I_n - P_\Omega)\boldsymbol{y}$$
$$= (I_n - P_\Omega)(\boldsymbol{\theta} + \boldsymbol{u}) = (I_n - P_\Omega)\boldsymbol{u}$$

$$E\hat{\boldsymbol{u}} = (I_n - P_\Omega)E\boldsymbol{u} = (I_n - P_\Omega)\boldsymbol{0} = \boldsymbol{0}$$

Note that $I_n - P_\Omega$ represents the projection on $\Omega^\perp$ (the orthogonal complement to Ω)

The residual sum of squares

$$\|\hat{\boldsymbol{u}}\|^2 = \|\boldsymbol{y} - \hat{\boldsymbol{\theta}}\|^2 = (\boldsymbol{y} - \hat{\boldsymbol{\theta}})^T(\boldsymbol{y} - \hat{\boldsymbol{\theta}})$$
$$= \boldsymbol{y}^T(I_n - P_\Omega)\boldsymbol{y} = \boldsymbol{u}^T(I_n - P_\Omega)\boldsymbol{u}$$

Assume $\Omega = R(X)$, i.e. $\underset{n \times p}{X}$ is the design matrix, $\underset{\sim}{\theta} = X\underset{\sim}{\beta}$

$X\hat{\underset{\sim}{\beta}} = \hat{\underset{\sim}{\theta}} \Rightarrow X^T X \hat{\underset{\sim}{\beta}} = X^T \hat{\underset{\sim}{\theta}} = X^T P_\Omega y = (P_\Omega X)^T y = X^T y$

hence $\hat{\underset{\sim}{\beta}}$ is a solution to $X^T X \underset{\sim}{\beta} = X^T y$

opposite $X^T X \hat{\underset{\sim}{\beta}} = X^T y \Rightarrow X^T (X\hat{\underset{\sim}{\beta}} - y) = \underset{\sim}{0}$

$\Rightarrow \hat{\underset{\sim}{\beta}}^T X^T (X\hat{\underset{\sim}{\beta}} - y) = 0 \Rightarrow (X\hat{\underset{\sim}{\beta}})^T (X\hat{\underset{\sim}{\beta}} - y) = 0$

$\Rightarrow X\hat{\underset{\sim}{\beta}} \perp y - X\hat{\underset{\sim}{\beta}} \Rightarrow X\hat{\underset{\sim}{\beta}} = P_\Omega y \; (= \hat{\underset{\sim}{\theta}})$, as

$y = X\hat{\underset{\sim}{\beta}} + (y - X\hat{\underset{\sim}{\beta}})$ apparently is an orthogonal decomposition of y, hence unique cf. B1.1

$X^T X \underset{\sim}{\beta} = X^T y$ is the normal equations

$$X^T X \hat{\underset{\sim}{\beta}} = X^T y \Rightarrow \hat{\underset{\sim}{\beta}} = \begin{cases} (X^T X)^{-1} X^T y & \text{for rank } X = p \\ (X^T X)^{-} X^T y & \text{for rank } X < p \end{cases}$$

$\hat{\beta}$ is not uniquely determined, when rank $X < p$
(this can be achieved by introducing "identifiable restrictions")

$X\hat{\beta} = \hat{\theta} = P_\Omega y$ is always uniquely determined, cf B 1.2

$$P_\Omega = \begin{cases} X(X^T X)^{-1} X^T & \text{for rank } X = p \quad \text{cf. B 1.8} \\ X(X^T X)^{-} X^T & \text{for rank } X < p \quad \text{cf. B 1.7} \end{cases}$$

$E\hat{\beta} = (X^T X)^{-1} X^T E y = (X^T X)^{-1} X^T X \underset{\sim}{\beta} = \underset{\sim}{\beta}$ (rank $X = p$),
hence $\hat{\beta}$ is an unbiased estimate

$\underset{\sim}{a}^T \underset{\sim}{\beta}$ is estimable, when one can find a linear unbiased estimate of the form $\underset{\sim}{c}^T y$

