Heteroskedastická jednofaktorová ANOVA - intervaly pre variančné komponenty

Barbora Arendacká
Ústav merania SAV, Bratislava

ROBUST 2010, 31.1.2010 - 5.2.2010, Králíky

Model

model:

$$Y_{ij} = \mu + A_i + \epsilon_{ij}$$
$$i = 1, \ldots, k, \quad j = 1, \ldots, n_i \quad (k > 1, n_i > 1)$$

$$A_i \sim N(0, \sigma_A^2),\, \epsilon_{ij} \sim N(0, \sigma_i^2),\, \sigma_A^2 \geq 0,\, \sigma_i^2 > 0$$
$A_1, \ldots, A_k, \epsilon_{11}, \ldots, \epsilon_{kn_k}$ sú navzájom nezávislé

maticový zápis:

$$Y = 1_n\mu + ZA + \epsilon$$

$$A \sim N(0, \sigma_A^2 I_k),\, \epsilon \sim N(0, diag\{\sigma_1^2 I_{n_1}, \ldots, \sigma_k^2 I_{n_k}\})$$

výskyt: medzilaboratórne porovnávania: k laboratórií, n_i meraní v i-tom laboratóriu
variabilita medzi laboratóriami $\longrightarrow \sigma_A^2$, variabilita v i-tom laboratóriu $\longrightarrow \sigma_i^2$

úloha: nájsť intervaly spoľahlivosti pre σ_A^2

Prístup v literatúre

inferencia o $\sigma_i^2, i = 1, \ldots, k$: $\quad S_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \sim \sigma_i^2 \chi^2_{n_i - 1}$

inferencia o σ_A^2:

1. vážená suma štvorcov

$$WS = \sum_{i=1}^{k} \frac{1}{\sigma_A^2 + \sigma_i^2/n_i} \left(\bar{Y}_i - \frac{\sum_{j=1}^{k} \bar{Y}_j/(\sigma_A^2 + \sigma_j^2/n_j)}{\sum_{j=1}^{k} 1/(\sigma_A^2 + \sigma_j^2/n_j)} \right)^2 \sim \chi^2_{k-1}$$

Wimmer, Witkovský (2003), Hartung, Knapp (2005)

2. nevážená suma štvorcov

$$S_0 = H_2^T Q_1 Y \sim N(0, \underbrace{\sigma_A^2 I_{k-1} + H_2^T diag(\sigma_1^2/n_1, \ldots, \sigma_k^2/n_k) H_2}_{\Sigma})$$

$$S_0^T S_0 = S_A^2 = \sum_{i=1}^{k-1} (\bar{Y}_i - \bar{Y})^2, \quad \bar{Y} = \sum_{i=1}^{k} \bar{Y}_i / k$$

$$S_0^T S_0 \sim w_0^T \Sigma w_0, \quad w_0 \sim N(0, I_{k-1})$$

$\rightarrow$ zovšeobecnené intervaly využívajúce, že trieda rozdelení S_A^2 je stochasticky rastúca vzhľadom na σ_A^2 (Li(2007))

Homoskedastický prípad, $\sigma_1^2 = \ldots = \sigma_k^2 = \sigma^2$

1. invariancia na posun v strednej hodnote: $\boldsymbol{Y} \to \widetilde{\boldsymbol{Y}}$

$$\widetilde{\boldsymbol{Y}} = \boldsymbol{B}^T\boldsymbol{Y} \sim N\left(\boldsymbol{0}, \sigma_A^2 \boldsymbol{B}^T\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{B} + \sigma^2 \boldsymbol{I}_{n-1}\right)$$

$\boldsymbol{B}^T\boldsymbol{B} = \boldsymbol{I}_{n-1}$, $\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n^T/n$

2. minimálna postačujúca štatistika: $(\widetilde{\boldsymbol{Y}}^T\boldsymbol{E}_1\widetilde{\boldsymbol{Y}}, \ldots, \widetilde{\boldsymbol{Y}}^T\boldsymbol{E}_{d-1}\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Y}}^T\boldsymbol{E}_d\widetilde{\boldsymbol{Y}})^T$

   ak $\boldsymbol{V} = \boldsymbol{B}^T\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{B}$, tak $\boxed{\boldsymbol{P}_V = \sum_{i=1}^{d-1}\boldsymbol{E}_i, \boldsymbol{P}_V + \boldsymbol{E}_d = \boldsymbol{I}_{n-1}}$

$$\boldsymbol{V} = \boldsymbol{B}^T\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{B} = \sum_{i=1}^{d-1}\lambda_i\boldsymbol{E}_i \text{ (spektr. rozklad)}, \sum_{i=1}^{d}\boldsymbol{E}_i = \boldsymbol{I}_{n-1}$$

