Zaawansowane aspekty praktycznego zastosowania modeli mieszanych

1. Wykład wstępny
2. Algebra macierzowa
3. **Best Linear Unbiased Prediction (BLUP): model jednocechowy z pojedynczym efektem losowym**
4. Best Linear Unbiased Prediction (BLUP): model wielocechowy
5. Metody redukcji wymiarów modeli wielocechowych BLUP
6. Analiza danych powtarzalnych w czasie
7. Predykcja genomowej wartości hodowlanej
8. Model jednostopniowy
9. Estymacja parametrów modelu
10. - 15. Prezentacje zaliczeniowe

1. BLUP
2. Model zwierzęcia (animal model)
3. Model ojcowski (sire model)
4. Model zwierzęcia z wykorzystaniem grup genetycznych
5. Solver

Copyright ©2022, Joanna Szyda

BLUP

Best **L**inear **U**nbiased **P**rediction

BLUP

- **Best** = maksymalizuje $r(\widehat{a}_i, a_i)$ i minimalizuje $var(\widehat{a}_i - a_i)$ ← prediction error variance (PEV)
- **Linear** = predyktor (a_i) jest liniową funkcją obserwacji (y_i)
- **Unbiased** = predyktor (a_i) jest nieobciążony $E(\hat{a}_i) = a_i$
- **Predictor** = funkcja pozwalająca na predykcję wartości hodowlanej (a_i)

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

BLUP

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e}$$

- $\boldsymbol{y}$ wektor n x 1 obserwacje fenotypowe
- $\boldsymbol{b}$ wektor p x 1 estymatory efektów stałych
- $\boldsymbol{a}$ wektor q x 1 predyktory efektów losowych addytywnie genetycznych
- $\boldsymbol{e}$ wektor n x 1 predyktory efektów losowych błędu
- $\boldsymbol{X}$ macierz n x p macierz wystąpień odnosząca y_i do poszczególnych klas efektów stałych
- $\boldsymbol{Z}$ macierz n x q macierz wystąpień odnosząca y_i do poszczególnych klas efektów losowych

n = liczba rekordów danych; p = liczba klas efektów stałych; q = liczba klas efektów losowych

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

BLUP

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e}$$

- $E(\boldsymbol{a}) = \boldsymbol{0}$
- $E(\boldsymbol{e}) = \boldsymbol{0}$
- $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{b}$
- $var(\boldsymbol{a}) = \boldsymbol{A}\sigma_a^2 = \boldsymbol{G}$
- $var(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2 = \boldsymbol{R}$
- $cov(\boldsymbol{a}, \boldsymbol{e}) = \boldsymbol{0}$
- $var(\boldsymbol{y}) = var(\boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e})$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

BLUP

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e}$$

- $var(\boldsymbol{y}) = var(\boldsymbol{Za} + \boldsymbol{e}) =$

  $\boldsymbol{Z}var(\boldsymbol{a})\boldsymbol{Z}^T + \boldsymbol{I}var(\boldsymbol{e})\boldsymbol{I}^T + cov(\boldsymbol{Za}, \boldsymbol{e}) = \boldsymbol{ZGZ}^T + \boldsymbol{R}$

- $cov(\boldsymbol{y}, \boldsymbol{a}) = cov(\boldsymbol{Za} + \boldsymbol{e}, \boldsymbol{a}) = cov(\boldsymbol{Za}, \boldsymbol{a}) + cov(\boldsymbol{e}, \boldsymbol{a}) = \mathbf{Z}cov(\boldsymbol{a}, \boldsymbol{a}) + 0 = \boldsymbol{ZG}$
- $cov(\boldsymbol{y}, \boldsymbol{e}) = cov(\boldsymbol{Za} + \boldsymbol{e}, \boldsymbol{e}) = cov(\boldsymbol{Za}, \boldsymbol{e}) + cov(\boldsymbol{e}, \boldsymbol{e}) = \boldsymbol{R}$

n = liczba rekordów danych; p = liczba klas efektów stałych; q = liczba klas efektów losowych