$E[\underset{\sim}{c}^T y] = \underset{\sim}{a}^T \underset{\sim}{\beta} \Leftrightarrow \underset{\sim}{c}^T X \underset{\sim}{\beta} = \underset{\sim}{a}^T \underset{\sim}{\beta} \Leftrightarrow \underset{\sim}{a} = X^T \underset{\sim}{c}$

Hence $\underset{\sim}{a}^T$ must be a linear combination of the row vectors in X, i.e. $\underset{\sim}{a} \in R(X^T)$

Theorem $\underset{\sim}{a}^T \underset{\sim}{\beta}$ estimable $\Rightarrow$ $\underset{\sim}{a}^T \hat{\underset{\sim}{\beta}}$ is uniquely determined

Proof $\underset{\sim}{a} = X^T \underset{\sim}{c}$ $\Rightarrow$ $\underset{\sim}{a}^T \underset{\sim}{\beta} = (X^T \underset{\sim}{c})^T \underset{\sim}{\beta} = \underset{\sim}{c}^T X \underset{\sim}{\beta} = \underset{\sim}{c}^T \underset{\sim}{\theta}$

$\Rightarrow$ $\underset{\sim}{a}^T \hat{\underset{\sim}{\beta}} = \underset{\sim}{c}^T \hat{\underset{\sim}{\theta}}$, which is unique

Assume that $Var\ \underset{\sim}{u} = \sigma^2 I_n$

Theorem (Gauss-Markov): $\underset{\sim}{c}^T \hat{\underset{\sim}{\theta}}$ is BLUE for $\underset{\sim}{c}^T \underset{\sim}{\theta}$

Proof $\underset{\sim}{c}^T \hat{\underset{\sim}{\theta}}$ is a linear unbiased estimate for $\underset{\sim}{c}^T \underset{\sim}{\theta}$, as

$\underset{\sim}{c}^T \hat{\underset{\sim}{\theta}} = \underset{\sim}{c}^T P_\Omega \underset{\sim}{y} = (P_\Omega \underset{\sim}{c})^T \underset{\sim}{y}$ and $E[\underset{\sim}{c}^T \hat{\underset{\sim}{\theta}}] = \underset{\sim}{c}^T E \hat{\underset{\sim}{\theta}} = \underset{\sim}{c}^T \underset{\sim}{\theta}$

Let $\underset{\sim}{d}^T \underset{\sim}{y}$ be any linear unbiased estimate for $\underset{\sim}{c}^T \underset{\sim}{\theta}$

$\underset{\sim}{c}^T \underset{\sim}{\theta} = E[\underset{\sim}{d}^T \underset{\sim}{y}] = \underset{\sim}{d}^T E \underset{\sim}{y} = \underset{\sim}{d}^T \underset{\sim}{\theta}$

$\Rightarrow (\underset{\sim}{c} - \underset{\sim}{d})^T \underset{\sim}{\theta} = 0 \Rightarrow \underset{\sim}{c} - \underset{\sim}{d} \in \Omega^\perp \Rightarrow P_\Omega(\underset{\sim}{c} - \underset{\sim}{d}) = \underset{\sim}{0}$

$\Rightarrow P_\Omega \underset{\sim}{c} = P_\Omega \underset{\sim}{d} \Rightarrow \underset{\sim}{c}^T \hat{\underset{\sim}{\theta}} = (P_\Omega \underset{\sim}{d})^T \underset{\sim}{y}$

$Var(\underset{\sim}{c}^T \hat{\underset{\sim}{\theta}}) = (P_\Omega \underset{\sim}{d})^T \sigma^2 I_n P_\Omega \underset{\sim}{d} = \sigma^2 \underset{\sim}{d}^T P_\Omega \underset{\sim}{d}$

$Var(\underset{\sim}{d}^T \underset{\sim}{y}) = \underset{\sim}{d}^T \sigma^2 I_n \underset{\sim}{d} = \sigma^2 \underset{\sim}{d}^T \underset{\sim}{d}$