$$\widetilde{\boldsymbol{Y}}^T\boldsymbol{E}_i\widetilde{\boldsymbol{Y}} \sim (\lambda_i\sigma_A^2 + \sigma^2)\chi^2_{\nu_i}, i = 1, \ldots, d-1, \widetilde{\boldsymbol{Y}}^T\boldsymbol{E}_d\widetilde{\boldsymbol{Y}} \sim \sigma^2\chi^2_{\nu_d}$$

navzájom nezávislé

3. zovšeobecnené pivoty pre σ_A^2: využijeme $\widetilde{\boldsymbol{Y}}^T\boldsymbol{E}_d\widetilde{\boldsymbol{Y}} \sim \sigma^2\chi^2_{\nu_d}$ a

Typ 1: $\sum_{i=1}^{d-1} c_i\widetilde{\boldsymbol{Y}}^T\boldsymbol{E}_i\widetilde{\boldsymbol{Y}} \sim \sum_{i=1}^{d-1} c_i(\lambda_i\sigma_A^2 + \sigma^2)\chi^2_{\nu_i}$, $c_i > 0$

pozn. $c_i = 1/\lambda_i \to S_A^2$

Typ2: $\sum_{i=1}^{d-1} c_i\widetilde{\boldsymbol{Y}}^T\boldsymbol{E}_i\widetilde{\boldsymbol{Y}}/(\lambda_i\sigma_A^2 + \sigma^2) \sim \sum_{i=1}^{d-1} c_i\chi^2_{\nu_i}$, $c_i > 0$

Analogický postup pre heteroskedastický prípad

Značenie:

$$ZZ^T = \begin{pmatrix} \mathbf{1}_{n_1}\mathbf{1}_{n_1}^T & & & \\ & \mathbf{1}_{n_2}\mathbf{1}_{n_2}^T & & \\ & & \ddots & \\ & & & \mathbf{1}_{n_k}\mathbf{1}_{n_k}^T \end{pmatrix}$$

Uvažujme matice $n \times n$ rozdelené do $k \times k$ blokov podľa ZZ^T. $(H)_{ij}$ označíme maticu, kde H $(n_i \times n_j)$ je na mieste i, j-teho bloku a ostatné bloky sú nulové.

$$P_{(i)} = (\mathbf{1}_{n_i}\mathbf{1}_{n_i}^T / n_i)_{ii}$$
$$M_{(i)} = (I_{n_i})_{ii} - P_{(i)}$$
$$J_{(ij)} = (\mathbf{1}_{n_i}\mathbf{1}_{n_j}^T)_{ij} + (\mathbf{1}_{n_j}\mathbf{1}_{n_i}^T)_{ji}$$
$$i = 1, \dots, k, \; j = 1, \dots, k, \; i < j$$

Analogický postup pre heteroskedastický prípad

1. invariancia na posun v strednej hodnote: $\boldsymbol{Y} \to \widetilde{\boldsymbol{Y}}$

$$\widetilde{Y} = B^T Y \sim N\left(0, \sigma_A^2 B^T Z Z^T B + \sigma_1^2 B^T (I_{n_1})_{11} B + \ldots + \sigma_k^2 B^T (I_{n_k})_{kk} B\right)$$

$$B^T B = I_{n-1}, \; B B^T = I_n - 1_n 1_n^T / n$$

2. minimálna postačujúca štatistika:

$$(\widetilde{Y}^T B^T J_{(12)} B \widetilde{Y}, \ldots, \widetilde{Y}^T B^T J_{(1k)} B \widetilde{Y}, \widetilde{Y}^T B^T J_{(23)} B \widetilde{Y}, \ldots,$$
$$\widetilde{Y}^T B^T J_{((k-1)k)} B \widetilde{Y}, \widetilde{Y}^T B^T M_{(1)} B \widetilde{Y}, \ldots, \widetilde{Y}^T B^T M_{(k)} B \widetilde{Y})^T$$

$S_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = \widetilde{Y}^T B^T M_{(i)} B \widetilde{Y} \sim \sigma_i^2 \chi^2_{n_i - 1}$, navzájom nezávislé

S_i^2 nezávislé na $\widetilde{Y}^T B^T J_{(kl)} B \widetilde{Y}$, $\quad \widetilde{Y}^T B^T J_{(kl)} B \widetilde{Y}$ NIE sú navzájom nezávislé

$P_V = -\sum_i \sum_{j; i<j} B^T J_{(ij)} B$
$P_V + B^T M_{(1)} B + \ldots + B^T M_{(k)} B = I_{n-1}$

