

Odhad genetických parametrov a plemenných hodnôt

Vzorové údaje 2.3

Vzorové údaje 2.4

Aplikácia programu SAS

Praktické použitie programu SAS (SAS Enterprise Guide)

1. úloha (SAS)

Výsledky úlohy 1

2. úloha (SAS)

Výsledky úlohy 2

Praktické použitie programov BLUPF90

3. úloha (BLUPF90)

Výsledky úlohy 3

4. úloha (BLUPF90)

Výsledky úlohy 4

Aplikácia programu programu R

Praktické použitie programu R (R Studio)

5. úloha (program R)

Výsledky úlohy 5

6. úloha (program R)

Výsledky úlohy 6

Odhad genetických parametrov a plemenných hodnôt

Odhad genetických parametrov a plemenných hodnôt vlastností hospodárskych zvierat vychádza z poznania, že celkový fenotypový prejav ľubovoľnej vlastnosti sa dá vyjadriť podľa základnej biologickej rovnice:

$$P = G + E$$

Pri hodnotení kvantitatívnych znakov nedokážeme priamo merať genotyp (G) ani celkový vplyv prostredia (E), ale pomerne presne zisťujeme hodnotu fenotypu (P), ktorá je ovplyvnená genotypom a prostredím.

Odhad plemennej hodnoty v skutočnosti predstavuje korekciu nameraného fenotypového prejavu vlastnosti o všetky známe negenetické činitele. Predpokladom odhadu plemennej hodnoty je aj poznanie, že namerané úžitkovosti hospodárskych zvierat sú navzájom korelované v rámci existujúcich príbuzenských vzťahov. Ďalším významným prvkom odhadu plemenných hodnôt je využívanie moderných výpočtových a štatistických metód v spojitosti s narastajúcou výkonnosťou výpočtovej techniky.

Plemenná hodnota môže byť definovaná ako **všeobecná plemenná hodnota** (vyjadruje najčastejšie aditívnu zložku, aditívny účinok génov), alebo ako **špecifická plemenná hodnota** (vyjadruje neaditívny účinok génov, dominantný resp. dominantný a epistatický účinok génov). V praxi je oveľa častejšie využívaná iba aditívna zložka, ktorá je základom odhadu väčšiny odhadovaných plemenných hodnôt.

Odhady plemenných hodnôt sa v prevažnej miere uskutočňujú pre kvantitatívne znaky (produkcia mlieka, produkcia mäsa, výkonnosť koní, telesné rozmery zvierat), ktoré sú podmienené polygénne. Kvantitatívny znak je charakterizovaný veľkým počtom génov a tým aj možných genotypov. Základným riešeným problémom kvantitatívnych znakov je poznanie aká časť premenlivosti (rozptylu) znaku je podmienená prostredím a aká časť je ovplyvnená geneticky.

Základná biologická rovnica $P = G + E$ sa dá preto vyjadriť pomocou rozptylov ako:

$$\begin{aligned}\sigma_P^2 &= \sigma_G^2 + \sigma_E^2 \quad \text{resp.} \\ \sigma_P^2 &= \sigma_A^2 + \sigma_D^2 + \sigma_I^2 + \sigma_E^2 \\ \sigma_P^2 &= \sigma_A^2 + \sigma_D^2 + \sigma_I^2 + \sigma_{PE}^2 + \sigma_{TE}^2\end{aligned}$$

- σ_P^2 - rozptyl fenotypový,
- σ_G^2 - rozptyl genetický,
- σ_E^2 - rozptyl spôsobený prostredím,
- σ_{PE}^2 - rozptyl trvalého (permanentného) efektu prostredia,
- σ_{TE}^2 - rozptyl dočasného efektu prostredia,
- σ_A^2 - rozptyl aditívny,
- σ_D^2 - rozptyl dominantný,
- σ_I^2 - rozptyl epistatický (interakcia medzi génmi)

Využitie a poznanie rôznych metód analýzy rozptylu je nevyhnutné pri správnej analýze premenlivosti konkrétnych kvantitatívnych vlastností hospodárskych zvierat a tým aj pri odhade genetických parametrov a plemenných hodnôt.

Všeobecne je možné rozdeliť najviac používané metódy odhadu plemenných hodnôt na dve skupiny: **metódy založené na teórii selekčných indexov** (Selection Index Theory, Hazel a Lusch, 1942, Hazel, 1943) a **metódy založené na princípe zmiešaných lineárnych modeloch** (metódy BLUP).