$Var(\underset{\sim}{d}^T \underset{\sim}{y}) - Var(\underset{\sim}{c}^T \hat{\underset{\sim}{\theta}}) = \sigma^2 \underset{\sim}{d}^T (I_n - P_\Omega) \underset{\sim}{d} \geq 0$

as $I_n - P_\Omega \geq 0$

$\Rightarrow Var(\underset{\sim}{d}^T \underset{\sim}{y}) \geq Var(\underset{\sim}{c}^T \hat{\underset{\sim}{\theta}})$

$Var(\underset{\sim}{d}^T \underset{\sim}{y}) - Var(\underset{\sim}{c}^T \hat{\underset{\sim}{\theta}}) = 0 \Leftrightarrow \underset{\sim}{d}^T (I_n - P_\Omega) \underset{\sim}{d} = 0$

$\Leftrightarrow (I_n - P_\Omega) \underset{\sim}{d} = \underset{\sim}{0} \Leftrightarrow \underset{\sim}{d} = P_\Omega \underset{\sim}{d} \Leftrightarrow \underset{\sim}{d} = P_\Omega \underset{\sim}{c}$

Corollary: $\hat{\theta}_i$ is BLUE for θ_i (choose $\underset{\sim}{c}$ suitable)

Corollary: $\underset{\sim}{a}^T \underset{\sim}{\beta}$ estimable $\Rightarrow$ $\underset{\sim}{a}^T \hat{\underset{\sim}{\beta}}$ is BLUE for $\underset{\sim}{a}^T \underset{\sim}{\beta}$

$(\underset{\sim}{a}^T \underset{\sim}{\beta} = \underset{\sim}{c}^T \underset{\sim}{\theta})$

The residual vector - continued

$$\text{Var}\, \hat{\underset{\sim}{u}} = (I_n - P_\Omega)\sigma^2 I_n (I_n - P_\Omega) = \sigma^2 (I_n - P_\Omega)$$

The residual sum of squares - continued

$$E[\|\hat{\underset{\sim}{u}}\|^2] = E[\underset{\sim}{u}^T (I_n - P_\Omega)\underset{\sim}{u}] = \text{tr}\,(I_n - P_\Omega)\sigma^2 \quad \text{cf. ex. } 1.4$$
$$= (n-p)\sigma^2 \quad \text{cf. A 6.1 and A 1.2 (a)}$$

$\dfrac{\|\hat{\underset{\sim}{u}}\|^2}{n-p} = \dfrac{\|\underset{\sim}{y} - \hat{\underset{\sim}{\theta}}\|^2}{n-p}$ is an unbiased estimate for σ^2

$$\text{Var}\, \hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T \sigma^2 I_n ((X^T X)^{-1} X^T)^T = \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1}$$
$$= \sigma^2 (X^T X)^{-1}$$

Assume $\underset{\sim}{u} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$, rank $X = p$

$\hat{\underset{\sim}{\theta}}$ and $\|\underset{\sim}{y} - \hat{\underset{\sim}{\theta}}\|^2$ independent, cf. A 6.6 modified

$(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta} = P_\Omega \underset{\sim}{u},\ \|\underset{\sim}{y} - \hat{\underset{\sim}{\theta}}\|^2 = \underset{\sim}{u}^T (I_n - P_\Omega)\underset{\sim}{u},\ P_\Omega (I_n - P_\Omega) = 0)$

$\Rightarrow$ $\hat{\underset{\sim}{\beta}}$ and $\|\underset{\sim}{y} - \hat{\underset{\sim}{\theta}}\|^2 = \|\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\|^2$ independent

$\hat{\underset{\sim}{\beta}} \sim N_p(\underset{\sim}{\beta}, \sigma^2 (X^T X)^{-1})$ cf. th. 2.1 (i)

The residual vector - continued

$\hat{\underset{\sim}{u}} \sim N_n(\underset{\sim}{0}, \sigma^2 (I_n - P_\Omega))$ singular $\Rightarrow$ $\hat{u}_i \sim N(0, \sigma^2(1 - p_{ii}))$, $i = 1, \dots, n$

