Ocena komponent disperzije

Milena Kovač

16. januar 2012

Analiza variance

- Določitev podrobne **strukture** varianc in kovarianc
    - naključni vplivi in ostanek
    - homogenost oz. heterogenost
    - določitev parametrov

$$\mathbf{V} = var\left(\mathbf{y}\right) = \mathbf{ZGZ}' + \mathbf{R}$$

- **Ocena** (izračun) komponent varianc in kovarianc
    - metoda ocenjevanja parametrov disperzije
    - metoda optimizacije

Statistični modeli za oceno parametrov disperzije

- praviloma isti kot za oceno lokacijskih parametrov
- lahko vpliva na ocene komponent variance
- preverimo strukturo podatkov - količino informacij za posamezne komponente
  - število nivojev in porazdelitev opazovanj po nivojih
  - pri genetskih: porazdelitev potomcev po prednikih
- zadostno število informacij $\Rightarrow$ parameter disperzije ocenimo
- **pri pomanjkanju informacij**
  - predpostavimo, da je nič in ni v modelu
  - vključi se v ostanek ali drugo komponento

Izpeljani parametri

- ocenjujemo (ko)variance, a primerjamo deleže varianc, razmerja med variancami in korelacije

Fenotipska varianca	$\sigma^2 = \sigma_a^2 + \sigma_g^2 + \cdots + \sigma_e^2$
Fenotipska kovarianca	$\sigma_{12} = \sigma_{a12} + \sigma_{g12} + \cdots + \sigma_{e12}$
Deleži varianc	$u^2 = \sigma_u^2/\sigma^2;\ u = a, g, \cdots, e$
Heritabiliteta	$h^2 = \sigma_a^2/\sigma^2$
Razmerja varianc	$\alpha = \sigma_e^2/\sigma_a^2$
Korelacije	$r_{uii'} = \sigma_{uii'}/\sqrt{\sigma_{ui}^2\sigma_{ui'}^2};\ u = a, g, \cdots, e$

Katere komponente (ko)variance?

- **Število lastnosti v modelu**
  - enolastnostni modeli: samo variance
  - večlastnostni modeli: tudi kovariance
- **Naključni vplivi**
  - ostanek: manjkajoče vrednosti
  - trivialni naključni vplivi: nivoji nekorelirani
  - naključni vplivi z koreliranimi (sorodnimi nivoji)
    - ⋆ aditivni genetski vpliv: matrika sorodstva $\mathbf{A}$
    - ⋆ dominanca: matrika $\mathbf{D}$
- **Porazdelitev naključnih spremenljivk**
  - homogene komponente (ko)variance
  - heterogene komponente (ko)variance

Struktura (ko)varianc v skalarni obliki

- Izhajamo iz statističnega modela

$$
\begin{aligned}
y_{ijklmno} = \; & \mu + P_i + S_j + b\left(x_{ijklmno} - \overline{x}\right) + \\
& + r_{jk} + g_{ijkl} + a_{ijklm} + p_{ijklmn} + e_{ijklmno}
\end{aligned}
$$

- Pričakovana vrednost

$$
\begin{aligned}
E\left(y_{ijklmno}\right) = \; & E\left(\mu + P_i + S_j + b\left(x_{ijklmno} - \overline{x}\right) + \right. \\
& \left. + r_{jk} + g_{ijkl} + a_{ijklm} + p_{ijklmn} + e_{ijklmno}\right) = \\
& = \mu + P_i + S_j + b\left(x_{ijklmno} - \overline{x}\right)
\end{aligned}
$$