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

BLUP

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e}$$

- MME **M**ixed **M**odel **E**quations
- Układ równań modelu mieszanego
- Henderson (1950)
- Układ równań dla predykcji **a** i estymacji **b**

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix}$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

BLUP

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e}$$

założenia

- $\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{e}$ mają rozkład normalny
  - → **a** jest determinowane przez nieskończenie wiele niezależnych genów o addytywnych efektach (infinitesimal model)
  - → **e** jest determinowane przez nieskończenie wiele niezależnych czynników
- $\boldsymbol{G}, \boldsymbol{R}$ są znane

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model zwierzęcia

animal model

Model zwierzęcia

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

$$y_{ij} = p_i + a_j + e_{ij}$$

- y_{ij} masa ciała osobnika j płci i
- p_i efekt i-tej płci
- a_j wartość hodowlana osobnika j
- e_{ij} błąd dla rekordu osobnika j płci i

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model zwierzęcia

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e}$$

$\mathbf{y}' =$

$\mathbf{X}' =$

$\mathbf{Z} =$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Model zwierzęcia

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e}$$

$$\mathbf{y}' = [4.5\ 2.9\ 3.9\ 3.5\ 5.0]$$

$$\mathbf{X}' =$$

$$\mathbf{Z} =$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Model zwierzęcia

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e}$$

$$\mathbf{y}' = [4.5\ 2.9\ 3.9\ 3.5\ 5.0]$$

$$\mathbf{X}' = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{Z} =$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Model zwierzęcia

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e}$$

$$\mathbf{y}' = [4.5\ 2.9\ 3.9\ 3.5\ 5.0]$$

$$\mathbf{X}' = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Model zwierzęcia

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}+\boldsymbol{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix}$$

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{A}^{-1}\dfrac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{bmatrix}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.833 & 0.500 & 0.000 & -0.667 & 0.000 & -1.000 & 0.000 & 0.000 \\ 0.500 & 2.000 & 0.500 & 0.000 & -1.000 & -1.000 & 0.000 & 0.000 \\ 0.000 & 0.500 & 2.000 & 0.000 & -1.000 & 0.500 & 0.000 & -1.000 \\ -0.667 & 0.000 & 0.000 & 1.833 & 0.500 & 0.000 & -1.000 & 0.000 \\ 0.000 & -1.000 & -1.000 & 0.500 & 2.500 & 0.000 & -1.000 & 0.000 \\ -1.000 & -1.000 & 0.500 & 0.000 & 0.000 & 2.500 & 0.000 & -1.000 \\ 0.000 & 0.000 & 0.000 & -1.000 & -1.000 & 0.000 & 2.000 & 0.000 \\ 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & -1.000 & 0.000 & 2.000 \end{bmatrix}$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model zwierzęcia

Rozwiązania

$$\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{A}^{-1}\dfrac{\sigma_e^2}{\sigma_a^2} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{bmatrix}$$

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{bmatrix} 3.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 1.000 & 1.000 \\ 0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 3.667 & 1.000 & 0.000 & -1.333 & 0.000 & -2.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 4.000 & 1.000 & 0.000 & -2.000 & -2.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 1.000 & 4.000 & 0.000 & -2.000 & 1.000 & 0.000 & -2.000 \\ 1.000 & 0.000 & -1.333 & 0.000 & 0.000 & 4.667 & 1.000 & 0.000 & -2.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & -2.000 & -2.000 & 1.000 & 6.000 & 0.000 & -2.000 & 0.000 \\ 0.000 & 1.000 & -2.000 & -2.000 & 1.000 & 0.000 & 0.000 & 6.000 & 0.000 & -2.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & -2.000 & -2.000 & 0.000 & 5.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & -2.000 & 0.000 & 0.000 & -2.000 & 0.000 & 5.000 \end{bmatrix}^{-1} \begin{bmatrix} 13.0 \\ 6.8 \\ 0.0 \\ 0.0 \\ 0.0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \end{bmatrix}$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model zwierzęcia

Rozwiązania

$$\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{A}^{-1}\dfrac{\sigma_e^2}{\sigma_a^2} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{bmatrix}$$