Teória selekčných indexov je v skutočnosti prvou metódou, ktorá sa využívala v praxi pri odhade plemenných hodnôt už od roku 1940. Metóda BLUP sa začala využívať oveľa neskôr, až s nástupom rozvoja výpočtovej techniky, hoci jej podstata bola odvodená v rovnakom období ako teória selekčných indexov. Za zakladateľa metódy BLUP sa považuje C.R. Henderson. Aj keď sa jedná o odlišné metódy hlavným a základným cieľom stále zostáva odhad plemennej hodnoty.

Ďalšou dôležitou spoločnou podmienkou obidvoch metód je nutnosť poznania hodnôt základných genetických parametrov jednotlivých vlastností a znakov zvierat.

Aktuálne najviac používanou metódou odhadu plemenných hodnôt hospodárskych zvierat na všetkých úrovniach výskumu a praxe je práve **metóda BLUP** (Best Linear Unbiased Prediction, najlepšia lineárna nestranná predpoveď). Metóda vychádza z teórie zmiešaných modelov pre súčasný odhad genetických efektov a efektov prostredia a úzko súvisí s metódou odhadu komponentov rozptylu v týchto modeloch (Henderson, 1953, 1954).

Zmiešané modely (modely s náhodnými a pevnými efektmi) sú často jedinou formou ako vyjadriť zložitý biologický dej a proces.

Metódy využívané pri odhade plemenných hodnôt hospodárskych zvierat Mrode (2014):

1. **BLUP jednorozmerné modely s jedným náhodným efektom**
(Animal model, Sire model, Reduced Animal model, Animal model with groups)
2. **BLUP modely s náhodnými efektmi prostredia**
(Repeatability model, Model with common environmental effects)
3. **BLUP viacrozmerné modely**
(Model with equal design matrices and no missing records, Model with canonical transformation, Model with equal design matrices with missing records, Model with Cholesky transformation, Model with unequal design matrices, Multivariate models with no environmental covariance, Model with different traits recorded on relatives, The Multi-trait across country evaluations model- MACE)
4. **Maternálne modely**
(Animal model for a maternal trait, Reduced animal model with maternal effects, Sire and Maternal Grand Sire maternal model, Multivariate maternal animal model)
5. **Sociálne interakčné modely**
(Animal model with Social Interaction effects, Model with partitioning evaluations from associative models, Model analysis with using correlated error structure)
6. **Modely analýzy dlhodobých časových údajov**
(Fixed regression models, Random regression model, Random regression model with spline function, Random regression model for maternal traits, Model with covariance functions, Model with equivalence of random regression model to the covariance function)
7. **Neadditívne animal modely**
(Animal model with dominance effect, Animal model with dominance and epistatic effect)
8. **Modely analýzy kategorických údajov**
(The threshold model, Model for joint analysis of quantitative and binary traits)
9. **Modely analýzy prežiteľnosti**
(Linear models, Random regression model for survival, Proportion hazard models, Non-parametric estimation of the survival function, Regression survival models, Mixed survival models, Group data survival model)

V rámci jednotlivých metód existuje množstvo prístupov a možností zabudovania genetických markerov a genomických informácií do odhadu plemenných hodnôt hospodárskych zvierat.

BLUP Animal model (bez zohľadnenia príbuzenských vzťahov medzi zvieratami)

Model môžeme zapísať v maticovej forme:

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

$\mathbf{y}$ stĺpcový vektor fenotypových pozorovaní (nameraná úžitkovosť)

$\mathbf{b}$ stĺpcový vektor neznámych pevných efektov

$\mathbf{a}$ stĺpcový vektor neznámych náhodných efektov (plemenné hodnoty), $N(0, \sigma_a^2)$

$\mathbf{X}, \mathbf{Z}$ matice vysvetľujúcich premenných priradujúce úžitkovosti pevným a náhodným efektom

$\mathbf{e}$ vektor náhodných chýb (rezíduum), $N(0, \sigma_e^2)$

Riešením uvedeného zápisu je súčasne získanie riešení pre pevné a náhodné efekty pomocou systému rovníc **MME** (Hendersonov Mixed Model Equations, model zmiešaných rovníc).