(note the $\hat{u}_i$'s are dep., $\text{Cov}(\hat{u}_i, \hat{u}_j) = -\sigma^2 p_{ij}$, $i \neq j$)

The residual sum of squares - continued

$$\|\hat{\underset{\sim}{u}}\|^2 = \|\underset{\sim}{y} - \hat{\underset{\sim}{\theta}}\|^2 = \underset{\sim}{u}^T (I_n - P_\Omega)\underset{\sim}{u} \sim \sigma^2 \chi^2(n-p) \quad \text{cf. A 6.5}$$

Test of $H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$, A $q \times p$ rank $A = q$, $\underset{\sim}{c}$ q-dim

$A\beta$ estimabel, i.e. $\exists M: A = MX$

$A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c} \sim N_q(\underset{\sim}{0}, \sigma^2 A(X^TX)^{-1}A^T)$ when H_0 is true

$$(A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c})^T (A(X^TX)^{-1}A^T)^{-1} (A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c}) \sim \sigma^2 \chi^2(q)$$

cf. th. 2.1 (vi)

$\underbrace{(A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c})^T (A(X^TX)^{-1}A^T)^{-1} (A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c})}_{\text{fct. of } \hat{\beta}}$ and $\|\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\|^2$ independent

Test statistic

$$\frac{n-p}{q} \frac{(A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c})^T (A(X^TX)^{-1}A^T)^{-1} (A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c})}{\|\underset{\sim}{y} - X\hat{\beta}\|^2} \sim F(q, n-p)$$

when H_0 is true

Multidimensional linear model

$$Y = \Theta + U, \quad EU = 0, \quad \underset{\sim}{\theta}^{(j)} \in \Omega, \quad j = 1, \dots, d$$

$$Y = [\underset{\sim}{y}_1\ \underset{\sim}{y}_2 \cdots \underset{\sim}{y}_n]^T = [\underset{\sim}{y}^{(1)}\ \underset{\sim}{y}^{(2)} \cdots \underset{\sim}{y}^{(d)}] \quad n \times d$$

analogous Θ and U

$\underset{\sim}{y}_i$ is the i'th observation of d variables

$\underset{\sim}{y}^{(j)}$ consist of n observations of the j'th variabel

P_Ω is the matrix of orthogonal projection of R^n on Ω

Let $\hat{\Theta} = P_\Omega Y$ or $\hat{\underset{\sim}{\theta}}^{(j)} = P_\Omega \underset{\sim}{y}^{(j)}$, $j = 1, \dots, d$

$\Rightarrow E\hat{\Theta} = E[P_\Omega Y] = P_\Omega EY = P_\Omega(\Theta + 0) = P_\Omega \Theta = \Theta$

or $E\hat{\underset{\sim}{\theta}}^{(j)} = \underset{\sim}{\theta}^{(j)}$, $j = 1, \dots, d$

Consider a symmetric matrix function $C(\Theta)$

Def. $C(\Theta)$ attains minimum for $\Theta = \hat{\Theta}$

$\Leftrightarrow \quad C(\hat{\Theta}) \leq C(\Theta)$ for all Θ

(hence $C(\Theta) - C(\hat{\Theta}) \geq 0$ for all Θ)

Least squares principle generalized

$$U^T U = (Y-\Theta)^T (Y-\Theta) = (Y-\hat{\Theta}+\hat{\Theta}-\Theta)^T (Y-\hat{\Theta}+\hat{\Theta}-\Theta)$$

$$(Y-\hat{\Theta})^T(\hat{\Theta}-\Theta) = (Y-P_\Omega Y)^T (P_\Omega Y - P_\Omega \Theta)$$
$$= Y^T (I_n - P_\Omega) P_\Omega (Y-\Theta)$$
$$= 0$$

$$U^T U = (Y-\hat{\Theta})^T (Y-\hat{\Theta}) + 0 + 0 + (\hat{\Theta}-\Theta)^T(\hat{\Theta}-\Theta)$$
$$\geq (Y-\hat{\Theta})^T (Y-\hat{\Theta}), \text{ as } (\hat{\Theta}-\Theta)^T(\hat{\Theta}-\Theta) \geq 0 \text{ cf. A 4.4},$$

