Statistični modeli

I. Enačba

II. Pričakovane vrednosti

III. Struktura varianc in kovarianc

IV. Predpostavke, omejitve ...

Milena Kovač

28. oktober 2019

Primer: preizkus mladic

Žival	Gn.	Pasma	Mesec	Reja	Masa (kg)	DP (g/dan)	DHS (mm)		DSS (mm)
1	1	SL	JAN	A	102	640	13	13	14
2	2	SL	JAN	B	98	650	16	14	17
3	1	SL	FEB	C	105	650	16	16	15
4	2	SL	FEB	A	102	680	15	12	13
5	4	SVB	JAN	B	95	620	20	17	22
6	5	SVB	FEB	C	101	600	24	24	27
7	4	SVB	FEB	A	101	590	27	25	29
8	5	SML	JAN	B	97	660	26	27	25
9	4	SML	JAN	C	100	650	22	19	24
10	6	SML	FEB	A	97	700	23	25	24
11	7	SML	FEB	B	102	710	24	22	26

Modela za DP v preizkusu mladic v skalarni in matrični obliki

Dnevni prirast

$$
\begin{array}{lccccc}
y_{ijklm} = & \mu + P_i + M_j + F_k & + g_{ikl} + a_{ijklm} & + e_{ijklm} & \\
\Uparrow & \Uparrow & \Uparrow & \Uparrow & \blacksquare \\
\mathbf{y} = & \mathbf{X}\beta & + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_a\mathbf{u}_a & + \mathbf{e} &
\end{array}
$$

... če delamo z več lastnostmi

$$
\begin{array}{lccccc}
y_{1ijklm} = & \mu_1 + P_{1i} + M_{1j} + F_{1k} & + g_{1ikl} + a_{1ijklm} & + e_{1ijklm} & \\
\Uparrow & \Uparrow & \Uparrow & \Uparrow & \blacksquare \\
\mathbf{y}_1 = & \mathbf{X}_1\beta_1 & + \mathbf{Z}_{g1}\mathbf{u}_{g1} + \mathbf{Z}_{a1}\mathbf{u}_{a1} & + \mathbf{e}_1 &
\end{array}
$$

Pričakovana vrednost za DP - skalarni zapis

- Ponovimo: pričakovana vrednost opazovanja je ...

$$E\left(y_{1ijklm}\right) = E\left(\mu_1 + P_{1i} + M_{1j} + F_{1k} + g_{1ikl} + a_{1ijklm} + e_{1ijklm}\right) =$$

- lahko razstavimo

$$E\left(y_{1ijklm}\right) = E\left(\mu_1\right) + E\left(P_{1i}\right) + E\left(M_{1j}\right) + E\left(F_{1k}\right) + \\ + E\left(g_{1ikl}\right) + E\left(a_{1ijklm}\right) + E\left(e_{1ijklm}\right) =$$

- v sistematskem delu modela (indeks za lastnost smo izpustili)

$$E\left(\mu\right) = \frac{\mu+\mu+\mu+\cdots+\mu}{n} = \frac{n\mu}{n} = \mu$$
$$E\left(P_i\right) = \frac{P_i+P_i+\cdots+P_i}{n_i} = \frac{n_i P_i}{n_i} = P_i$$
$$E\left(M_j\right) = \frac{M_j+M_j+\cdots+M_j}{n_j} = \frac{n_j M_j}{n_j} = M_j$$
$$E\left(P_k\right) = \frac{P_k+P_k+\cdots+P_k}{n_k} = \frac{n_j P_k}{n_j} = P_k$$

Pričakovana vrednost za DP - skalarni zapis II

- v naključnem delu modela (indeks za lastnost smo izpustili)

$$E(g_{ik}) = \frac{\sum_{ij} g_{ik}}{n_g} = \frac{0}{n_g} = 0$$

$$E(a_{ijkl}) = \frac{\sum_{ij} a_{ijkl}}{m} = \frac{0}{m} = 0$$

- za ostanek

$$E(e_{ijklm}) = \frac{\sum_{ij} e_{ij}}{n} = \frac{0}{n} = 0$$

- vstavimo rezultate

$$\begin{aligned} = \quad & E(\mu_1) + E(P_{1i}) + E(M_{1j}) + E(F_{1k}) + \\ & + E(g_{1ikl}) + E(a_{1ijklm}) + E(e_{1ijklm}) = \end{aligned}$$

$$E(y_{1ijklm}) \quad == \quad \mu_1 + P_{1i} + M_{1j} + F_{1k} + 0 + 0 + 0$$

Pričakovana vrednost za DP - matrični zapis

- Iščemo pričakovano vrednost vseh opazovanj za prvo lastnost

$$E(\mathbf{y_1}) = E(\mathbf{X}_1\beta_1 + \mathbf{Z}_{g1}\mathbf{u}_{g1} + \mathbf{Z}_{a1}\mathbf{u}_{a1} + \mathbf{e}_1) =$$