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hline \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{array}{cr} 1 & 4.358 \\ 2 & 3.404 \\ \hline 1 & 0.098 \\ 2 & -0.019 \\ 3 & -0.041 \\ 4 & -0.009 \\ 5 & -0.186 \\ 6 & 0.177 \\ 7 & -0.249 \\ 8 & 0.183 \end{array}$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model zwierzęcia

Rozwiązania

1	4.358
2	3.404
1	0.098
2	−0.019
3	−0.041
4	−0.009
5	−0.186
6	0.177
7	−0.249
8	0.183

- Osobniki męskie (i=1) mają wyższe przyrosty niż osobniki żeńskie
- Najwyższą wartością hodowlaną charakteryzuje się osobnik 8

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model zwierzęcia

Dokładność predykcji wartości hodowlanej osobnika *i*

$$r_i = r(\hat{a}_i, a_i) = \sqrt{1 - c_i^{22}\frac{\sigma_e^2}{\sigma_a^2}}$$

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

$$\left[\begin{array}{cc|cccccccc}
3.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 1.000 & 1.000 \\
0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 \\
\hline
0.000 & 0.000 & 3.667 & 1.000 & 0.000 & -1.333 & 0.000 & -2.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.000 & 4.000 & 1.000 & 0.000 & -2.000 & -2.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 1.000 & 4.000 & 0.000 & -2.000 & 1.000 & 0.000 & -2.000 \\
1.000 & 0.000 & -1.333 & 0.000 & 0.000 & 4.667 & 1.000 & 0.000 & -2.000 & 0.000 \\
0.000 & 1.000 & 0.000 & -2.000 & -2.000 & 1.000 & 6.000 & 0.000 & -2.000 & 0.000 \\
0.000 & 1.000 & -2.000 & -2.000 & 1.000 & 0.000 & 0.000 & 6.000 & 0.000 & -2.000 \\
1.000 & 0.000 & 0.000 & 0.000 & 0.000 & -2.000 & -2.000 & 0.000 & 5.000 & 0.000 \\
1.000 & 0.000 & 0.000 & 0.000 & -2.000 & 0.000 & 0.000 & -2.000 & 0.000 & 5.000
\end{array}\right]$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model zwierzęcia

Dokładność predykcji wartości hodowlanej osobnika *i*

$$r_i = r(\hat{a}_i, a_i) = \sqrt{1 - c_i^{22}\frac{\sigma_e^2}{\sigma_a^2}}$$

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix}^{-1} = \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$$

$$\left[\begin{array}{cc|cccccccc}
0.596 & 0.157 & -0.164 & -0.084 & -0.131 & -0.265 & -0.148 & -0.166 & -0.284 & -0.238 \\
0.157 & 0.802 & -0.133 & -0.241 & -0.112 & -0.087 & -0.299 & -0.306 & -0.186 & -0.199 \\
\hline
-0.164 & -0.133 & 0.471 & 0.007 & 0.033 & 0.220 & 0.045 & 0.221 & 0.139 & 0.134 \\
-0.084 & -0.241 & 0.007 & 0.492 & -0.010 & 0.020 & 0.237 & 0.245 & 0.120 & 0.111 \\
-0.131 & -0.112 & 0.033 & -0.010 & 0.456 & 0.048 & 0.201 & 0.023 & 0.126 & 0.218 \\
-0.265 & -0.087 & 0.220 & 0.020 & 0.048 & 0.428 & 0.047 & 0.128 & 0.243 & 0.123 \\
-0.148 & -0.299 & 0.045 & 0.237 & 0.201 & 0.047 & 0.428 & 0.170 & 0.220 & 0.178 \\
-0.166 & -0.306 & 0.221 & 0.245 & 0.023 & 0.128 & 0.170 & 0.442 & 0.152 & 0.219 \\
-0.284 & -0.186 & 0.139 & 0.120 & 0.126 & 0.243 & 0.220 & 0.152 & 0.442 & 0.168 \\
-0.238 & -0.199 & 0.134 & 0.111 & 0.218 & 0.123 & 0.178 & 0.219 & 0.168 & 0.422
\end{array}\right]$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model zwierzęcia