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + I_k \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad k = (1 - h^2)/h^2$$

Praktické riešenie je založené na riešení pomocou inverzie ľavej strany modelu so zmiešanými rovnicami, ktoré sa dá zapísať nasledovne:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + I_k \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Uvedený model nám odhadne jednotlivé úrovne neznámych pevných efektov $\mathbf{b}$ a neznámych náhodných efektov (plemenných hodnôt) $\mathbf{a}$. Pre riešenie veľkých problémov (veľké počty zvierat, veľké počty rovníc) nie je možné väčšinou použiť predchádzajúci postup pomocou inverzie matice koeficientov. Bolo vytvorených niekoľko iteračných postupov a metód, ktoré pomerne presne riešia uvedený problém v praxi.

BLUP Animal model (so zohľadnením príbuzenských vzťahov medzi zvieratami)

Zohľadnenie všetkých príbuzenských vzťahov medzi zvieratami je hlavnou prednosťou odhadu plemenných hodnôt na základe Animal modelu. Príbuzenské vzťahy sú vyjadrené pomocou matice príbuznosti ($\mathbf{A}$), ktorá je v skutočnosti maticou aditívnej genetickej podobnosti. Správnosť a pravdivosť rodokmeňových informácií zohráva v tomto prípade využitia Animal modelu veľmi dôležitú úlohu.

Základný model zmiešaných rovníc je preto upravený nasledovne:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}k \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Matica aditívnej genetickej podobnosti je v modeli použitá v inverznom ($\mathbf{A}^{-1}$) tvare.

Presnosť odhadovaných plemenných hodnôt

Animal model so zohľadnením príbuzenských vzťahov medzi zvieratami umožňuje súčasne pomocou riešenia sústavy rovníc vypočítať aj presnosť odhadovaných plemenných hodnôt zvierat

Model v tomto prípade vychádza z inverznej matice koeficientov:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}k \end{bmatrix}^{-1} = \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$$

C^{11} predstavuje časť vyjadrenia pevných efektov a C^{22} predstavuje časť vyjadrujúcu náhodné efekty. Predpoveď chyby variance (PEV, Prediction error variance) pre zviera i sa vypočíta:

$$PEV_i = c^{ii} \sigma_e^2$$

Presnosť odhadovanej plemennej hodnoty (r_{TI}) konkrétneho zvieraťa i je vyjadrená:

$$r_{TI}^2 = 1 - c^{ii}k \quad r_{TI} = \sqrt{1 - c^{ii}k}$$

Okrem predchádzajúcich Animal modelov existuje ešte veľa rôznych typov a modifikácií ďalších používaných modelov.

Zohľadnenie genomických informácií

Existuje niekoľko metód zohľadňovania genomických informácií pri odhade plemenných hodnôt. Väčšina týchto metód je vytvorená na základe resp. úpravy metódy BLUP. Všeobecne môžeme všetky tieto metódy rozdeliť na dve skupiny: prvá využíva pri zohľadnení genomických informácií **viackrokové prístupy** a druhá skupina **jednokrokové prístupy**.

Podstata viackrového prístupu:

- potreba získania východiskových údajov tzv. referenčnej populácie zvierat (najčastejšie preverených plemenníkov)
- odhad regresných koeficientov jednotlivých SNP v referenčnej populácii zvierat
- podkladom pre odhad regresných koeficientov sú odchýlky úžitkovosti dcér plemenníkov (DYD, Daughter Yield Deviation) a ich spoľahlivosti vyjadrené vo forme efektívnych príspevkov dcér (EDC, Effective Daughter Contribution)
- odhad priamych genomických plemenných hodnôt (GPH)
- vypočítanie genomicky upravených plemenných hodnôt (GEPH, kombinácia rodokmeňovej a priamej genomickej plemennej hodnoty pomocou indexu)
- odhadnuté regresné koeficienty rovnakých SNP sú následne využívané hlavne pri mladých genotypovaných plemenníkoch, ktorým je takto stanovená ich genomická plemenná hodnota

Podstata jednokrového prístupu:

- základným princípom je úprava matice aditívnej genetickej podobnosti maticou genomickej podobnosti (úprava spočíva v inverzii matice additívnej genetickej podobnosti pomocou genomickej matice príbuznosti)
- na samotnej metóde odhadu plemenných hodnôt sa okrem zohľadnenia genomickej matice podobnosti v podstate nič nemení
- výhodou je, že genomicky upravené plemenné hodnoty sú odhadované aj pre zvieratá, ktoré nie sú genotypované
- jednokrovový postup je v prípade veľkého počtu genotypovaných zvierat výpočtovo pomerne dosť náročný (súčasná výkonnosť výpočtovej techniky tento problém ale čiastočne eliminuje)