equality for $\Theta = \hat{\Theta}$ cf. A 4.5

The residual matrix

$$\hat{U} = Y - \hat{\Theta} = Y - P_\Omega Y = (I_n - P_\Omega) Y$$
$$= (I_n - P_\Omega)(\Theta + U) = (I_n - P_\Omega) U$$

$$E\hat{U} = (I_n - P_\Omega) EU = (I_n - P_\Omega) 0 = 0$$

The generalized residual sum of squares

$$E = \hat{U}^T \hat{U} = (Y-\hat{\Theta})^T (Y-\hat{\Theta}) = Y^T (I_n - P_\Omega) Y$$
$$= U^T (I_n - P_\Omega) U$$

Assume $\Omega = R(X)$, X $n \times p$ is the designmatrix

$\underset{\sim}{\theta}^{(j)} = X \underset{\sim}{\beta}^{(j)}$, $j = 1, \ldots, d \quad \Leftrightarrow \quad \Theta = XB$

$B = [\underset{\sim}{\beta}_1 \ \underset{\sim}{\beta}_2 \cdots \underset{\sim}{\beta}_p]^T = [\underset{\sim}{\beta}^{(1)} \ \underset{\sim}{\beta}^{(2)} \cdots \underset{\sim}{\beta}^{(d)}]$ $\quad p \times d$

$$X\hat{B} = \hat{\Theta} \Rightarrow X^TX\hat{B} = X^T\hat{\Theta} = X^TP_\Omega Y = (P_\Omega X)^T Y = X^TY$$

hence $\hat{B}$ is a solution to $X^TXB = X^TY$

$\Leftrightarrow \hat{\underset{\sim}{\beta}}^{(j)}$ is a solution to $X^TX\underset{\sim}{\beta}^{(j)} = X^T\underset{\sim}{y}^{(j)}$, $j = 1, \ldots, d$

$X^TXB = X^TY$ is the normal equations

solution : $\begin{cases} \hat{B} = (X^TX)^{-1}X^TY & \text{for rank } X = p \\ \hat{B} = (X^TX)^{-}X^TY & \text{for rank } X < p \end{cases}$

$\hat{B}$ is not uniquely determined when rank $X < p$

(can be achieved by introducing "identifiable restrictions")

$X\hat{B} = \hat{\Theta} = P_\Omega Y$ is always uniquely determined, cf. B 1.2

$P_\Omega = \begin{cases} X(X^TX)^{-1}X^T & \text{for rank } X = p, \text{ cf. B 1.8} \\ X(X^TX)^{-}X^T & \text{for rank } X < p, \text{ cf. B 1.7} \end{cases}$

$E\hat{B} = (X^TX)^{-1}X^T EY = (X^TX)^{-1}X^TXB = B$ (rank $X = p$),

hence $\hat{B}$ is an unbiased estimate

Note that $\sum_j (\underset{\sim}{a}_j^T)^T \underset{\sim}{\beta}^{(j)} = \text{tr}(AB)$, A $p \times d$

Def. $\text{tr}(AB)$ is estimable when one can find a linear unbiased estimate of the form $\text{tr}(CY)$

$E[\text{tr}(CY)] = \text{tr}(AB) \Leftrightarrow \text{tr}(CXB) = \text{tr}(AB)$

$\Leftrightarrow A = CX \Leftrightarrow A^T = X^TC^T$

$\Leftrightarrow \underset{\sim}{a}_j^T = X^T\underset{\sim}{c}_j^T$, $j = 1, \ldots, d$, hence $\underset{\sim}{a}_j^T$ must be a linear combination of the row vectors in X, $j = 1, \ldots, d$

i.e. $\underset{\sim}{a}_j^T \in \mathcal{R}(X^T)$, $j = 1, \ldots, d$

Note that $(\underset{\sim}{a}_j^T)^T \underset{\sim}{\beta}^{(j)}$, $j = 1, \ldots, d$ estimable