- Razstavimo (prva vrstica) in izpostavimo konstante (druga vrstica)

$$\begin{aligned} E(\mathbf{y_1}) &= E(\mathbf{X}_1\beta_1) + E(\mathbf{Z}_{g1}\mathbf{u}_{g1}) + E(\mathbf{Z}_{a1}\mathbf{u}_{a1}) + E(\mathbf{e}_1) = \\ &\quad \curvearrowleft \qquad\qquad \curvearrowleft \qquad\qquad \curvearrowleft \\ &= \mathbf{X}_1 E(\beta_1) + \mathbf{Z}_{g1} E(\mathbf{u}_{g1}) + \mathbf{Z}_{a1} E(\mathbf{u}_{a1}) + E(\mathbf{e}_1) = \end{aligned}$$

- potrebujemo pričakovano vrednost za neznane parametre in ostanke

Pričakovana vrednost za parametre v sistematskem delu modela

- Ponovimo pričakovano vrednost v skalarni obliki

$$E(\mu_1) = \frac{\mu_1+\mu_1+\mu_1+\cdots+\mu_1}{n_1} = \frac{n_1\mu_1}{n_1} = \mu_1$$
$$E(P_{1i}) = \frac{P_{1i}+P_{1i}+\cdots+P_{1i}}{n_{1i}} = \frac{n_iP_{1i}}{n_{1i}} = P_{1i}\ldots$$

- Prenesimo v matrično obliko

$$E(\beta_1) = E\left(\begin{bmatrix} \mu_1 \\ P_{11} \\ P_{12} \\ P_{13} \\ M_{11} \\ M_{12} \\ F_{11} \\ F_{12} \\ F_{13} \end{bmatrix}\right) = \begin{bmatrix} E(\mu_1) \\ E(P_{11}) \\ E(P_{12}) \\ E(P_{13}) \\ E(M_{11}) \\ E(M_{12}) \\ E(F_{11}) \\ E(F_{12}) \\ E(F_{13}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ P_{11} \\ P_{12} \\ P_{13} \\ M_{11} \\ M_{12} \\ F_{11} \\ F_{12} \\ F_{13} \end{bmatrix} = \beta_{19\times 1}$$

Pričakovana vrednost za skupno okolje v gnezdu

- v skalarni obliki

$$E\left(g_{1ik}\right) = \frac{\sum_{ik} g_{1ik}}{n_{1g}} = \frac{0}{n_{1g}} = 0$$

- v matrični obliki

$$E\left(\mathbf{u}_{g1}\right) = E\left(\begin{bmatrix} g_{11} \\ g_{12} \\ g_{13} \\ g_{14} \\ g_{15} \\ g_{16} \\ g_{17} \end{bmatrix}\right) = \begin{bmatrix} E\left(g_{11}\right) \\ E\left(g_{12}\right) \\ E\left(g_{13}\right) \\ E\left(g_{14}\right) \\ E\left(g_{15}\right) \\ E\left(g_{16}\right) \\ E\left(g_{17}\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}_{7\times 1}$$

Pomni! Pričakovana vrednost za vektor parametrov za skupno okolje je ničelni vektor!

Pričakovana vrednost za plemensko vrednost

- v skalarni obliki

$$E\left(a_{1ijkl}\right) = \frac{\sum_{ijkl} a_{1ijkl}}{m} = \frac{0}{m} = 0$$

- število nivojev (m) je pri naključnih vplivih iz istega vira enako

- v matrični obliki

$$E\left(\mathbf{u}_{a1}\right) = E\left(\begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \\ a_{15} \\ \vdots \\ a_{1m} \end{bmatrix}\right) = \begin{bmatrix} E\left(a_{11}\right) \\ E\left(a_{12}\right) \\ E\left(a_{13}\right) \\ E\left(a_{14}\right) \\ E\left(a_{15}\right) \\ \vdots \\ E\left(a_{1m}\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}_{m\times 1}$$

Pomni! Pričakovana vrednost za vektor plemenskih vrednosti je pogosto ničelni vektor!