Genomická matica príbuznosti je zostavovaná na základe SNP informácií, ktoré prakticky popisujú celú genetickú informáciu jedinca (podľa hustoty konkrétneho genomického čipu). Ak vychádzame zo základného zápisu metódy BLUP so zohľadnením príbuzenských vzťahov, kde matica **A** predstavuje maticu aditívnej genetickej podobnosti, ktorá vznikla klasickým spôsobom z rodokmeňových informácií. Matica aditívnej genetickej podobnosti môže, ale nemusí presne vyjadrovať skutočnú genetickú podobnosť medzi jedincami (existuje totiž náhodný mendelistický výber pohlavných buniek, ktorého efekt je síce v populácii v priemere nulový, ale pri individuálnom jedincovi môže byť významný). Na základe uvedeného je možné spresniť maticu **A** genomickými („SNP“) informáciami.

Zaujímavosťou genomickej matice príbuznosti je to, že môže nadobúdať aj záporné hodnoty (od -1 do + 1), čo v prípade aditívnej matice genetickej podobnosti nie je možné (všetky prvky matice nadobúdajú hodnoty od 0 do 1). Záporné hodnoty genomickej matice sú v prípade ak jedinci majú v jednotlivých lokusoch opačné alely.

V maticovom zápise je zabudovanie genomickej matice zohľadnené v matici H:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + H^{-1}k \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Matica H sa vypočíta podľa nasledovného vzťahu:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

G^{-1} je inverzná genomická matica a A_{22}^{-1} je inverzná matica aditívnej príbuznosti genotypovaných zvierat vypočítaná na základe rodokmeňových informácií.

Redukovaný Animal model (reduced Animal model)

V modeli sú použité iba rovnice pre rodičov, ktorí majú potomstvo. Odhad plemenných hodnôt je z dôvodu menšieho počtu rovníc výpočtovo menej náročný oproti plnému modelu so všetkými zvieratami. Nevýhodou je ale, že absentujú plemenné hodnoty potomstva. Uvedený problém sa dá ale riešiť následným odhadom plemenných hodnôt potomkov z plemenných hodnôt rodičov (jedná sa v skutočnosti o výpočet tzv. rodokmeňových plemenných hodnôt potomstva, alebo len výpočet priemernej plemennej hodnoty rodičov pre každého potomka).

Veľmi špecifický prípadom v rámci animal modelov je existencia zvierat bez známych rodičov. Najčastejšie sa tento problém rieši pomocou **genetických skupín neznámych rodičov** (fantómové skupiny neznámych rodičov). Pri vytváraní genetických skupín neznámych rodičov ale musíme poznať časové obdobie, krajinu a tak isto aj približnú genetickú úroveň do ktorej je takéto zviera zaradené. Inak by sme sa mohli dopustiť pomerne veľkej chyby, ktorá je veľmi podobná ako nesprávna rodokmeňová informácia.

Sire model (otcovský model)

Je typickým jednoduchším využitím individuálne Animal modelu, pomocou ktorého odhadujeme plemenné hodnoty iba pre plemenníkov. Odhadovaná plemenná hodnota je v skutočnosti iba polovičnou plemennou hodnotou, pretože potomstvo má v hodnotení iba polovicu génov od otca. Matka sa vôbec neberie do úvahy.

Sire-maternal- grandsire model

Pri odhade je zohľadňovaný otec a otec matky. Uvedený model sa v minulosti intenzívne používal pri niektorých ukazovateľoch hospodárskych zvierat.

Modely s maternálnym efektom

Existujú znaky a vlastnosti hospodárskych zvierat, ktoré sú výrazne ovplyvnené maternálnym efektom. Takýmto modelom hovoríme modely s maternálnym efektom. Jedná sa hlavne o odhad plemenných hodnôt znakov, kde matka pomerne výrazne vplýva na svojho potomka počas gravidity, čo je typické iba pre samice cicavcov. Modely s maternálnymi efektmi sú typické pri odhade plemenných hodnôt ukazovateľov plodnosti a obtiažnosti pôrodov.

