XIV. Simultaneous Inference in a Linear Model

ASSUMPTIONS in the whole chapter:

$(Y_i, X_i^T)^T$, $i = 1, \dots, n$ satisfy

$$Y \mid X \sim N_n(X\beta, \sigma^2 I_n), \quad \text{rank}(X_{n \times k}) = k < n$$

$$L = \begin{pmatrix} l_1^T \\ \vdots \\ l_m^T \end{pmatrix}, \quad l_j \neq 0_k \qquad \theta := L\beta = (l_1^T\beta, \dots, l_m^T\beta)^T = (\theta_1, \dots, \theta_m)^T$$

INTEREST:

(i) confidence region for θ

(ii) test of $H_0: \theta = \theta^0$ $(\theta^0 \in \mathbb{R}^m$ given$)$

14.1 Basic simultaneous inference

$\mathbb{L}_{m \times k}$, $m \le k$, rows of $\mathbb{L}$: $l_1^T, \dots, l_m^T \in \mathbb{R}^k$

→ linearly independent ($\neq 0_k$)

$$\hat{\theta} := \mathbb{L}\hat{\beta} = (l_1^T\hat{\beta}, \dots, l_m^T\hat{\beta})^T, \quad \hat{\beta} = (X^TX)^{-1}X^TY$$

→ BLUE of θ

$$MS_e := \frac{1}{n-k}(Y - X\hat{\beta})^T(Y - X\hat{\beta})$$

→ unbiased estim of σ^2

$$CR(\alpha) = \{\theta \in \mathbb{R}^m : (\theta - \hat{\theta})^T\{MS_e \mathbb{L}(X^TX)^{-1}\mathbb{L}^T\}^{-1}(\theta - \hat{\theta}) < m\, F_{m,n-k}(1-\alpha)\}$$

confidence region with a coverage of $1-\alpha$

$\equiv$ conf ellipsoid with a coverage of $1-\alpha$

Test of $H_0: \theta = \theta^0$ ($\in \mathbb{R}^m$)

→ test based on

$$Q_0 = \frac{1}{m}(\hat{\theta} - \theta^0)^T\{MS_e \mathbb{L}(X^TX)^{-1}\mathbb{L}^T\}^{-1}(\hat{\theta} - \theta^0)$$

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

← critical region

Remark: Confidence region and test are dual

$H_0: \theta = \theta^0$ rejected on a level of α

$$\Longleftrightarrow \quad \theta^0 \notin CR(\alpha)$$

WILL FOLLOW (added value):

- rows of $\mathbb{L}$ possibly linearly dependent
- $m > k$
- conf. region  and not 

= product of intervals

14.2 Multiple comparison procedures

Interest in testing $H_0: \Theta = \Theta^0$

$\equiv H_0: \theta_1 = \theta_1^0 \ \& \dots \& \ \theta_m = \theta_m^0$

14.2.1 Multiple testing

Def 14.1 Multiple testing problem, elementary null hypotheses, global null hypothesis

A testing problem with the null hypothesis

$H_0: \theta_1 = \theta_1^0 \ \& \dots \& \ \theta_m = \theta_m^0$

is called the multiple testing problem with the m elementary hypotheses

$H_1: \theta_1 = \theta_1^0, \dots, H_m: \theta_m = \theta_m^0$.

Hypothesis H_0 is called in this context also as a global null hypothesis.

Notation: $H_0 \equiv H_1 \ \& \dots \& \ H_m$

$\equiv H_1, \dots, H_m$

$\equiv \bigcap_{j=1}^{m} H_j$

Alternative to the global null hypothesis:

$H_0^C: \theta_1 \neq \theta_1^0$ OR $\dots$ OR $\theta_m \neq \theta_m^0$

$\equiv H_1^C$ OR $\dots$ OR H_m^C, $H_j^C: \theta_j \neq \theta_j^0$

$\equiv \bigcup_{j=1}^{m} H_j^C$

Sometimes also double subscript will be used to index the elementary hyp.