- struktura varianc

- Določimo strukturo varianc za opazovanja

$$var\left(y_{ijklmno}\right) =$$

- Vstavimo model

$$var\left(y_{ijklmno}\right) = \quad var\left(\mu + P_i + S_j + b\left(x_{ijklmno} - \overline{x}\right) + \right.$$

$$\left. + r_{jk} + g_{ijkl} + a_{ijklm} + p_{ijklmn} + e_{ijklmno}\right) =$$

- Zadostuje samo naključni del modela in ostanek

$$= var\left(r_{jk} + g_{ijkl} + a_{ijklm} + p_{ijklmn} + e_{ijklmno}\right) =$$

- Kovariance med naključnimi vplivi so vse enake 0, zato

$$= \sigma_r^2 + \sigma_g^2 + \sigma_a^2 + \sigma_p^2 + \sigma_e^2$$

Kovarianca med opazovanji

- Vzemite,
    - da sta živali iz iste črede r_{jk}
    - da imata živali a_{ijklm} in $a_{ijkl'm'}$ istega očeta, a različno mater,
    - zato sta tudi iz različnega gnezda g_{ijkl} in $g_{ijkl'}$
    - opazovanji iz različnih permanentneh okolij: p_{ijklmn} in $p_{ijkl'm'n}$

$$cov\left(y_{ijklmno}, y_{ijkl'm'no}\right) =$$

- lahko uporabimo samo naključni del modela

$$= cov\left(r_{jk} + g_{ijkl} + a_{ijklm} + p_{ijklmn} + e_{ijklmno},\right.$$
$$\left. r_{jk} + g_{ijkl'} + a_{ijkl'm'} + p_{ijkl'm'n} + e_{ijkl'm'no}\right) =$$

Kovarianca med opazovanji - II

$$= cov\left(r_{jk} + g_{ijkl} + a_{ijklm} + p_{ijklmn} + e_{ijklmno},\right.$$

$$\left. r_{jk} + g_{ijkl'} + a_{ijkl'm'} + p_{ijkl'm'n} + e_{ijkl'm'no}\right) =$$

- razčlenimo

$$= cov\left(r_{jk},\ r_{jk}\right) + cov\left(g_{ijkl},\ g_{ijkl'}\right) + cov\left(a_{ijklm},\ a_{ijkl'm'}\right) +$$

$$+cov\left(p_{ijklmn},\ p_{ijkl'm'n}\right) + cov\left(e_{ijklmno},\ e_{ijkl'm'no}\right) + 0 + \cdots =$$

- Vse kovariance med vplivi so enake 0

Kovarianca med opazovanji - III

- Preverimo posamezne člene

živali iz iste črede:

$$cov\left(r_{jk},\ r_{jk}\right) + \quad \Rightarrow \quad var\left(r_{jk}\right) = \sigma_r^2$$

živali iz različnega gnezda:

$$cov\left(g_{ijkl},\ \ g_{ijkl'}\right) + \quad \Rightarrow \quad 0$$

živali sta polsestri/polbrata:

$$cov\left(a_{ijklm},\ \ a_{ijkl'm'}\right) + \quad \Rightarrow \quad {}^1/_4\sigma_a^2$$

meritvi nimata skupnega permanentnega okolja

$$cov\left(p_{ijklmn},\ \ p_{ijkl'm'n}\right) + \quad \Rightarrow \quad 0$$

ostanka nista korelirana

$$cov\left(e_{ijklmno},\ \ e_{ijkl'm'no}\right) \quad \Rightarrow \quad 0$$

Kovarianca med opazovanji - IV

- Rezultat:

$$= \sigma_r^2 + 0 + {}^1\!/_4\sigma_a^2 + 0 + 0$$

- Kovarianco za izbrani opazovanji povzroča
  - skupno okolje v gnezdu in
  - genetska komponenta, ki je odvisna od stopnje sorodstva

- Konstanta pri aditivni genetski varianci je odvisna od deleža skupnih genov

Homogene variance

- Komponente variance za eno lastnost v populaciji iste
  - Meritve opravljene z isto natančnostjo, istim instrumentom
  - Populacija v podobnih, ne zelo različnih pogojih
- Porazdelitev naključnih vplivov: žival ($\mathbf{a}$), gnezdo($\mathbf{g}$) in ostanek ($\mathbf{e}$)

$$\mathbf{a} \sim \mathbf{A}\widehat{\sigma}_a^2 \qquad \mathbf{g} \sim \mathbf{I}\widehat{\sigma}_g^2, \qquad \cdots, \qquad \mathbf{e} \sim \mathbf{I}\widehat{\sigma}_e^2$$