Model sa dá zapísať ako:

$$y = Xb + Z_1a + Z_2m + Z_3p + e,$$

y - hodnotený ukazovateľ, a - aditívny genetický efekt, m - maternálny genetický efekt, p - permanentný permanentný efekt prostredia, e - rezíduum

Viacznakové Animal modely

Predstavujú metódy pri ktorých sú odhadované plemenné hodnoty viacerých znakov, alebo vlastností súčasne. Pri existencii vzťahov medzi znakmi vyjadrených najčastejšie vo forme genetických korelácií sa odhadnuté plemenné hodnoty navzájom ovplyvňujú a spoľahlivosť ich odhadu je vyššia. Problémom môže byť ale v prípade veľkého počtu zvierat a veľkého počtu ukazovateľov praktické výpočtové a ekonomické zvládnutie viacznakového prístupu. Uvedené sa ale v poslednom období s narastajúcou výkonnosťou výpočtovej techniky mení a stáva sa čoraz menej limitujúcim. Problémy sú skôr pri odhade základných genetických parametrov, ktoré sú nevyhnutné pre praktické odhady plemenných hodnôt a výpočtová náročnosť je výrazne vyššia ako samotný odhad plemenných hodnôt. V prípade viacznakového prístupu nie je často krát možné odhadovať genetické parametre na celých populáciách zvierat. Z tohto dôvodu sa vyberajú reprezentatívne (početne menšie) vzorky zvierat a údajov na ktorých sa dajú zložit a časovo náročné výpočty uskutočniť.

Viacznakový model s dvoma znakmi môže byť v maticovom zápise uvedený nasledovne:

$$Y_i = X_i b_i + Z_i a_i + e_i$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Výhodou viacznakovým modelov je to, že jedince nemusia mať zákonite všetky hodnotené ukazovatele. Absencia niektorých ukazovateľov pri niektorých jedincoch je možná a pritom plemenné hodnoty sú odhadnuté pre všetky ukazovatele a všetky jedince (umožňuje to zohľadnenie vzťahov, genetických korelácií medzi ukazovateľmi).

Nelineárne modely

Odhad plemenných hodnôt na základe nelineárnych modelov nie je až tak typický, ale existujú ukazovatele, pri ktorých je výhodnejšie použiť práve nelineárne prístupy. Najčastejšie sa nelineárnymi modelmi hodnotia niektoré nepriame úžitkové vlastnosti. S priamymi úžitkovými vlastnosťami sú známe nelineárne systémy hodnotenia rastu a vývoja hospodárskych zvierat. Rast zvierat sa všeobecne správa nelineárne, hlavne pokiaľ sledujeme rast dlhšie časové obdobie.

V časovo krátkom období sa rast zvierat spravidla správa lineárne. V rámci nepriamych úžitkových vlastností hospodárskych zvierat má v poslednom období stále väčší význam v modernom šľachtení hodnotenie **dlhovekosti a prežiteľnosti** zvierat. Dlhovekosť sa v skutočnosti javí ako súhrnný ukazovateľ, ktorý vo svojej podstate vyjadruje všetky vlastnosti zvierat.

Coxov model

Coxov proporcionálny regresný model sa vo veľkej miere využíva pre analýzu údajov o prežiteľnosti. Patrí do skupiny semiparametrických modelov, pri ktorých nie je známy tvar základnej hazardnej funkcie, to znamená, že funkcia nevykazuje žiadne známe rozdelenie. Coxov model má v porovnaní s inými neparametrickými metódami širšie využitie, pretože predstavuje jednoduchú metódu na interpretáciu vzťahu hazardnej funkcie k nezávislým premenným (faktorom ovplyvňujúcim dlhovekosť). Vzťah medzi hazardnou funkciou a faktormi je nelineárny, pomer rizika vyradenia medzi dvomi pozorovaniami je konštantný pre časovo nezávislé premenné. Coxov model je najpopulárnejší model vhodný pre Analýzu prežiteľnosti