$\iff$ tr(AB) estimable

Theorem: tr(AB) estimable $\Rightarrow$ tr$(A\hat{B})$ is uniquely determined

Proof: $A^T = X^T C^T \Rightarrow A = CX \Rightarrow AB = CXB \Rightarrow AB = C\Theta$

$\Rightarrow A\hat{B} = C\hat{\Theta}$, which is unique

Alternative representations of $Y = XB + U$

$$\underset{\sim}{y}_i = B^T \underset{\sim}{x}_i + \underset{\sim}{u}_i, \quad i = 1, \ldots, n, \quad \text{here } X^T = [\underset{\sim}{x}_1 \ldots \underset{\sim}{x}_n]$$

$$\underset{\sim}{y}^{(j)} = X \underset{\sim}{\beta}^{(j)} + \underset{\sim}{u}^{(j)}, \quad j = 1, \ldots, d$$

Assume that $\underset{\sim}{y}_i$, $i = 1, \ldots, n$, are uncorrelated

and that Var $\underset{\sim}{y}_i = \Sigma$, $i = 1, \ldots, n$

hence $\mathrm{Cov}(\underset{\sim}{y}_i, \underset{\sim}{y}_r) = \delta_{ir} \Sigma$, $i, r = 1, \ldots, n$

$$\mathrm{Cov}(\underset{\sim}{y}^{(j)}, \underset{\sim}{y}^{(k)}) = \mathrm{Cov}(\underset{\sim}{u}^{(j)}, \underset{\sim}{u}^{(k)}) = \sigma_{jk} I_n$$

$$\mathrm{Cov}(\hat{\underset{\sim}{\beta}}^{(j)}, \hat{\underset{\sim}{\beta}}^{(k)}) = (X^T X)^{-1} X^T \mathrm{Cov}(\underset{\sim}{y}^{(j)}, \underset{\sim}{y}^{(k)}) X (X^T X)^{-1}$$

$$= \sigma_{jk} (X^T X)^{-1}$$

$$\mathrm{Var}\, \hat{\underset{\sim}{\beta}}^{(j)} = \sigma_{jj} (X^T X)^{-1}$$

Note that $\sum_j (\underset{\sim}{c}_j^T)^T \underset{\sim}{\theta}^{(j)} = \mathrm{tr}(C\Theta)$

$tr(C\hat{\Theta}) = tr(CP_\Omega Y) = tr((CP_\Omega)Y)$, hence a lin. transf. of Y, and $E[tr(C\hat{\Theta})] = tr(CE\hat{\Theta}) = tr(C\Theta)$

$\Rightarrow$ $tr(C\hat{\Theta})$ is a linear unbiased estimate for $tr(C\Theta)$

Lemma: $Var(tr(AY)) = tr(A^T \Sigma A)$

Proof: $Var(tr(AY)) = Var(\sum_j (\underset{\sim}{a}_{\cdot j}^T)^T y^{(j)})$

$= Cov(\sum_j (\underset{\sim}{a}_{\cdot j}^T)^T y^{(j)}, \sum_k (\underset{\sim}{a}_{\cdot k}^T)^T y^{(k)})$

$= \sum_j \sum_k (\underset{\sim}{a}_{\cdot j}^T)^T Cov(y^{(j)}, y^{(k)}) \underset{\sim}{a}_k^T$

$= \sum_j \sum_k (\underset{\sim}{a}_{\cdot j}^T)^T \sigma_{jk} I_n \underset{\sim}{a}_{\cdot k}^T = \sum_j \sum_k (\underset{\sim}{a}_{\cdot j}^T)^T \underset{\sim}{a}_{\cdot k}^T \sigma_{kj}$

$= \sum_j \sum_k (AA^T)_{jk} \sigma_{kj} = \sum_j (AA^T\Sigma)_{jj} = tr(AA^T\Sigma)$
$= tr(A^T \Sigma A)$

Theorem (Gauss-Markov)