- Za vsak naključni vpliv le ena, skupna komponenta
- Tudi za ostanek le ena, skupna komponenta

Heterogene komponente variance

- Kdaj: različna natančnost merjenja ali izvedbe preizkusa, različni instrumenti, različni ocenjevalci ...
  - Lahko so heterogene samo nekatere komponente (okolje)
  - Populacija razdeljena na **subpopulacije** (le zmerna delitev)
  - Heteregene genetske variance med subpopulacijami z nesorodnimi živalmi
- Ena ali več naključnih spremenljivk več komponent
- Skupine so lahko iste ali različno formirane

(Ko)varianca genetskih naključnih spremenljivk

- Vpliv živali: r subpopulacij (npr. genotipov)

$$var(\mathbf{a}) = \begin{bmatrix} \mathbf{A}_1 \widehat{\sigma}^2_{a1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{A}_r \widehat{\sigma}^2_{ar} \end{bmatrix} = \sum_{i=1}^{r} {}^{\oplus} \mathbf{A}_i \widehat{\sigma}^2_{ai}$$

- Subpopulacije nesorodne in zaprte, v njih so različne živali
- Kadar subpopulacije niso genetsko ločene, je zagovarjanje heterogenih komponent genetske variance težko. Spremeniti bi se morali koeficienti sorodstva. Tako potomca iz različnih subpopulacij ne bi bila s staršem enako sorodna.

(Ko)varianca okoljskih naključnih spremenljivk

- Vpliv skupnega okolja v gnezdu: q subpopulacij (čreda, genotipi ...)

$$var\left(\mathbf{g}\right) = \begin{bmatrix} \mathbf{I}_1\widehat{\sigma}^2_{g1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I}_q\widehat{\sigma}^2_{gq} \end{bmatrix} = \sum_{i=1}^{q}{}^{\oplus}\widehat{\sigma}^2_{gi}$$

- Ostanek: s subpopulacij (območje, čreda, genotipi ...)

$$var\left(\mathbf{e}\right) = \begin{bmatrix} \mathbf{I}_1\widehat{\sigma}^2_{e1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I}_s\widehat{\sigma}^2_{es} \end{bmatrix} = \sum_{i=1}^{s}{}^{\oplus}\widehat{\sigma}^2_{ei}$$

- Okoljske komponente med skupinami - nekorelirane ($\mathbf{0}$)

Večlastnostni modeli

- Za vpliv živali vključeno r lastnosti
- Vektor plemenskih vrednosti za $t-$ to lastnost

$$\mathbf{a}_t' = \left[\begin{array}{cccc} a_{t1} & a_{t2} & \cdots & a_{tm} \end{array} \right]$$

- in za vse lastnosti ($t = 1, 2 \cdots, r$)

$$\mathbf{a}' = \left[\begin{array}{ccc} \mathbf{a}_1' & \cdots & \mathbf{a}_r' \end{array} \right]$$

- Razvrstitev primerna za algebro
- Matrika genetskih varianc in kovarianc

$$var\left(\mathbf{a}\right) = var \left[\begin{array}{c} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \end{array} \right] = \left[\begin{array}{ccc} \mathbf{A}\widehat{\sigma}_{a1}^2 & \cdots & \mathbf{A}\widehat{\sigma}_{a1r} \\ \vdots & \ddots & \vdots \\ \mathbf{A}\widehat{\sigma}_{a1r} & \cdots & \mathbf{A}\widehat{\sigma}_{ar}^2 \end{array} \right] = \widehat{\mathbf{G}}_{a0} \otimes \mathbf{A}$$