Pričakovana vrednost za ostanek

- v skalarni obliki

$$E\left(e_{ijklm}\right) = \frac{\sum_{ij} e_{ij}}{n} = \frac{0}{n} = 0$$

- v matrični obliki

$$E\left(\mathbf{e}_1\right) = E\left(\begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{15} \\ \vdots \\ e_{110} \\ e_{111} \end{bmatrix}\right) = \begin{bmatrix} E\left(e_{11}\right) \\ E\left(e_{12}\right) \\ E\left(e_{13}\right) \\ E\left(e_{14}\right) \\ E\left(e_{15}\right) \\ \vdots \\ E\left(e_{110}\right) \\ E\left(e_{111}\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}_{11\times 1}$$

Pomni! Pričakovana vrednost za vektor ostankov je ničelni vektor!

... dokončajmo delo

- Prepišimo enačbo, s katero se ukvarjamo

$$
\begin{aligned}
E(\mathbf{y_1}) &= E(\mathbf{X}_1\beta_1) + E(\mathbf{Z}_{g1}\mathbf{u}_{g1}) + E(\mathbf{Z}_{a1}\mathbf{u}_{a1}) + E(\mathbf{e}_1) = \\
&\quad \curvearrowleft \qquad\qquad \curvearrowleft \qquad\qquad \curvearrowleft \\
&= \mathbf{X}_1 E(\beta_1) + \mathbf{Z}_{g1} E(\mathbf{u}_{g1}) + \mathbf{Z}_{a1} E(\mathbf{u}_{a1}) + E(\mathbf{e}_1) =
\end{aligned}
$$

- Vstavimo rezultate

$$
\begin{aligned}
E(\mathbf{y_1}) &= \mathbf{X}_1\beta_1 + \underbrace{\mathbf{Z}_{g1}\mathbf{0}}_{=\mathbf{0}} + \underbrace{\mathbf{Z}_{a1}\mathbf{0}}_{=\mathbf{0}} + \mathbf{0} = \\
&= \mathbf{X}_1\beta_1
\end{aligned}
$$

Modela za DHS v preizkusu mladic v skalarni in matrični obliki

Debelina hrbtne slanine

$$
\begin{array}{lcccc}
y_{ijklmn} = & \mu + P_i + M_j + F_k + b_i\,(x_{ijklm} - 100) & +g_{ikl} + a_{ijklm} & +e_{ijklmn} \\
\Uparrow & \Uparrow & \Uparrow & \Uparrow \\
\mathbf{y} = & \mathbf{X}\beta & +\mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_a\mathbf{u}_a & +\mathbf{e}
\end{array}
$$

... če delamo z več lastnostmi

$$
\begin{array}{lcccc}
y_{2ijklmn} = & \mu_2 + P_{2i} + M_{2j} + F_{2k} + b_{2i}\,(x_{ijklm} - 100) & +g_{2ikl} + a_{2ijklm} & +e_{2ijklmn} \\
\Uparrow & \Uparrow & \Uparrow & \Uparrow \\
\mathbf{y}_2 = & \mathbf{X}_2\beta_2 & +\mathbf{Z}_{g2}\mathbf{u}_{g2} + \mathbf{Z}_{a2}\mathbf{u}_{a2} & +\mathbf{e}_2
\end{array}
$$

... pričakovana vrednost DHS v skalarni obliki

- Izvrednotimo pričakovano vrednost opazovanj za DHS v skalarni obliki

$$\begin{aligned} E\left(y_{2ijklmn}\right) &= E\left(\mu_2 + P_{2i} + M_{2j} + F_{2k} + b_{2i}\left(x_{ijklm} - 100\right) + \right. \\ &\quad \left. + g_{2ikl} + a_{2ijklm} + e_{2ijklmn}\right) = \end{aligned}$$

- Razčlenimo

$$\begin{aligned} E\left(y_{2ijklmn}\right) &= E\left(\mu_2\right) + E\left(P_{2i}\right) + E\left(M_{2j}\right) + E\left(F_{2k}\right) + E\left(b_{2i}\left(x_{ijklm} - 100\right)\right) + \\ &\quad + E\left(g_{2ikl}\right) + E\left(a_{2ijklm}\right) + E\left(e_{2ijklmn}\right) = \end{aligned}$$