$V = B^T Z Z^T B$, $V = \sum_{i=}^{d-1} \lambda_i E_i$, E_i =lin. kombinácia $B^T J_{(ij)} B$
$E_d = B^T M_{(1)} B + \ldots + B^T M_{(k)} B$

Analogický postup pre heteroskedastický prípad

1. invariancia na posun v strednej hodnote: $\boldsymbol{Y} \to \widetilde{\boldsymbol{Y}}$

$$\widetilde{Y} = B^T Y \sim N\left(0, \sigma_A^2 B^T Z Z^T B + \sigma_1^2 B^T (I_{n_1})_{11} B + \ldots + \sigma_k^2 B^T (I_{n_k})_{kk} B\right)$$

$$B^T B = I_{n-1},\ B B^T = I_n - 1_n 1_n^T / n$$

2. minimálna postačujúca štatistika:

$$(\widetilde{Y}^T B^T J_{(12)} B \widetilde{Y}, \ldots, \widetilde{Y}^T B^T J_{(1k)} B \widetilde{Y}, \widetilde{Y}^T B^T J_{(23)} B \widetilde{Y}, \ldots,$$
$$\widetilde{Y}^T B^T J_{((k-1)k)} B \widetilde{Y}, \widetilde{Y}^T B^T M_{(1)} B \widetilde{Y}, \ldots, \widetilde{Y}^T B^T M_{(k)} B \widetilde{Y})^T$$

$S_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = \widetilde{Y}^T B^T M_{(i)} B \widetilde{Y} \sim \sigma_i^2 \chi^2_{n_i - 1}$, navzájom nezávislé

S_i^2 nezávislé na $\widetilde{Y}^T B^T J_{(kl)} B \widetilde{Y}$, $\quad \widetilde{Y}^T B^T J_{(kl)} B \widetilde{Y}$ NIE sú navzájom nezávislé

$$P_V = -\sum_i \sum_{j; i<j} B^T J_{(ij)} B$$
$$P_V + B^T M_{(1)} B + \ldots + B^T M_{(k)} B = I_{n-1}$$

$V = B^T Z Z^T B$, $V = \sum_{i=}^{d-1} \lambda_i E_i$, E_i =lin. kombinácia $B^T J_{(ij)} B$

$$E_d = B^T M_{(1)} B + \ldots + B^T M_{(k)} B$$

Prvé pozorovanie:

$\longrightarrow$ doteraz navrhnuté zovšeobecnené pivoty sú založené na min. postač. štatistike

$$S_1^2, \ldots, S_k^2$$
$$S_A^2 = \sum_{i=1}^{k} (\bar{Y}_i - \bar{Y})^2 = \sum_{i=1}^{d-1} \frac{1}{\lambda_i} \widetilde{Y}^T E_i \widetilde{Y} \qquad (E_i = \text{lin. komb. } B^T J_{(ij)} B)$$
$$WS = \sum_{i=1}^{k} w_i (\bar{Y}_i - \bar{Y}_w)^2 = Y^T A Y = Y^T B B^T A B B^T Y = \widetilde{Y}^T B^T A B \widetilde{Y}$$

Druhé pozorovanie:

môžme konštruovať rovnaké typy pivotov ako v homoskedastickom prípade

$$\sum_{i=1}^{d-1} c_i \widetilde{Y}^T E_i \widetilde{Y} \rightarrow \sum_{i=1}^{d-1} c_i \widetilde{Y}^T E_i \widetilde{Y} = \widetilde{Y}^T B_V C B_V^T \widetilde{Y}$$

$$\sum_{i=1}^{d-1} c_i \frac{\widetilde{Y}^T E_i \widetilde{Y}}{\lambda_i \sigma_A^2 + \sigma^2} \overset{c_i=1}{\longrightarrow} \sum_{i=1}^{d-1} \frac{\widetilde{Y}^T E_i \widetilde{Y}}{\lambda_i \sigma_A^2 + \sigma^2} = \widetilde{Y}^T B_V (B_V^T \Sigma B_V)^{-1} B_V^T \widetilde{Y}$$

$B_V = (vl.vektoryE_1, \ldots, vl.vektoryE_{d-1})$, $B_V B_V^T = P_V$, $C = diag(c_1 1_{\nu_1}, \ldots, c_{d-1} 1_{\nu_{d-1}})$, $c_i > 0$

Analógia - pokračujeme...