- Kronecker produkt

Matriko **A** izpostavimo zadaj

$$\begin{bmatrix} \mathbf{A}\widehat{\sigma}^2_{a1} & \cdots & \mathbf{A}\widehat{\sigma}_{a1r} \\ \vdots & \ddots & \vdots \\ \mathbf{A}\widehat{\sigma}_{a1r} & \cdots & \mathbf{A}\widehat{\sigma}^2_{ar} \end{bmatrix} = \begin{bmatrix} \widehat{\sigma}^2_{a1} & \cdots & \widehat{\sigma}_{a1r} \\ \vdots & \ddots & \vdots \\ \widehat{\sigma}_{a1r} & \cdots & \widehat{\sigma}^2_{ar} \end{bmatrix} \otimes \mathbf{A}$$

Preostanek skrčimo v G_o

- matriko izpostavino vedno zadaj
- skalarji ostanejo v prvi matriki, ki se skrči

- matrika aditivnih genetskih varianc

$$var\left(\mathbf{a}\right) = \widehat{\mathbf{G}}_{a0} \otimes \mathbf{A}$$

- Matrika genetskih varianc in kovarianc za lastnosti merjene na isti živali

$$\widehat{\mathbf{G}}_{a0} = \begin{bmatrix} \widehat{\sigma}^2_{a1} & \cdots & \widehat{\sigma}_{a1r} \\ \vdots & \ddots & \vdots \\ \widehat{\sigma}_{a1r} & \cdots & \widehat{\sigma}^2_{ar} \end{bmatrix}$$

- Lastnosti so genetsko korelirane ($\widehat{\sigma}_{arr'}$)
- Kadar so vse lastnosti nekorelirane, je rezultat večlastnostnega modela enak rezultatom več enolastnostnih modelov
- Navadno variance ocenimo iz podatkov, zato smo uporabili ocene parametrov disperzije

- matrike (ko)varianc za okoljske vplive in ostanek

- Vpliv skupnega okolja v gnezdu: q lastnosti

$$var\left(\mathbf{g}\right)=\begin{bmatrix}\mathbf{I}_g\widehat{\sigma}^2_{g1} & \cdots & \mathbf{I}_g\widehat{\sigma}_{g1q}\\ \vdots & \ddots & \vdots\\ \mathbf{I}_g\widehat{\sigma}_{g1q} & \cdots & \mathbf{I}_g\widehat{\sigma}^2_{gq}\end{bmatrix}=\widehat{\mathbf{G}}_{g0}\otimes\mathbf{I}_g$$

- Ostanek: r lastnosti (vse lastnosti

$$var\left(\mathbf{e}\right)=\begin{bmatrix}\mathbf{I}\widehat{\sigma}^2_{e1} & \cdots & \mathbf{I}\widehat{\sigma}_{e1r}\\ \vdots & \ddots & \vdots\\ \mathbf{I}\widehat{\sigma}_{e1r} & \cdots & \mathbf{I}\widehat{\sigma}^2_{er}\end{bmatrix}=\widehat{\mathbf{R}}_{0}\otimes\mathbf{I}$$

- Okoljske komponente med skupinami - nekorelirane ($\mathbf{0}$)

Druga razporeditev opazovanj

- Lastnosti razvrščene znotraj živali

$$\mathbf{a}'_i = \left[\begin{array}{cccc} a_{i1} & a_{i2} & \cdots & a_{ir} \end{array}\right]; \text{za vse lastnosti } i = 1, 2 \cdots, r$$

- Genetske (ko)variance za lastnosti merjene na isti živali

$$\widehat{\mathbf{G}}_{a0} = \left[\begin{array}{ccc} \widehat{\sigma}^2_{a1} & \cdots & \widehat{\sigma}_{a1r} \\ \vdots & \ddots & \vdots \\ \widehat{\sigma}_{a1r} & \cdots & \widehat{\sigma}^2_{ar} \end{array}\right]$$

- Vektor plemenskih vrednosti za vse živali

$$\mathbf{a}' = \left[\begin{array}{ccccc} \mathbf{a}'_1 & \cdots & \mathbf{a}'_i & \cdots & \mathbf{a}'_m \end{array}\right]$$

