Genetika ve šlechtění zvířat

část 8. (rough draft version)

9 Odhad PH – BLUP AM

V animal modelu (AM) se hodnotí každé zvíře samostatně a současně v závislosti na užitkovosti příbuzných jedinců hodnocené populace. Veškeré příbuzenské vztahy (pokud jsou známé) jsou zahrnuty do aditivně genetické matice příbuznosti. V AM je pro každé zvíře sestavena zvlášť rovnice a to je velký problém – mnoho rovnic.

BLUP - AM se provádí pomocí <u>distribuční funkce</u> **f(T/y)**
T – hledané veličiny (vektor)
y – naměřené užitkovosti (vektor)

Parciální **derivací** distribuční funkce (první derivace = 0 ⇒ směrnice tečny je nulová – extrém) ⇒ hledáme průběh a extrém funkce ⇒ pomocí soustavy normálních rovnic (maticová soustava) – ***Mixed Model Equation*** (**MME**) – **SMÍŠENÉ MODELY**:

$$(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{H}^{-1})\mathbf{T} = \mathbf{W}'\mathbf{R}^{-1}\mathbf{y}$$

W – matice plánu experimentu, incidenční, designová (odhad PH) – rozepisuje se na matice **X** a **Z** !
R – kovarianční matice reziduí (chyb v datech)
H – kovarianční matice mezi hledanými veličinami
T – hledaná veličina

Dále se řeší modelová rovnice (maticový zápis):

- smíšený lineární model $\mathbf{y}_{ijk} = \mathbf{b}_i + \boldsymbol{u}_j + \mathbf{e}_{ijk}$ [1]

užitkovost = **součet faktorů**, které ji ovlivňují

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

Aditivní plemenné hodnoty jsou náhodnými efekty se známou variančně -kovarianční maticí. U vektorů **u** a **e** se předpokládá, že mají normální rozdělení a tedy odhadovaná střední hodnota je E(**u**) = E(**e**) = 0. Vektor pozorování **y** má multivariátní normální rozdělení s průměrem **Xb** (E(**y**) = **Xb**) a variancí **V** (variančně kovarianční matice vektoru pozorování) vypočítanou jako:

$$\mathbf{V} = \mathbf{V}(\mathbf{Zu} + \mathbf{e}) = \mathbf{ZGZ}` + \mathbf{R} \qquad !$$

kde **G** je variančně kovarianční matice vektoru náhodných efektů **u** ~ V(**u**) a **R** je variančně kovarianční matice reziduálních chyb ~ V(**e**).

Nejsou-li otcové příbuzní pak je $\mathbf{G} = \mathbf{I}\sigma_0^2$, kde **I** je jednotková matice a σ_0^2 je ¼ aditivní genetické variance (protože každý otec dá ½ svých genů dcerám a po umocnění pro získání variance získáme ¼ aditivní genetické variance).

[1] P_i – efekt plemene i (pevný efekt), u_j – efekt otce j na produkci jeho dcer (náhodný efekt)

Jsou-li otcové příbuzní pak $\mathbf{G} = \mathbf{A}\,\sigma_0^2$, kde **A** je matice příbuznosti mezi otci s jedničkami na diagonále a mimo diagonální prvky zobrazují podíl genů, které dva jedinci mají od společného předka. Obě matice jsou symetrické (**G** =**G**\` a **A** =**A**\`).

Nejlepší lineární nevychýlená předpověď (BLUP) vektoru **u** je:

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \qquad \textbf{!}$$

kde $\hat{\mathbf{b}}$ je odhadce pevných efektů získaných metodou zobecněných nejmenších čtverců (GLS) – viz níže.

Rovnice $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \Rightarrow$ se rozepisuje do soustavy normálních rovnic smíšeného modelu (MME)

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{H}^{-} \end{bmatrix} \cdot \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Matice **T**[2] **S**

$$\mathbf{M} . \mathbf{T} = \mathbf{S}$$

$$\mathbf{T} = \mathbf{M}^{-1} . \mathbf{S}$$ [3]

y	– vektor naměřených užitkovostí (*n*)	(*n* x 1)
X	– incidenční matice udávající plán pokusu pevných efektů X	(*n* x *p*)
Z	– incidenční matice udávající plán pokusu náhodných efektů Z	(*n* x *q*)
b	– vektor odhadů pevných efektů (odhad úrovní *p*)	(*p* x 1)
u	– vektor odhadů náhodných efektů; **u** ~ PH (odhad úrovní *q*)	(*q* x 1)
e	– vektor nekontrolovatelných náhodných reziduálních efektů (vektor reziduálních odchylek, u kterých se předpokládá, že jsou nezávislé na náhodných genetických efektech (*n* x 1)	
$\mathbf{H}^{-}$	– kovarianční matice inverzní	