- Izpostavimo konstanto $\left(x_{ijklm} - 100\right)$

$$\begin{aligned} E\left(y_{2ijklmn}\right) &= E\left(\mu_2\right) + E\left(P_{2i}\right) + E\left(M_{2j}\right) + E\left(F_{2k}\right) + \left(x_{ijklm} - 100\right) E\left(b_{2i}\right) + \\ &\quad + E\left(g_{2ikl}\right) + E\left(a_{2ijklm}\right) + E\left(e_{2ijklmn}\right) = \end{aligned}$$

... pričakovana vrednost DHS v skalarni obliki

- Preverimo pričakovano vrednost za regresijski koeficient

$$E\left(b_{2i}\right) = \frac{b_{2i}+b_{2i}+b_{2i}+\cdots+b_{2i}}{n_{1i}} = \frac{n_i b_{2i}}{n_{1i}} = b_{2i} \ldots$$

- Vstavimo rezultat v nedokončano enačbo

$$E\left(y_{2ijklmn}\right) = \mu_2 + P_{2i} + M_{2j} + F_{2k} + b_{2i}\left(x_{ijkl} - 100\right) + 0 + 0 + 0$$

Pomni! Pričakovana vrednost opazovanj je odvisna od sistematskega dela modela!

Razlikujte: matrika oz. vektor $\mathbf{0}$ in skalar 0

... pričakovana vrednost DHS v matrični obliki

- Vektor opazovanj zamenjamo z desno stranjo enačbe modela

$$E(\mathbf{y_2}) = E(\mathbf{X}_2\beta_2 + \mathbf{Z}_{g2}\mathbf{u}_{g2} + \mathbf{Z}_{a2}\mathbf{u}_{a2} + \mathbf{e}_2) =$$

- Razčlenimo

$$E(\mathbf{y_2}) = E(\mathbf{X}_2\beta_2) + E(\mathbf{Z}_{g2}\mathbf{u}_{g2}) + E(\mathbf{Z}_{a2}\mathbf{u}_{a2}) + E(\mathbf{e}_2) =$$

- Izpostavimo konstante - matrike dogodkov

$$E(\mathbf{y_2}) = \overset{\curvearrowleft}{\mathbf{X}_2 E(\beta_2)} + \overset{\curvearrowleft}{\mathbf{Z}_{g2} E(\mathbf{u}_{g2})} + \overset{\curvearrowleft}{\mathbf{Z}_{a2} E(\mathbf{u}_{a2})} + E(\mathbf{e}_2) =$$

$$E(\mathbf{y_2}) = \mathbf{X}_2\beta_2 + \mathbf{0} + \mathbf{0} + \mathbf{0}$$

Pomni! Matrike/vektorje izpostavimo na isto stran, kot so v izhodiščnem izrazu. V teh primerih so na levi, zato gredo na levo!

Modela za DSS v preizkusu mladic v skalarni in matrični obliki

Debelina stranske slanine

$$
\begin{array}{ccccccc}
y_{ijklm} = & \mu + P_i + M_j + F_k + b_i\,(x_{ijklm} - 100) & & + g_{ikl} + a_{ijklm} & & + e_{ijklm} \\
\Uparrow & \Uparrow & & \Uparrow & & \Uparrow & \\
\mathbf{y} = & \mathbf{X}\beta & & + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_a\mathbf{u}_a & & + \mathbf{e} &
\end{array}
$$

... če delamo z več lastnostmi

$$
\begin{array}{ccccccc}
y_{3ijklm} = & \mu_3 + P_{3i} + M_{3j} + F_{3k} + b_{3i}\,(x_{ijklm} - 100) & & + g_{3ikl} + a_{3ijklm} & & + e_{3ijklm} \\
\Uparrow & \Uparrow & & \Uparrow & & \Uparrow & \\
\mathbf{y}_3 = & \mathbf{X}_3\beta_3 & & + \mathbf{Z}_{g3}\mathbf{u}_{g3} + \mathbf{Z}_{a3}\mathbf{u}_{a3} & & + \mathbf{e}_3 &
\end{array}
$$