Matrika genetskih varianc in kovarianc

$$
var\left(\mathbf{a}\right) = \begin{bmatrix} a_{11}\widehat{\mathbf{G}}_{0a} & \cdots & a_{1i}\widehat{\mathbf{G}}_{0a} & \cdots & a_{1m}\widehat{\mathbf{G}}_{0a} \\ & \ddots & \vdots & \cdots & \vdots \\ & & a_{ii}\widehat{\mathbf{G}}_{0a} & \cdots & a_{im}\widehat{\mathbf{G}}_{0a} \\ & & & \ddots & \vdots \\ sim. & & & \cdots & a_{mm}\widehat{\mathbf{G}}_{0a} \end{bmatrix} =
$$

$$
= \mathbf{A} \otimes \widehat{\mathbf{G}}_{0a}
$$

- razvrstitev uporabljena pogosto pri praktičnem delu
- razvrstitev ne vpliva na rezultate, le na razporeditev
- tudi pri drugih naključnih spremenljivkah se struktura spremeni

- matrike (ko)varianc za okoljske vplive in ostanek

- Vpliv skupnega okolja v gnezdu: q lastnosti

$$var\left(\mathbf{g}\right)=\begin{bmatrix}\widehat{\mathbf{G}}_{g0} & \cdots & \mathbf{0}\\ \vdots & \ddots & \vdots\\ \mathbf{0} & \cdots & \widehat{\mathbf{G}}_{g0}\end{bmatrix}=\mathbf{I}_g\otimes\widehat{\mathbf{G}}_{g0}=\sum_{i=1}^{n_g}{}^{\oplus}\widehat{\mathbf{G}}_{g0}$$

- Ostanek: s subpopulacij (območje, čreda, genotipi ...)

$$var\left(\mathbf{e}\right)=\begin{bmatrix}\widehat{\mathbf{R}}_{0} & \cdots & \mathbf{0}\\ \vdots & \ddots & \vdots\\ \mathbf{0} & \cdots & \widehat{\mathbf{R}}_{0}\end{bmatrix}=\mathbf{I}\otimes\widehat{\mathbf{R}}_{0}=\sum_{i=1}^{n}{}^{\oplus}\widehat{\mathbf{R}}_{0}$$

Direktna vsota

$$
\begin{bmatrix}
\widehat{\mathbf{R}}_{01} & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \ddots & \vdots & \vdots \\
\mathbf{0} & \cdots & \widehat{\mathbf{R}}_{0i} & \cdots \\
\mathbf{0} & \cdots & \cdots & \ddots
\end{bmatrix}
= \sum_{i=1}^{n}{}^{\oplus} \widehat{\mathbf{R}}_{0i} = \bigoplus_{i=1}^{n} \widehat{\mathbf{R}}_{0i}
$$

- za blok-diagonalne matrike, med bloki ni povezav
- bloki na diagonali so lahko različnih oblik: skalarji, vektorji, matrike
- blokov ne seštevamo, to je le način prikazovanja blok-diagonalne matrike

Manjkajoče meritve

- Matrika (ko)varianc za ostanek

$$var\left(\mathbf{e}\right) = \begin{bmatrix} \widehat{\mathbf{R}}_{01} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \widehat{\mathbf{R}}_{0i} & \cdots \\ \mathbf{0} & \cdots & \cdots & \ddots \end{bmatrix} = \sum_{i=1}^{n}{}^{\oplus}\widehat{\mathbf{R}}_{0i}$$

- Matrike $\widehat{\mathbf{R}}_{0i}$ in $\mathbf{0}$ so različnih velikosti

- Matrike (ko)varianc za naključne vplive so nespremenjene

Variance in kovariance - parametri disperzije

- Enačbe za komponente variance:
  - poiščemo **vsote kvadratnih odklonov** za posamezne naključne spremenljivke
  - jih izenačimo s pričakovano vrednostjo

$$vsota\,kvadratov = kvadratna\,oblika = E\left(kvadratne\,oblike\right)$$

- Enačba za komponente kovariance:
  - poiščemo **vsote produktov odklonov** za posamezne naključne spremenljivke in
  - jih izenačimo s pričakovano vrednostjo

$$vsota\,produktov = bilinearne\,oblike = E\left(bilinearne\,oblike\right)$$

Pričakovana vrednost kvadratne oblike

- Vzemimo, da so opazovanja porazdeljena

$$\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

- Potem je pričakovana vrednost kvadratne oblike enaka pričakovani vrednosti sledi te kvadratne oblike

$$E\left(\mathbf{y}'\mathbf{Q}\mathbf{y}\right) = E\left(tr\left(\mathbf{y}'\mathbf{Q}\mathbf{y}\right)\right) =$$