Example: One-way classification

- $Z \in \mathcal{Z} = \{1, \dots, G\}$ categorical covariate
- $\mathbb{X} = n \times G$ matrix derived from $\mathbb{C} = \begin{pmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_G^T \end{pmatrix}$

(pseudo) contrast matrix

- $m_g = \mathbb{E}(Y \mid Z = g) = \beta_0 + \mathbf{c}_g^T \boldsymbol{\beta}^Z, \quad g = 1, \dots, G$
- $H_0: \ m_1 = \dots = m_G$

$$\equiv H_{1,2}: \underbrace{m_1 - m_2}_{\theta_{1,2}} = 0, \quad \dots, \quad H_{G-1,G}: \underbrace{m_{G-1} - m_G}_{\theta_{G-1,G}} = 0$$

$$= (\mathbf{c}_1^T - \mathbf{c}_2^T)\boldsymbol{\beta}^Z \qquad\qquad = (\mathbf{c}_{G-1}^T - \mathbf{c}_G^T)\boldsymbol{\beta}^Z$$

$$\theta_{g,h} := m_g - m_h = (\mathbf{c}_g - \mathbf{c}_h)^T \boldsymbol{\beta}^Z,$$

$$g = 1, \dots, G-1$$
$$h = g+1, \dots, G$$

In a linear model parameterization:

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_{1,2} \\ \vdots \\ \theta_{G-1,G} \end{pmatrix} = \mathbb{L}\boldsymbol{\beta}, \quad \mathbb{L} = \underbrace{\begin{pmatrix} 0 & \mathbf{c}_1^T - \mathbf{c}_2^T \\ \vdots & \vdots \\ 0 & \mathbf{c}_{G-1}^T - \mathbf{c}_G^T \end{pmatrix}}_{G} \Bigg\} \binom{G}{2}$$

$$\mathbb{L} = \left.\underbrace{\begin{pmatrix} 0 & C_1^T - C_2^T \\ \vdots & \vdots \\ 0 & C_{G-1}^T - C_G^T \end{pmatrix}}_{G}\right\} \binom{G}{2}$$

$$\mathbb{C} = \left.\underbrace{\begin{pmatrix} C_1^T \\ \vdots \\ C_G^T \end{pmatrix}}_{G-1}\right\} G$$

$\operatorname{rank}(\mathbb{C}) = G-1$

$\Rightarrow \operatorname{rank}(\mathbb{L}) = G-1$

- $G \geq 3 \Rightarrow$ rows in $\mathbb{L}$ are linearly dependent
- $G \geq 4 \Rightarrow \binom{G}{2} > \operatorname{rank}(\mathbb{X}) = G$

NOT POSSIBLE:

- elliptical confidence region for

$$\theta = (\underbrace{m_1 - m_2}_{\theta_{1,2}}, \ldots, \underbrace{m_{G-1} - m_G}_{\theta_{G-1,G}})^T$$

- test of $\theta = \theta^0$ (directly by using the F-statistic Q_0 applied to $\hat{\theta}$)

$$Q_0 = \frac{1}{m}(\hat{\theta} - \theta^0)^T (\ldots)^{-1} (\hat{\theta} - \theta^0)$$

$\sim F_{m, n-G}$

$m = \binom{G}{2}$

14.2.2 Simultaneous confidence intervals

$D\ (= D(Y, X))$: random vector, its distribution depends on a (vector) parameter

$\theta = (\theta_1, \dots, \theta_m)^T \in \Theta_1 \times \dots \times \Theta_m = \Theta \subseteq \mathbb{R}^m$

Def. 14.2 Simultaneous confidence intervals

(Random) intervals (θ_j^L, θ_j^U), $j = 1, \dots, m$, where $\theta_j^L = \theta_j^L(D)$, $\theta_j^U = \theta_j^U(D)$, $j = 1, \dots, m$, are called simultaneous confidence intervals for parameter θ with a coverage of $1-\alpha$ if for any $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \Theta$,

$$P\left((\theta_1^L, \theta_1^U) \times \dots \times (\theta_m^L, \theta_m^U) \ni \theta^0 ;\ \theta = \theta^0 \right) \geq 1-\alpha .$$

$$P\left(\forall j = 1, \dots, m \ \ (\theta_j^L, \theta_j^U) \ni \theta_j^0 ;\ \theta = \theta^0 \right) \geq 1-\alpha$$

$$= P\left(\bigcap_{j=1}^m \left[(\theta_j^L, \theta_j^U) \ni \theta_j^0 \right] ;\ \theta = \theta^0 \right) \leq P\left((\theta_j^L, \theta_j^U) \ni \theta_j^0 ;\ \theta = \theta^0 \right) \quad \text{for any } j$$

(a) simultaneous confidence intervals form a classical confidence region for θ

(b) $\forall \theta^0$ $\forall j = 1, \dots, m$ $P\left((\theta_j^L, \theta_j^U) \ni \theta_j^0 ;\ \theta = \theta^0 \right) \geq 1-\alpha$

Example: Bonferroni simultaneous confidence intervals

Let (θ_j^L, θ_j^U) be a classical CI for θ_j with a coverage $1-\frac{\alpha}{m}$.