$tr(C\hat{\Theta})$ is BLUE for $tr(C\Theta)$

Proof: Let $tr(DY)$ be any linear unbiased estimate for $tr(C\Theta)$

$tr(C\Theta) = E[tr(DY)] = tr(DEY)$
$= tr(D\Theta)$

$\Rightarrow tr((C-D)\Theta) = 0$ for all Θ

$\Leftrightarrow \sum_j (\underset{\sim}{c}_{\cdot j}^T - \underset{\sim}{d}_{\cdot j}^T)^T \underset{\sim}{\theta}^{(j)} = 0$ for all $\Theta = [\underset{\sim}{\theta}^{(1)} \dots \underset{\sim}{\theta}^{(d)}]$, $\underset{\sim}{\theta}^{(j)} \in \Omega$, $j = 1, \dots, d$

$\Leftrightarrow (\underset{\sim}{c}_{\cdot j}^T - \underset{\sim}{d}_{\cdot j}^T)^T \underset{\sim}{\theta}^{(j)} = 0$ for all $\underset{\sim}{\theta}^{(j)} \in \Omega$, $j = 1, \dots, d$

$\Rightarrow \underset{\sim}{c}_{\cdot j}^T - \underset{\sim}{d}_{\cdot j}^T \in \Omega^\perp$, $j = 1, \dots, d$

$\Rightarrow P_\Omega(\underset{\sim}{c}_{\cdot j}^T - \underset{\sim}{d}_{\cdot j}^T) = \underset{\sim}{0}$, $j = 1, \dots, d$

$\Leftrightarrow P_\Omega(C^T - D^T) = 0 \Leftrightarrow CP_\Omega = DP_\Omega$

$\Rightarrow tr(C\hat{\Theta}) = tr(CP_\Omega Y) = tr(DP_\Omega Y)$

$$Var(tr(C\hat{\Theta})) = tr((DP_\Omega)^T \Sigma D P_\Omega) \quad \text{cf. the lemma}$$

$$= tr(P_\Omega D^T \Sigma D P_\Omega) = tr(P_\Omega D^T \Sigma D)$$

$$Var(tr(DY)) = tr(D^T \Sigma D) \quad \text{cf. the lemma}$$

$$Var(tr(DY)) - Var(tr(C\hat{\Theta}))$$

$$= tr(D^T\Sigma D - P_\Omega D^T \Sigma D) = tr((I_n - P_\Omega) D^T \Sigma D)$$

$$= tr((I_n - P_\Omega) D^T \Sigma D (I_n - P_\Omega))$$

$$= tr((D(I_n - P_\Omega))^T \Sigma D(I_n - P_\Omega)) \geq 0 \quad \text{cf. A 4.4 and A 4.2}$$

$$\Rightarrow Var(tr(DY)) \geq Var(tr(C\hat{\Theta}))$$

$$Var(tr(DY)) - Var(tr(C\hat{\Theta})) = 0$$

$$\Leftrightarrow tr((D(I_n - P_\Omega))^T \Sigma D(I_n - P_\Omega)) = 0$$

$$\Leftrightarrow (D(I_n - P_\Omega))^T \Sigma D(I_n - P_\Omega) = 0$$

$$\Leftrightarrow D(I_n - P_\Omega) = 0 \quad \text{cf. A 4.5}$$

$$\Leftrightarrow D = DP_\Omega \Leftrightarrow D = CP_\Omega$$

Corollary: $\hat{\theta}_{ij}$ is BLUE for Θ_{ij} (choose C suitable)

Corollary: $tr(AB)$ estimable
$\Rightarrow tr(A\hat{B})$ is BLUE for $tr(AB)$
($AB = C\Theta$ cf. page 8)

Theorem: $\frac{E}{n-r}$ is an unbiased estimate for Σ, $r = \dim \Omega$

Proof: $E[E] = E[U^T(I_n - P_\Omega)U] = (tr(I_n - P_\Omega))\Sigma + O^T(I_n - P_\Omega)O$ cf. corollary to lemma 1.1

$= (n-r)\Sigma$

$E > 0$ a.s. for $n - r \geq d$ cf. A 5.13

$\frac{E}{n-r}$ is usually named S, i.e. S is an unbiased estimate for Σ