Dokładność
Powtarzalność

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

Animal	Diagonals of inverse	r^2	r
1	0.471	0.058	0.241
2	0.492	0.016	0.126
3	0.456	0.088	0.297
4	0.428	0.144	0.379
5	0.428	0.144	0.379
6	0.442	0.116	0.341
7	0.442	0.116	0.341
8	0.422	0.156	0.395

Rodzice (1,2,3)

→ brak własnych fenotypów

→ niższa dokładność

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model ojcowski

sire model

Model ojcowski

- Oceniana jest tylko wartość hodowlana ojców
- Historycznie wcześniejszy od modelu zwierzęcia
- Mniejsze wymiary MME

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model ojcowski

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{s} + \boldsymbol{e}$$

- $\boldsymbol{y}$ wektor n x 1 obserwacje fenotypowe
- $\boldsymbol{b}$ wektor p x 1 estymatory efektów stałych
- $\boldsymbol{s}$ wektor q x 1 predyktory efektów losowych addytywnie genetycznych
- $\boldsymbol{e}$ wektor n x 1 predyktory efektów losowych błędu
- $\boldsymbol{X}$ macierz n x p macierz wystąpień odnosząca y_i do poszczególnych klas efektów stałych
- $\boldsymbol{Z}$ macierz n x q macierz wystąpień odnosząca y_i do poszczególnych klas efektów losowych

n = liczba rekordów danych; p = liczba klas efektów stałych; q = liczba klas efektów losowych

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model ojcowski

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{s} + \boldsymbol{e}$$

- $E(\boldsymbol{s}) = \boldsymbol{0}$
- $E(\boldsymbol{e}) = \boldsymbol{0}$
- $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{b}$
- $var(\boldsymbol{s}) = \boldsymbol{A}\sigma_s^2$
- $var(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2 = \boldsymbol{R}$
- $cov(\boldsymbol{s}, \boldsymbol{e}) = \boldsymbol{0}$
- $var(\boldsymbol{y}) = var(\boldsymbol{Z}\boldsymbol{s} + \boldsymbol{e}) = \boldsymbol{Z}\boldsymbol{A}\boldsymbol{Z}^{\boldsymbol{T}}\sigma_s^2 + \boldsymbol{R}$
- $\sigma_s^2 = \frac{1}{4}\sigma_a^2$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model ojcowski

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

$$y_{ij} = p_i + s_j + e_{ij}$$

- y_{ij} masa ciała osobnika j płci i
- p_i efekt i-tej płci
- s_j wartość hodowlana **ojca** osobnika j (1,3,4)
- e_{ij} błąd dla rekordu osobnika j płci i

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model ojcowski

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

$$\mathbf{y}' = [4.5\ 2.9\ 3.9\ 3.5\ 5.0]$$

$$\mathbf{X}' = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zs} + \boldsymbol{e}$$

$$\mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} ? & & \\ & & \\ & & \\ & & \end{bmatrix}$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model ojcowski

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{A}^{-1}\sigma_s^2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{bmatrix}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.333 & 0.0 & -0.667 \\ 0.000 & 1.0 & 0.000 \\ -0.667 & 0.0 & 1.333 \end{bmatrix}$$

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{s}_1 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} \begin{bmatrix} 3.000 & 0.000 & 1.000 & 1.000 & 1.000 \\ 0.000 & 2.000 & 1.000 & 1.000 & 0.000 \\ 1.000 & 1.000 & 16.666 & 0.000 & -7.334 \\ 1.000 & 1.000 & 0.000 & 13.000 & 0.000 \\ 1.000 & 0.000 & -7.334 & 0.000 & 15.666 \end{bmatrix}^{-1} = \begin{bmatrix} 13.00 \\ 6.80 \\ 8.40 \\ 7.90 \\ 3.50 \end{bmatrix}$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model ojcowski

Rozwiązania

$$\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{A}^{-1}\sigma_s^2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{bmatrix}$$