$$\rightarrow \forall j=1,\dots,m \quad \forall \theta_j^0 \in \Theta_j: \quad P\left((\theta_j^L, \theta_j^U) \ni \theta_j^0; \theta_j = \theta_j^0\right) \geq 1-\frac{\alpha}{m}$$

$$\forall j=1,\dots,m \quad \forall \theta_j^0 \in \Theta_j: \quad P\left((\theta_j^L, \theta_j^U) \not\ni \theta_j^0; \theta_j = \theta_j^0\right) \leq \frac{\alpha}{m}$$

$\Rightarrow \forall \theta^0 \in \Theta$:

$$P\left(\exists j=1,\dots,m \quad (\theta_j^L, \theta_j^U) \not\ni \theta_j^0; \theta = \theta^0\right)$$

$$\leq \sum_{j=1}^{m} P\left((\theta_j^L, \theta_j^U) \not\ni \theta_j^0; \theta = \theta^0\right) \leq \sum_{j=1}^{m} \frac{\alpha}{m} = \alpha.$$

$\Rightarrow \forall \theta^0 \in \Theta$:

$$P\left(\forall j=1,\dots,m: \quad (\theta_j^L, \theta_j^U) \ni \theta_j^0; \theta = \theta^0\right) \geq 1-\alpha$$

Hence: (θ_j^L, θ_j^U), $j=1,\dots,m$ are simultaneous confidence intervals for θ.

PROBLEM:

$$P\left(\forall j=1,\dots,m: \quad (\theta_j^L, \theta_j^U) \ni \theta_j^0; \theta = \theta^0\right) \overset{\text{often}}{>>} 1-\alpha$$

14.2.3 Multiple comparison procedure, P-values adjusted for multiple comparison

Let for each $0<\alpha<1$ a procedure be given to construct the simultaneous confidence intervals $(\theta_j^L(\alpha), \theta_j^U(\alpha))$, $j=1,..,m$, for parameter θ with a coverage $1-\alpha$.
Let for each $j=1,..,m$, the procedure creates intervals satisfying a monotonicity condition

$$1-\alpha_1 < 1-\alpha_2 \Rightarrow (\theta_j^L(\alpha_1), \theta_j^U(\alpha_1)) \subseteq (\theta_j^L(\alpha_2), \theta_j^U(\alpha_2)).$$

---

being motivated by duality between tests and confidence intervals

Def 14.3 Multiple comparison procedure

→ see slide

NOTE: H_0 rejected $\Leftrightarrow$ $\exists j$ H_j rejected

Construction guarantees ($\forall \theta^0 \in \Theta$)

$$P(H_0 \text{ rejected};\ \theta=\theta^0) \leq \alpha$$

$$\forall j \quad P(H_j \text{ rejected};\ \theta_j=\theta_j^0) \leq \alpha$$

$\Rightarrow$ The procedure controls type I error rate for (i) test of (global) H_0

(ii) test of each (elementary) H_j $\forall j=1,..,m$

Remind (a possible) definition of a P-value

$p = \inf\{\alpha :$ H_0 rejected by a given procedure on a signif. level of $\alpha \}$

Def 12.4 P-values adjusted for multiple comparison

P-values adjusted for multiple comparison for a multiple testing problem with the elementary null hypotheses $H_j: \theta_j = \theta_j^0$, $j = 1,..,m$, based on a given procedure for construction of simultaneous confidence intervals for parameter θ are values $p_1^{adj}, \dots, p_m^{adj}$ defined as

$$p_j^{adj} = \inf\{\alpha : (\theta_j^L(\alpha), \theta_j^U(\alpha)) \not\ni \theta_j^0\}, \quad j = 1,..,m.$$

- For given α, $0 < \alpha < 1$
- MCP rejects $H_j: \theta_j = \theta_j^0$ $(j = 1,..,m)$ $\Leftrightarrow$ $p_j^{adj} \leq \alpha$.
- MCP rejects $H_0: \theta = \theta^0$

$\Leftrightarrow$ at least one elementary hypothesis rejected

$\Leftrightarrow$ $\exists j$ $p_j^{adj} \leq \alpha$

$\Leftrightarrow$ $\min_j p_j^{adj} \leq \alpha$

That is, P-value of the test of a (global) H_0:

$$p^{adj} := \min\{p_1^{adj}, \dots, p_m^{adj}\}.$$

Example: Bonferroni MCP, Bonferroni adjusted P-values

For given $0<\alpha<1$:

- $\forall j=1,\dots,m$: $(\theta_j^L(\alpha),\ \theta_j^U(\alpha))$: a classical CI for θ_j with a coverage of $1-\frac{\alpha}{m}$,

  i.e. $\forall j=1,\dots,m\ \forall \theta_j^0 \in \Theta_j\ P\big((\theta_j^L(\alpha), \theta_j^U(\alpha)) \ni \theta_j^0;\ \theta_j=\theta_j^0\big) \geq 1-\frac{\alpha}{m}$

  $\Rightarrow \forall \theta^0 \in \Theta\ P\big(\forall j=1,\dots,m,\ (\theta_j^L(\alpha), \theta_j^U(\alpha)) \ni \theta_j^0;\ \theta=\theta^0\big) \geq 1-\alpha$

  $\equiv$ Bonferroni simultaneous conf. intervals for Θ with a coverage of $1-\alpha$.

- Let p_j^{uni} = p-value related to the (single) test of $H_j: \theta_j=\theta_j^0$, dual to CI $(\theta_j^L(\alpha), \theta_j^U(\alpha))$

  i.e. $p_j^{uni} = \inf\{\frac{\alpha}{m} : (\theta_j^L(\alpha), \theta_j^U(\alpha)) \not\ni \theta_j^0\}$

$\Rightarrow \min\{\underbrace{m\,p_j^{uni}}_{\text{can be } >1}, 1\} = \inf\{\alpha : (\theta_j^L(\alpha), \theta_j^U(\alpha)) \not\ni \theta_j^0\}$

$\Rightarrow p_j^B = \min\{m\,p_j^{uni}, 1\}$

$\equiv$ Bonferroni adjusted P-values

Bonferroni procedure

- rather universal
- BUT conservative (is "afraid" to reject H_0)

  – to reject (a global) $H_0: \Theta = \Theta^0$ on a level of α, there must exist $p_j^{uni} \leq \frac{\alpha}{m}$

  small for m large

- global H_0 rejected on a level of α

  $\Leftrightarrow$ $\exists j$: H_j rejected on a (small) level $\frac{\alpha}{m}$

14.2.4 Bonferroni simultaneous inference in a normal linear model

Consider a full-rank normal linear model
$Y|X \sim N_n(X\beta, \sigma^2 I_n)$, $\text{rank}(X_{n\times k}) = k < n$.

- $\Theta := L\beta = (l_1^T\beta, \dots, l_m^T\beta)^T = (\theta_1, \dots, \theta_m)^T$

$\Rightarrow$ Θ LSE: $\hat{\Theta} = L\hat{\beta} = (l_1^T\hat{\beta}, \dots, l_m^T\hat{\beta})^T$
$= (\hat{\theta}_1, \dots, \hat{\theta}_m)^T$

MS_e = residual mean square

Under $H_j: \theta_j = \theta_j^0$: $T_j(\theta_j^0) = T_{j,0} = \dfrac{\hat{\theta}_j - \theta_j^0}{\sqrt{MS_e\, l_j^T (X^TX)^{-1} l_j}} \sim t_{n-k}$

$\Rightarrow$ Bonferroni simultaneous confidence intervals (coverage $1-\alpha$):

$$\begin{matrix}\theta_j^L(\alpha) \\ \theta_j^U(\alpha)\end{matrix} = l_j^T\hat{\beta} \mp \sqrt{MS_e\, l_j^T(X^TX)^{-1}l_j}\; t_{n-k}\left(1 - \frac{\alpha}{2m}\right)$$

$\rightarrow$ are dual to tests of $H_j: \theta_j = \theta_j^0$ based on statistics T_{j0} and critical regions $C_j = \left(-\infty, -t_{n-k}\left(1-\frac{\alpha}{2m}\right)\right)$
$\cup \left(t_{n-k}\left(1-\frac{\alpha}{2m}\right), \infty\right)$

(test on a level of $\frac{\alpha}{m}$)

That is, $\theta_j^0 \notin (\theta_j^L(\alpha), \theta_j^u(\alpha))$

$$\Longleftrightarrow \quad |T_{j,0}| > t_{n-k}\left(1 - \frac{\alpha}{2m}\right)$$

$\hat{=}$ level of a single test

Hence $\quad p_j^{uni} = 2 \cdot CDF_{t,n-k}(-|t_{j,0}|)$

$t_{j,0}$ = value of $T_{j,0}$ attained with given data

$\Rightarrow$ Bonferroni adjusted P-values for $H_j: \theta_j = \theta_j^0$, $j = 1, \dots, m$

$$p_j^B = \min\{2m\, CDF_{t,n-k}(-|t_{j,0}|), 1\}, \quad j = 1, \dots, m$$