Stejně jako v modelu s pevnými efekty se obecně předpokládá, že rezidua jsou nekorelována a mají stejnou, konstantní varianci. Pak $\mathbf{R} = \mathbf{I}\,\sigma_e^2$, kde σ_e^2 je reziduální variance a $\mathbf{R}^{-1} = \mathbf{I}/\sigma_e^2$. Takže soustava normálních rovnic může být zjednodušena vynásobením obou stran reziduální variancí (maticí **R**) a získáme rovnice:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1}\sigma_e^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Pokud předpovídáme plemennou hodnotu pomocí AM (jsou známy příbuzenské vztahy mezi jedinci) nabývají rovnice tvar:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}K \end{bmatrix} \cdot \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

X\`**X** – diagonální matice s řádky a sloupci rovno počtu úrovní pevného efektu (např. počtu plemen), diagonální prvky jsou počty záznamů v korespondující úrovni efektu (konkrétního plemene), mimodiagonální prvky jsou rovny nule

[2] **T** chceme určit!

[3] **M** – matice koeficientů; **T** – vektor řešení

Z`Z – diagonální matice, kde každý diagonální prvek je roven počtu záznamů (počtu dcer) každé úrovně náhodného efektu (otce)

X`Z – matice s počtem řádků rovno počtu úrovní pevných efektů (počet plemen) a počtem sloupců rovno počtu úrovní náhodných efektů (počet otců); každý prvek bude číslo záznamů v odpovídající kombinaci pevného x náhodného efektu (plemeno x otec)

Z`X** – matice je transponovaná matice **X`Z

X`y – vektor jehož délka bude rovna počtu úrovní pevného efektu (počtu plemen) a každý prvek je součet hodnot v odpovídající úrovni efektu (plemene)

Z`y – vektor jehož délka bude rovna počtu úrovní náhodného efektu (počtu otců) a každý prvek je součet hodnot v odpovídající úrovni efektu (užitkovost všech dcer po každém otci)

A – aditivně genetická matice příbuznosti, jejíž prvky $\mathbf{a_{ii}}$ jsou rovny (1 + Fz) (Fz koeficient inrídingu) a prvky $\mathbf{a_{ij}}$ jsou rovny koeficientům příbuznosti R*ij* mezi jedinci *i* a *j*.

koeficient inbrídingu: $F_Z = a_{ii} = \sum \left(\frac{1}{2}\right)^{n_1+n_2+1} .(1+F_A)$

koeficient příbuznosti: $R_{XY} = a_{ij} = a_{ji} = \sum \left(\frac{1}{2}\right)^{n_1+n_2} (1+F_A)$

n_1 - počet generací mezi rodičem X a jedincem Z a společným předkem A

n_2 - počet generací mezi rodičem Y a jedincem Z a společným předkem A

F_A - koeficient inbrídingu předka A (společného předka)

Σ - sumace příbuznosti pro více úseků jedinců X a Y ke společným předkům

$$K = \frac{\sigma_e^2}{\sigma_u^2} = \frac{\sigma_e^2}{\sigma_A^2}$$ (někdy označováno jako λ nebo α)

$$= \frac{1-h^2}{h^2}$$ např. při sledování vlastní užitkovosti

$$= \frac{4-h^2}{h^2}$$ např. při sledování užitkovostí dcer (polosester) po býcích

Řešení rovnic smíšeného modelu může být obtížné získat, protože řešení rovnic je prováděno invertováním matic, což je pro běžné výpočty u velkých populací HZ méně vhodné: **T** = **M**$^{-1}$.**S**. Přesné řešení vyžaduje inverzi matice koeficientů (**M**), která je zpravidla menší než matice **V**. Počet řádků a sloupců matice **V** je rovna celkovému počtu pozorování, zatímco v matici **M** jsou rovny počtu úrovní efektů zahrnutých do modelu, což je obecně mnohem méně.

Je-li zahrnuto do modelové rovnice více faktorů, může být získáno přibližné řešení iterativními postupy (**iteracemi**) – řešení jednotlivých rovnic izolovaně, a jejich řešení je využito v dalších rovnicích s cílem stabilizace řešení, kdy se již nemění od jednoho iteračního kola k druhému. Existuje mnoho iteračních metod aplikovatelné na rovnice smíšeného modelu. Nejpoužívanější metodou jsou iterace Gauss-Seidel, protože je relativně rychlá a spolehlivě konverguje a poskytuje řešení rovnic. Animal model vyžaduje několik stovek kol iterací, než se získá přibližná konvergence.