Weibullov model

Základná hazardná funkcia modelu má známy parametrický tvar, alebo sa aspoň trochu približuje k nejakému teoretickému rozdeleniu, odhad regresných koeficientov a základnej hazardnej funkcie je vo všeobecnosti jednoduchší a menej náročný na čas. Najvhodnejšou parametrickou formou základnej hazardnej funkcie je Weibullovo rozdelenie. Vo Weibullovom regresnom modeli je základná hazardná funkcia charakterizovaná dvomi parametrami λ a p , ktoré definujú rozsah a tvar základnej hazardnej funkcie. Ak $p = 1$, riziko je konštantné, ak je $p > 1$, riziko narastá, ak je $p < 1$ riziko klesá. Aproximáciou základnej hazardnej funkcie Weibullovým rozdelením sa značne zjednodušujú ďalšie analýzy. Výhodou Weibullovho modelu je jeho rozšírenie na zmiešané modely, ktoré môžu zahŕňať korelované náhodné genetické efekty, ako napríklad rodokmeňové vzťahy medzi otcami. Takéto zmiešané modely sú uvádzané pod názvom "slabé" (frailty). Slabé modely sa pokúšajú odhadnúť nepozorovateľné rôznorodé náhodné vplyvy, ktoré sa vyskytujú počas života. Základnou myšlienkou slabých modelov je zaradiť do rizika náhodný parameter, ktorý vysvetľuje časť rozptylu, ktorá nemôže byť vypočítaná v modeli s fixnými efektmi. Pri náhodných faktoroch predpokladáme, že majú gamma rozdelenie.

Vzorové údaje 2.3 (zdroj: Databáza údajov USDA)

Vzorové údaje 2.3 pre odhad genetických parametrov a odhad plemenných hodnôt tvoria údaje z oficiálneho systému genetického mliekovej úžitkovosti hodnotenia v USA (<http://www.aipl.arsusda.gov/>), ktoré publikoval Vazquez et al. (2010). Súbor tvorí 3,397 laktácií, 1,359 kráv holštajského plemena na prvej až piatej laktácii, počet otcov (plemenníkov) 38 a celkový počet stád 57.

Súbor údajov **Milk01.xlsx** obsahuje nasledovné premenné:

id	číslo zvierat'a
lact	poradie laktácie
herd	stádo
sire	plemenník (otec)
dim	počet dní laktácie
milk	produkcia mlieka (kg)
fat	produkcia tuku (kg)
prot	produkcia bielkovín (kg)
scs	skóre za somatické bunky
sdmilk	štandardizované mlieko ($s_x=0$)

Vzor súboru údajov (10 záznamov)

id	lact	herd	sire	dim	milk	fat	prot	scs	sdMilk
6489	1	89	2	286	18420	784	607	2.01	4.1187558
6489	2	89	2	305	21592	954	644	1.64	4.82802254
6489	3	89	2	203	15834	779	490	1.86	3.54052005
6490	1	89	2	281	20683	785	610	2.71	4.62476798
6490	2	89	2	277	20050	841	582	2.83	4.48322767
6490	3	89	2	289	21891	884	650	2.63	4.89487965
6491	1	89	2	305	23646	808	729	5.11	5.28730182
6492	1	89	2	281	19773	767	622	2.35	4.42128982
6492	2	89	2	270	20738	872	639	2.29	4.63706611
6492	3	89	2	256	20510	829	688	1.46	4.58608477

Vzorové údaje 2.4 (zdroj: Henderson, 1984)

Vzorové údaje predstavujú príklad z oblasti šľachtenia hospodárskych zvierat. Údaje obsahujú dve hodnotené vlastnosti troch zvierat (druhá vlastnosť pri treťom zvierati chýba). Vzorové údaje sú nasledovné:

Trait	ukazovateľ
Animal	jedinec
Y	hodnota ukazovateľa

Trait	Animal	Y
1	1	6
1	2	8
1	3	7
2	1	9
2	2	5
2	3	.

Matica G (aditívna genetická podobnosť)							
Row	Col1	Col2	Col3	Col4	Col5	Col6	
1	2	1	1	2	1	.	
2	1	2	0.5	1	2	0.3	
3	1	0.5	2	1	0.5	2	
4	2	1	1	3	1.5	1.5	
5	1	2	0.5	1.5	3	0.75	
6	1	0.5	2	1.5	0.75	3	

Aplikácia programu SAS – procedúry GLM a MIXED

Pomocou procedúry MIXED je možné uskutočňovať anlyzy na základe zmiešaných lineárnych modelov. Zmiešaný lineárny model je zovšeobecnením štandardného lineárneho modelu používaného v postupe GLM (Generalized Linear Model, zovšeobecnený lineárny model). Zmiešané lineárne modely nám poskytujú pomerne veľkú flexibilitu modelovania biologických procesov a javov spolu s odhadom veľkého množstva parametrov geneticko-biologického charakteru (rozptyly, variačno-kovariačné parametre, koeficienty dedivosti, fenotypové a genetické korelácie). Modely môžu obsahovať dva typy základných efektov: fixné a náhodné, ktoré môžu byť kombinované resp. zoskupované rôznym spôsobom.