1	4.358
2	3.404
1	0.098
2	−0.019
3	−0.041
4	−0.009
5	−0.186
6	0.177
7	−0.249
8	0.183

Animal model

1	4.336
2	3.382
1	0.022
3	0.014
4	−0.043

Sire model

- Estymatory efektu płci różne
- Różnice pomiędzy efektami płci takie same

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model ojcowski

Rozwiązania

$$\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{A}^{-1}\sigma_s^2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{bmatrix}$$

1	4.358
2	3.404
1	0.098
2	−0.019
3	−0.041
4	−0.009
5	−0.186
6	0.177
7	−0.249
8	0.183

Animal model

1	4.336
2	3.382
1	0.022
3	0.014
4	−0.043

Sire model

- Predyktory s_i różne → brak informacji o a_i matek
- Rankingi s_i jednakowe

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2020, Joanna Szyda

Model zwierzęcia z wykorzystaniem grup genetycznych

animal model with phantom parent groups

Model zwierzęcia z wykorzystaniem grup genetycznych

- Osobniki o nieznanych rodzicach → osobniki reprezentujące populację bazową
- Założenie populacji bazowej → średnia a_i=0
- W praktyce poszczególne „komponenty” populacji bazowej mają różne średnie a_i z uwagi na różne intensywności selekcji:
  - Ojcowie buhajów
  - Ojcowie krów
  - Matki buhajów
  - Matki krów
  - Osobniki importowane z różnych krajów
  - Osobniki urodzone w różnych latach

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Model zwierzęcia z wykorzystaniem genetycznych

- Westell & Van Vleck (1987)

- **phantom parents** → „sztuczni” rodzice osobników z populacji bazowej

- **phantom parents** → definiowane wg „kategorii”
  - Ojcowie buhajów
  - Ojcowie krów
  - Matki buhajów
  - Matki krów
  - Kraj pochodzenia osobnika z populacji bazowej
  - Rok urodzenia osobnika z populacji bazowej

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Model zwierzęcia z wykorzystaniem genetycznych

- Legarra et al. (2015)
- **metafounders** → „sztuczni” rodzice osobników z populacji bazowej zdefiniowani na podstawie informacji genotypowej (SNP)

Ancestral Relationships Using Metafounders: Finite Ancestral Populations and Across Population Relationships FREE

Andres Legarra ✉, Ole F Christensen, Zulma G Vitezica, Ignacio Aguilar, Ignacy Misztal

Genetics, Volume 200, Issue 2, 1 June 2015, Pages 455–468, https://doi.org/10.1534/genetics.115.177014

Copyright ©2024, Joanna Szyda

Model zwierzęcia z wykorzystaniem grup genetycznych

$$y_{ij} = h_j + a_i + \sum_{k=1}^{K} t_{ik} g_k + e_{ij}$$

- y_{ij} fenotyp krowy i ze stada j
- h_j stały efekt j-tego stada
- a_i losowy efekt wartości hodowlana osobnika i
- g_k stały efekt k-tej grupy genetycznej
- t_{ik} współczynnik spokrewnienia między osobnikiem i, a grupą k
- e_{ij} błąd

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Model zwierzęcia z wykorzystaniem grup genetycznych

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{ZQg} + \boldsymbol{e}$$

$$\boldsymbol{Q} = \boldsymbol{TQ}^*$$

- $\boldsymbol{Q}^*$ macierz wystąpień odnosząca przodka do danej „sztucznej” grupy
- $\boldsymbol{T} \leftarrow \boldsymbol{A} = \boldsymbol{TDT}^T$

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'ZQ} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\alpha & \mathbf{Z'ZQ} \\ \mathbf{Q'Z'X} & \mathbf{Q'Z'Z} & \mathbf{Q'Z'ZQ} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{Q'Z'y} \end{bmatrix}$$

- Ostateczny predyktor wartości hodowlanej osobnika i: $\hat{a}_i^* = \boldsymbol{Q}\hat{\boldsymbol{g}} + \hat{a}_i$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Model zwierzęcia z wykorzystaniem grup genetycznych