Př. **BLUP –AM** **B**est **L**inear **U**nbiased **P**rediction – **A**nimal **M**odel

$$y_{ijk} = \mu + SRO_i + g_j + e_{ijk}$$

y_{ijk} [~ **y**] naměřená užitkovost
μ [~ **X**] populační průměr
SRO_i [~ **b**] **s**tádo x **r**ok x **o**bdobí (působení chovatele na zvířata, na jejich užitkovost)
g_j [~ **u**] efekt jedince (genetický) – ten chceme určit - **PH !**
e_{ijk} [~ **e**] reziduum

BLUP
AM (animal model) } pro 1 nebo více vlastností
RAM (redukovaný AM)
(gametický model)

AM – **A**nimal **M**odel – individuální model

- vylepšený BLUP
- nebereme v úvahu jen naměřenou užitkovost jedince, ale i všechny příbuzenské vztahy (např. odhad mléčné užitkovosti býka na základě užitkovosti dcer)

Porovnání zobecněného lineárního modelu GLM (*general linear model*), který obsahuje jen pevné efekty, se smíšeným modelem MM (*mixed model*):

GLM	**MM**
$y_{ij} = \mu + b_i + e_{ij}$	$y_{ijk} = \mu + b_i + u_j + e_{ijk}$
$\mathbf{y} = \mathbf{Xb} + \mathbf{e}^*$	$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$
$\mathbf{e}^* \sim (0, \mathbf{V})$	$\mathbf{u} \sim (0, \mathbf{G})\ \mathbf{e} \sim (0, \mathbf{R})$
$\mathbf{y} \sim (\mathbf{Xb}, \mathbf{V})$	$\mathbf{y} \sim (\mathbf{Xb}, \mathbf{V}) = (\mathbf{Xb}, \mathbf{ZGZ'} + \mathbf{R})$

~ (a, b) znamená, že náhodná proměnná má průměr a a varianci b
ve smíšeném modelu je vektor reziduálních efektů rozdělen do dvou komponent $\mathbf{e}^* = \mathbf{Zu} + \mathbf{e}$

Ve smíšeném modelu pozorujeme **y**, **X** a **Z**, zatímco **b**, **u**, **R** a **G** jsou obecně neznámé. Takže smíšené modely nám umožňují:

1) odhadovat vektory pevných **b** a náhodných efektů **u**
2) odhadovat kovarianční matice **G** a **R** (u kterých se předpokládá, že jsou funkcemi několika neznámých komponent variance)

Pro **pevné efekty** platí **BLUE** (nejlepší lineární nevychýlený odhad): $E(\hat{b} - b)^2 = \min$.

Řešení **náhodných efektů** je **BLUP** (nejlepší lineární nevychýlená předpověď) $E(\hat{u} - u)^2 = \min$, protože náhodné efekty nejsou parametry a jejich řešení se nazývají prediktory *„predictors"* a naopak u pevných efektů hovoříme o odhadcích *„estimators"*.[4]

[4] Odhadujeme pevné efekty a předpovídáme náhodné efekty !!

BLUE a **BLUP** jsou *nejlepší*, protože minimalizují výběrovou varianci; *lineární* v tom smyslu, že jsou lineárními funkcemi pozorovaných fenotypů **y**; *nevychýlené* ve smyslu, že E[BLUE(b)] = b a E[BLUP(u)] = u.

Pro smíšený model $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$ platí:

BLUE pro pevné efekty **b**: $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$

kde $\mathbf{V} = \mathbf{ZGZ}` + \mathbf{R}$

Jedná se odhadce zobecněných nejmenších čtverců (GLS)

BLUP pro náhodné efekty **u**: $\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$

(Henderson, 1963)

Praktická aplikace obou rovnic vyžaduje známé komponenty variance. Před analýzami BLUP je nutné odhadnout komponenty variance pomocí ANOVA nebo REML.

- historie (výpočty ovlivňuje úroveň výpočetní techniky)
  - Bayes (1763) – BLUP (myšlenka již 200 let stará)
  - Henderson (1949) – AM
    - nelineární postupy

Materiály určené pro studenty specializace „Genetika a šlechtění hospodářských zvířat“ pro předmět **Genetika ve šlechtění zvířat** (letní semestr 2006).

Dr. Ing. Tomáš Urban
ÚMFGZ – pracoviště genetiky MZLU v Brně
http://www.af.mendelu.cz/genetika/
urban@mendelu.cz

březen ’06