- Za vajo izpeljite pričakovano vrednost opazovanj v skalarni in matrični obliki

Pričakovana vrednost ob pogoju - skalarna oblika

- Srednja vrednost in parametri pri sistematskih vplivih

$$E\left(\mu_3 \middle| u_{a3ijkl}\right) = \mu_3$$

- Ostanek

$$E\left(e_{3ijkl} \middle| u_{aijkl}\right) = E\left(e_{ijkl}\right) = 0$$

- Naključni vplivi, ki niso izbrani

$$E\left(u_{vij} \middle| u_{aijkl}\right) = E\left(u_{vij}\right) = 0$$

- Izbrani naključni vpliv ni več naključni...

$$E\left(u_{aijkl} \middle| u_{aijkl}\right) = u_{aijkl}$$

Pričakovana vrednost ob pogoju - vpliv živali

- Izberem skupno okolje v gnezdu

$$E(\mathbf{y}_3) = E(\mathbf{X}_3\beta_3 + \mathbf{Z}_{g3}\mathbf{u}_{g3} + \mathbf{Z}_{a3}\mathbf{u}_{a3} + \mathbf{e}_3|\mathbf{u}_{a3})$$

- Razčlenimo

$$E(\mathbf{y}_3) = E(\mathbf{X}_3\beta_3|\mathbf{u}_{a3}) + E(\mathbf{Z}_{g3}\mathbf{u}_{g3}|\mathbf{u}_{a3}) + E(\mathbf{Z}_{a3}\mathbf{u}_{a3}|\mathbf{u}_{a3}) + E(\mathbf{e}_3|\mathbf{u}_{a3})$$

- Izpostavimo konstante - matrike dogodkov

$$E(\mathbf{y_3}) = \overset{\curvearrowleft}{\mathbf{X}_3 E(\beta_3|\mathbf{u}_{a3})} + \overset{\curvearrowleft}{\mathbf{Z}_{g3} E(\mathbf{u}_{g3}|\mathbf{u}_{a3})} + \overset{\curvearrowleft}{\mathbf{Z}_{a3} E(\mathbf{u}_{a3}|\mathbf{u}_{a3})} + E(\mathbf{e}_3|\mathbf{u}_{a3})$$

- Pričakovana vrednost se spremeni samo izbranemu naključnemu vplivu

$$E(\mathbf{y_3}) = \mathbf{X}_3\beta_3 + \mathbf{Z}_{g3}\mathbf{0} + \mathbf{Z}_{a3}\underbrace{E(\mathbf{u}_{a3}|\mathbf{u}_{a3})}_{=\mathbf{u}_{a3}} + \mathbf{0} = \mathbf{X}_3\beta_3 + \mathbf{Z}_{a3}\mathbf{u}_{a3}$$

Pričakovana vrednost ob pogoju - skupno okolje

- Izberem skupno okolje v gnezdu

$$E(\mathbf{y}_3) = E(\mathbf{X}_3\beta_3 + \mathbf{Z}_{g3}\mathbf{u}_{g3} + \mathbf{Z}_{a3}\mathbf{u}_{a3} + \mathbf{e}_3|\mathbf{u}_{g3})$$

- Razčlenimo

$$E(\mathbf{y}_3) = E(\mathbf{X}_3\beta_3|\mathbf{u}_{g3}) + E(\mathbf{Z}_{g3}\mathbf{u}_{g3}|\mathbf{u}_{g3}) + E(\mathbf{Z}_{a3}\mathbf{u}_{a3}|\mathbf{u}_{g3}) + E(\mathbf{e}_3|\mathbf{u}_{g3})$$

- Izpostavimo konstante - matrike dogodkov

$$E(\mathbf{y_3}) = \overset{\curvearrowleft}{\mathbf{X}_3 E(\beta_3|\mathbf{u}_{g3})} + \overset{\curvearrowleft}{\mathbf{Z}_{g3} E(\mathbf{u}_{g3}|\mathbf{u}_{g3})} + \overset{\curvearrowleft}{\mathbf{Z}_{a3} E(\mathbf{u}_{a3}|\mathbf{u}_{g3})} + E(\mathbf{e}_3|\mathbf{u}_{g3})$$

- Pričakovana vrednost se spremeni samo izbranemu naključnemu vplivu

$$E(\mathbf{y_3}) = \mathbf{X}_3\beta_3 + \mathbf{Z}_{g3}\underbrace{E(\mathbf{u}_{g3}|\mathbf{u}_{g3})}_{=\mathbf{u}_{g3}} + \mathbf{Z}_{a3}\mathbf{0} + \mathbf{0} = \mathbf{X}_3\beta_3 + \mathbf{Z}_{g3}\mathbf{u}_{g3}$$