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

Calf	Sire	Dam
1	Unknown	Unknown
2	Unknown	Unknown
3	Unknown	Unknown
4	1	Unknown
5	3	2
6	1	2
7	4	5
8	3	6

Calf	Sire	Dam
1	M1	M2
2	M3	M4
3	M5	M6
4	1	M7
5	3	2
6	1	2
7	4	5
8	3	6

Calf	Sire	Dam
1	G1	G2
2	G1	G2
3	G1	G2
4	1	G2
5	3	2
6	1	2
7	4	5
8	3	6

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Model zwierzęcia z wykorzystaniem grup genetycznych

$$\boldsymbol{Q} = \boldsymbol{T}\boldsymbol{Q}^*$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.25 & 0.75 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.375 & 0.625 \\ 0.5 & 0.5 \end{bmatrix} =$$

	M1	M2	M3	M4	M5	M6	M7	1	2	3	4	5	6	7	8
M1	1.000	0.000	0.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M2	0.000	1.000	0.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M3	0.000	0.000	1.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M4	0.000	0.000	0.000	1.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M5	0.000	0.000	0.000	0.000	1.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M6	0.000	0.000	0.000	0.000	0.000	1.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M7	0.000	0.000	0.000	0.000	0.000	0.000	1.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
1	0.500	0.500	0.000	0.000	0.000	0.000	0.00	1.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
2	0.000	0.000	0.500	0.500	0.000	0.000	0.00	0.00	1.00	0.00	0.0	0.0	0.0	0.0	0.0
3	0.000	0.000	0.000	0.000	0.500	0.500	0.00	0.00	0.00	1.00	0.0	0.0	0.0	0.0	0.0
4	0.250	0.250	0.000	0.000	0.000	0.000	0.50	0.50	0.00	0.00	1.0	0.0	0.0	0.0	0.0
5	0.000	0.000	0.250	0.250	0.250	0.250	0.00	0.00	0.50	0.50	0.0	1.0	0.0	0.0	0.0
6	0.250	0.250	0.250	0.250	0.000	0.000	0.00	0.50	0.50	0.00	0.0	0.0	1.0	0.0	0.0
7	0.125	0.125	0.125	0.125	0.125	0.125	0.25	0.25	0.25	0.25	0.5	0.5	0.0	1.0	0.0
8	0.125	0.125	0.125	0.125	0.250	0.250	0.00	0.25	0.25	0.50	0.0	0.0	0.5	0.0	1.0

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

R.A. Mrode 2005 Linear Models for the Prediction of Animal Breeding Values

Copyright ©2022, Joanna Szyda

Algorytm obliczeniowy do rozwiązania MME

Rozwiazanie układu równań MME

$$\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{A}^{-1}\sigma_s^2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{bmatrix}$$

$$\boldsymbol{\theta} = \boldsymbol{W}^{-1}\boldsymbol{y}$$

- Gauss-Seidel
- Jacobi
- Conjugate gradient
- Preconditioned conjugate gradient

TABLE 2. Central processing unit time in minutes used for preprocessing (CPU_{Pre}) and iterating 1000 rounds (CPU_{1000}) with different iterative methods by model[1].

	Univariate model		Multivariate model	
Iterative method[2]	CPU_{Pre}	CPU_{1000}	CPU_{Pre}	CPU_{1000}
SOR(IMIT)	2.3	19.9	41.5	102.3
JCG(IMIT)	7.1	18.1	52.6	78.2
GSJ(IOD)	1.5	142.9	2.4	365.9
GSSJ(IOD)	1.4	42.9	2.0	111.7
GSSJ(IOD,RR)[3]	2.1	34.2	2.2	95.6
PCG(IOD,RR)	1.3	22.4	1.5	69.8
PCG(IOD,RR,U)	1.3	19.6	1.5	52.6

[1]Some of the programs had the mixed model equations in the memory (IMIT), and others used iteration on data (IOD) algorithm when performing the necessary computations. Random regression test-day models (RR) were allowed by some of the programs. One of the programs used the new computing technique (U).