- Pri opearciji $tr\left(\right)$ lahko matrike rotiramo

$$= E\left(tr\left(\mathbf{Q}\mathbf{y}\mathbf{y}'\right)\right) =$$

- Ker je sled ($tr\left(\right)$) vsota in $E\left(\right)$ povprečje, vrstni red operacij lahko zamenjamo

$$tr\left(E\left(\mathbf{Q}\mathbf{y}\mathbf{y}'\right)\right) =$$

Pričakovana vrednost kvadratne oblike - II

- Matrika kvadratne oblike $\mathbf{Q}$ je povezana s strukturo podatkov
- Ko podatke imamo, je konstantna in jo lahko izpostavimo pri pričakovani vrednosti

$$= tr\left(\mathbf{Q}E\left(\mathbf{y}\mathbf{y}'\right)\right) =$$

- Iz definicije za varianco opazovanj (fenotipsko varianco):

$$\mathbf{V} = E\left(\mathbf{y}\mathbf{y}'\right) - E\left(\mathbf{y}\right)E\left(\mathbf{y}'\right)$$

- lahko izpeljemo enačbo za $E\left(\mathbf{y}\mathbf{y}'\right)$

$$E\left(\mathbf{y}\mathbf{y}'\right) = \mathbf{V} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'$$

- v $= tr\left(\mathbf{Q}E\left(\mathbf{y}\mathbf{y}'\right)\right) =$ nadomestimo izraz z desno stranjo enačbe

$$= tr\left(\mathbf{Q}\left(\mathbf{V} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'\right)\right) =$$

Pričakovana vrednost kvadratne oblike - III

- lahko pomnožimo in razstavimo

$$= tr\left(\mathbf{QV}\right) + tr\left(\mathbf{QX}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'\right) =$$

- matrike v drugem členu lahko ponovno rotiramo

$$= tr\left(\mathbf{QV}\right) \quad + tr\underbrace{\left(\boldsymbol{\beta}'\mathbf{X}'\mathbf{QX}\boldsymbol{\beta}\right)}_{skalar} =$$

- drugi člen je skalar, zato lahko sled izpustimo

$$= tr\left(\mathbf{QV}\right) + \quad \underbrace{\boldsymbol{\beta}'\mathbf{X}'}_{E\left(\mathbf{y}'\right)} \quad \mathbf{Q} \ \underbrace{\mathbf{X}\boldsymbol{\beta}}_{E\left(\mathbf{y}\right)}$$

- prvi člen enačbe je povezan z naključnim delom modela, drugi člen s sistematskim

Pričakovana vrednost vsote kvadratov za ostanek

$$E\left(RSS\right) = E\left(\mathbf{y}'\mathbf{Q}\mathbf{y}\right) =$$

- uporabimo pravkar razvito enačbo

$$= tr\left(\mathbf{QV}\right) + \boldsymbol{\beta}'\mathbf{X}'\mathbf{QX}\boldsymbol{\beta} =$$

- preverimo produkt matrike kvadratne oblike $\mathbf{Q}$ z $\mathbf{V}$ in $\mathbf{X}$

$$\mathbf{QV} = \quad \mathbf{QI}\sigma_e^2 = \left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\right)\sigma_e^2$$

$$\mathbf{QX} = \quad \left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\right)\mathbf{X} = \mathbf{X} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{X} = \mathbf{X} - \mathbf{X} = \mathbf{0}$$