Prvé pozorovanie:

$\longrightarrow$ doteraz navrhnuté zovšeobecnené pivoty sú založené na min. postač. štatistike

$S_1^2, \ldots, S_k^2$

$S_A^2 = \sum_{i=1}^{k} (\bar{Y}_i - \bar{Y})^2 = \sum_{i=1}^{d-1} \frac{1}{\lambda_i} \widetilde{Y}^T E_i \widetilde{Y}$ (E_i =lin. komb. $B^T J_{(ij)} B$)

$WS = \sum_{i=1}^{k} w_i (\bar{Y}_i - \bar{Y}_w)^2 = Y^T A Y = Y^T B B^T A B B^T Y = \widetilde{Y}^T B^T A B \widetilde{Y}$

Druhé pozorovanie:

môžme konštruovať rovnaké typy pivotov ako v homoskedastickom prípade

$$\sum_{i=1}^{d-1} c_i \widetilde{Y}^T E_i \widetilde{Y} \longrightarrow \sum_{i=1}^{d-1} c_i \widetilde{Y}^T E_i \widetilde{Y} = \widetilde{Y}^T B_V C B_V^T \widetilde{Y}$$

$$\sum_{i=1}^{d-1} c_i \frac{\widetilde{Y}^T E_i \widetilde{Y}}{\lambda_i \sigma_A^2 + \sigma^2} \stackrel{c_i=1}{\longrightarrow} \sum_{i=1}^{d-1} \frac{\widetilde{Y}^T E_i \widetilde{Y}}{\lambda_i \sigma_A^2 + \sigma^2} = \widetilde{Y}^T B_V (B_V^T \Sigma B_V)^{-1} B_V^T \widetilde{Y}$$

$B_V = (vl.vektoryE_1, \ldots, vl.vektoryE_{d-1})$, $B_V B_V^T = P_V$, $C = diag(c_1 1_{\nu_1}, \ldots, c_{d-1} 1_{\nu_{d-1}})$, $c_i > 0$

Analógia - pokračujeme...

Typ 1, $c_i > 0$

$$\sum_{i=1}^{d-1} c_i \widetilde{Y}^T E_i \widetilde{Y} = \widetilde{Y}^T B_V C B_V^T \widetilde{Y} \sim w_0^T C^{1/2} B_V^T \Sigma B_V C^{1/2} w_0$$

$w_0 \sim N(0, I_{k-1})$, $C^{1/2} B_V^T \Sigma B_V C^{1/2} = \sigma_A^2 *$ diag. matica$+ \sum_{i=1}^{k} \sigma_i^2 *$príslušná matica

trieda rozdelení stochasticky rastúca v σ_A^2

• $c_i = 1/\lambda_i \longrightarrow S_A^2$ (Li (2007)), • $c_i = 1 \longrightarrow S_I^2$

Typ 2, $c_i = 1$

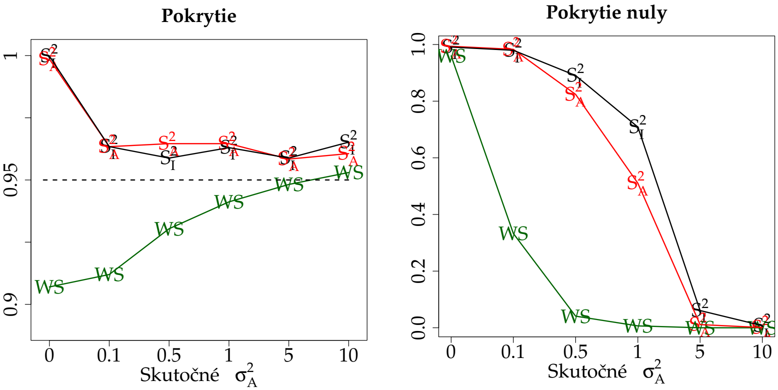
$$\underbrace{\widetilde{Y}^T B_V (B_V^T \Sigma B_V)^{-1} B_V^T \widetilde{Y}}_{= WS \text{ (Wimmer, Witkovský (2003))}} \sim \chi^2_{k-1}$$

$$\widetilde{Y}^T B_V (B_V^T \Sigma B_V)^{-1} B_V^T \widetilde{Y} = \widetilde{Y}^T (P_V^T \Sigma P_V)^+ \widetilde{Y}$$

$$\Sigma = Var(\widetilde{Y}) = \sigma_A^2 B^T Z Z^T B + \sigma_1^2 B^T (I_{n_1})_{11} B + \dots + \sigma_k^2 B^T (I_{n_k})_{kk} B$$

monotónnosť v σ_A^2

Vlastnosti



$n_i = (20, 18, 16, 14, 12, 10, 8, 6, 4, 2)$
$\sigma_i^2 = (.001, .01, .1, 1, 2, 5, 7, 10, 15, 20)$

$n_i = (2, 4, 6, 8, 10, 12, 14, 16, 18, 20)$
$\sigma_i^2 = (.001, .01, .1, 1, 2, 5, 7, 10, 15, 20)$

Ďakujem za pozornosť