[2]SOR = Successive overrelaxation in PEST Version 2.8, JCG = Jacobi conjugate gradient as in DMU Version 4, GSJ = Jacobi on animal effects and Gauss-Siedel on others with relaxation as in PEST Version 2.8, GSSJ = Gauss-Seidel on herd effect and second-order Jacobi on others as in DMU Version 4 and DMUIOD Version 2.1, and PCG = preconditioned conjugate gradient method.

Journal of Dairy Science Volume 82, Issue 12, December 1999, Pages 2779-2787

Copyright ©2022, Joanna Szyda

Preconditioned conjugate gradient

$$\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{A}^{-1}\sigma_s^2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{bmatrix}$$

$$\boldsymbol{\theta} = \boldsymbol{W}^{-1}\boldsymbol{y}$$

PCG wykorzystuje “preconditioner” → macierz **P**

$$\boldsymbol{\theta} = \boldsymbol{W}^{-1}\boldsymbol{y}$$

$$\boldsymbol{P}^{-1}\boldsymbol{\theta} = \boldsymbol{P}^{-1}\boldsymbol{W}^{-1}\boldsymbol{y}$$

Front. Genet., 28 July 2021 | https://doi.org/10.3389/fgene.2021.655638

Copyright ©2022, Joanna Szyda

PCG solver

Genetics Selection Evolution

A second-level diagonal preconditioner for single-step SNPBLUP

The preconditioned conjugate gradient (PCG) method is an iterative solver of linear equations systems commonly used in animal breeding. However, the PCG method has been shown to encounter convergence issues wh...

Jeremie Vandenplas, M

Genetics Selection Evolu

Research Article | Publ

› Full Text › PDF

Convergence behavior of single-step GBLUP and SNPBLUP for different termination criteria

The preconditioned conjugate gradient (PCG) method is the current method of choice for iterative solving of genetic evaluations. The relative difference between two successive iterates and the relative residua...

Jeremie Vandenplas, Mario P. L
Vuik

Genetics Selection Evolution 202

Research Article | Published on

› Full Text › PDF

Computational strategies for the preconditioned conjugate gradient method applied to ssSNPBLUP, with an application to a multivariate maternal model

The single-step single nucleotide polymorphism best linear unbiased prediction (ssSNPBLUP) is one of the single-step evaluations that enable a simultaneous analysis of phenotypic and pedigree information of ge...

Jeremie Vandenplas, Herwin Eding, Maarten Bosmans and Mario P. L. Calus

Genetics Selection Evolution 2020 52:24

Research Article | Published on: 13 May 2020

› Full Text › PDF

Copyright ©2023, Joanna Szyda

1. BLUP
2. Model zwierzęcia (animal model)
3. Model ojcowski (sire model)
4. Model zwierzęcia z wykorzystaniem grup genetycznych
5. Solver

1. Co oznacza skrót BLUP ?
2. Który model ma większe rozmiary - animal czy sire ?
3. Czy można uzyskać predyktory wartości hodowlanej dla osobników bez potomstwa ?
4. Po co wyróżniamy grupy genetyczne?

1. BLUP
2. Model zwierzęcia (animal model)
3. Model ojcowski (sire model)
4. Model zwierzęcia z wykorzystaniem grup genetycznych
5. Solver

That BLUP is a Good Thing: The Estimation of Random Effects

G. K. Robinson

Statist. Sci. 6(1): 15-32 (February, 1991). DOI: 10.1214/ss/1177011926

ABOUT | FIRST PAGE | CITED BY | RELATED ARTICLES

Abstract

In animal breeding, Best Linear Unbiased Prediction, or BLUP, is a technique for estimating genetic merits. In general, it is a method of estimating random effects. It can be used to derive the Kalman filter, the method of Kriging used for ore reserve estimation, credibility theory used to work out insurance premiums, and Hoadley's quality measurement plan used to estimate a quality index. It can be used for removing noise from images and for small-area estimation. This paper presents the theory of BLUP, some examples of its application and its relevance to the foundations of statistics. Understanding of procedures for estimating random effects should help people to understand some complicated and controversial issues about fixed and random effects models and also help to bridge the apparent gulf between the Bayesian and Classical schools of thought.