- upoštevajmo rezultata

$$= tr\left[\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\right)\widehat{\sigma}_e^2\right] + \boldsymbol{\beta}'\mathbf{X}'\mathbf{0}\boldsymbol{\beta} =$$

- drugi člen odpade, prvega razdelimo na dve komponenti

Pričakovana vrednost vsote kvadratov za ostanek

$$= \left[tr\left(\mathbf{I}\right) - tr\left(\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\right)\right] \widehat{\sigma}_e^2 + 0 =$$

- in še izračunajmo sledi pri obeh matrikah

$$= \left[n - tr\left(\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{X}\right)\right] \widehat{\sigma}_e^2 + 0 = \left(n - p\right) \widehat{\sigma}_e^2$$

- Vsoto kvadratov za ostanek izenačiti s pričakovano vrednostjo kvadratne oblike

Ocena variance za ostanek

$$RSS = E\left(RSS\right)$$

- napišimo kvadratno obliko za vsoto kvadratov za ostanek,

$$\mathbf{y}'\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\right)\mathbf{y} =$$

- jo izenačimo s pričakovano vrednostjo kvadratne oblike

$$\mathbf{y}'\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\right)\mathbf{y} = (n-p)\,\widehat{\sigma}_e^2$$

- ker iščemo varianco za ostanek, lahko enačbo preoblikujemo

$$\widehat{\sigma}_e^2 = \frac{1}{n-p}\mathbf{y}'\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\right)\mathbf{y}$$

Preverimo

$$\widehat{\sigma}_e^2 = \tfrac{1}{n-p}\mathbf{y}'\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\right)\mathbf{y}$$

- spomnimo se še skalarne oblike

$$\widehat{\sigma}_e^2 = \tfrac{1}{n-p}\sum\left(y_{ijk} - E\left(\widehat{y}_{ijk}\right)\right)^2$$

- dobili nismo nič novega, preveritev pa je uspela
- model je bil enostaven $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$
- pri bolj kompleksnih modelih in podatkih iz prireje skalarne oblike niso enostavne

Analiza komponent variance

- **Komponente variance** dobimo tako, da
  - poiščemo primerno vsoto kvadratov,
  - jo zapišemo v obliki kvadratne oblike ($\mathbf{y}'\mathbf{Q}\mathbf{y}$) in
  - jo izenačimo z njeno pričakovano vrednostjo ($E\left(\mathbf{y}'\mathbf{Q}\mathbf{y}\right)$).
- ocenjujemo po metodi največjega verjetja / največje zanesljivosti
- sistem normalnih enačb z najnovejšimi ocenami parametrov disperzije

$$\left(\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{X}\right)\widetilde{\boldsymbol{\beta}} = \mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{y} \tag{1}$$

- napovemo naključne vplive in ostanek

$$\widetilde{\mathbf{u}} = \widetilde{\mathbf{G}}\mathbf{Z}\widetilde{\mathbf{V}}^{-1}\left(\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}\right) \tag{2}$$

$$\widetilde{\mathbf{e}} = \widetilde{\mathbf{R}}\widetilde{\mathbf{V}}^{-1}\left(\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}\right) \tag{3}$$

- poiščemo vsote kvadratov za naključne vplive in jih izenačimo s pričakovano vrednostjo

$$\widetilde{\mathbf{u}}'\widetilde{\mathbf{G}}^{-\mathbf{1}}\widetilde{\mathbf{u}} = tr\left(\mathbf{V}^{-1}\mathbf{Z}\widetilde{\mathbf{G}}\mathbf{Z}'\right) \tag{4}$$

- poiščemo vsote kvadratov za ostanek in jih izenačimo s pričakovano vrednostjo

$$\widehat{\mathbf{e}}'\widetilde{\mathbf{R}}^{-1}\widehat{\mathbf{e}} = tr\left(\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{R}}\right) \qquad (5)$$

- sistem nelinearnih enačb rešujemo iterativno, začnemo ponovno pri enačbi 1, dokler ne izpolnimo kriterija za zaustavitev ("*stopping criteria*")