Pevné efekty sa spájajú so známymi vysvetľujúcimi premennými ako to je v štandardnom lineárnom modeli. Tieto premenné môžu byť buď kvalitatívne (ako pri tradičnej analýze rozptylu) alebo kvantitatívne (ako pri štandardnej lineárnej regresii). Oproti tomu náhodné efekty, ktoré vytvárajú kovariančné parametre odlišuje zmiešaný lineárny model od štandardného lineárneho modelu. Náhodné efekty sa spájajú najčastejšie s genetickými (rodokmeňovými) informáciami.

Dôležitým predpokladom použitia zmiešaných modelov je podobne ako pri väčšine všeobecných analýz rozptylu podmienka zachovania normality údajov.

Štandardný lineárny model (GLM):

$$y = Xb + e$$

y - vektor pozorovaní

b - neznámy vektor parametrov pevných efektov

X - známa matica vysvetľujúcich premenných

e - neznáma náhodná chyba (rezíduum)

Predpokladom je, že náhodné (reziduálne) chyby sú nezávislé a majú normálne rozdelenie s nulovou strednou hodnotou a rozptylom.

Zmiešaný lineárny model (MIXED):

$$y = Xb + Za + e$$

y - vektor pozorovaní

b - neznámy vektor parametrov pevných efektov

a - neznámy vektor parametrov náhodných efektov

X - známa matica vysvetľujúcich premenných

Z - známa matica vysvetľujúcich premenných

e - neznáma náhodná chyba

Praktické použitie programu SAS (SAS Enterprise Guide)

1. úloha (SAS)

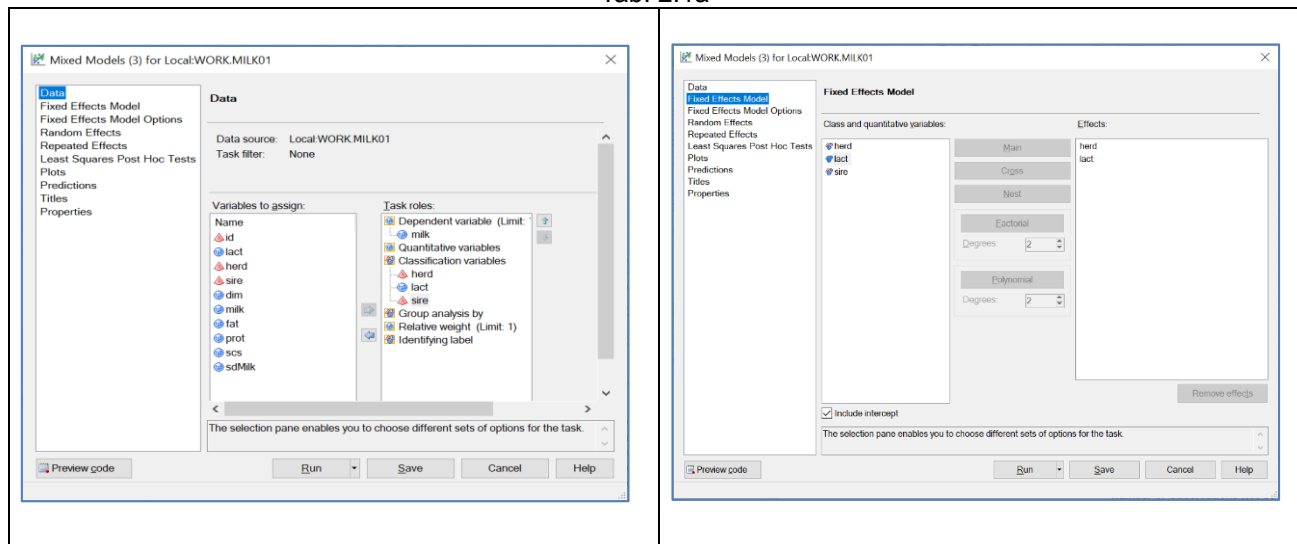
Odhadnite pomocou zmiešaného lineárneho modelu základné genetické parametre ukazovateľov mliekovej úžitkovosti (vzorové údaje 2.3).

Model 1: **milk** = herd + lact + sire + e

Herd (stádo) a lact (poradie laktácie) sú pevné efekty, sire (plemenník) a rezíduum (e) sú náhodné efekty.

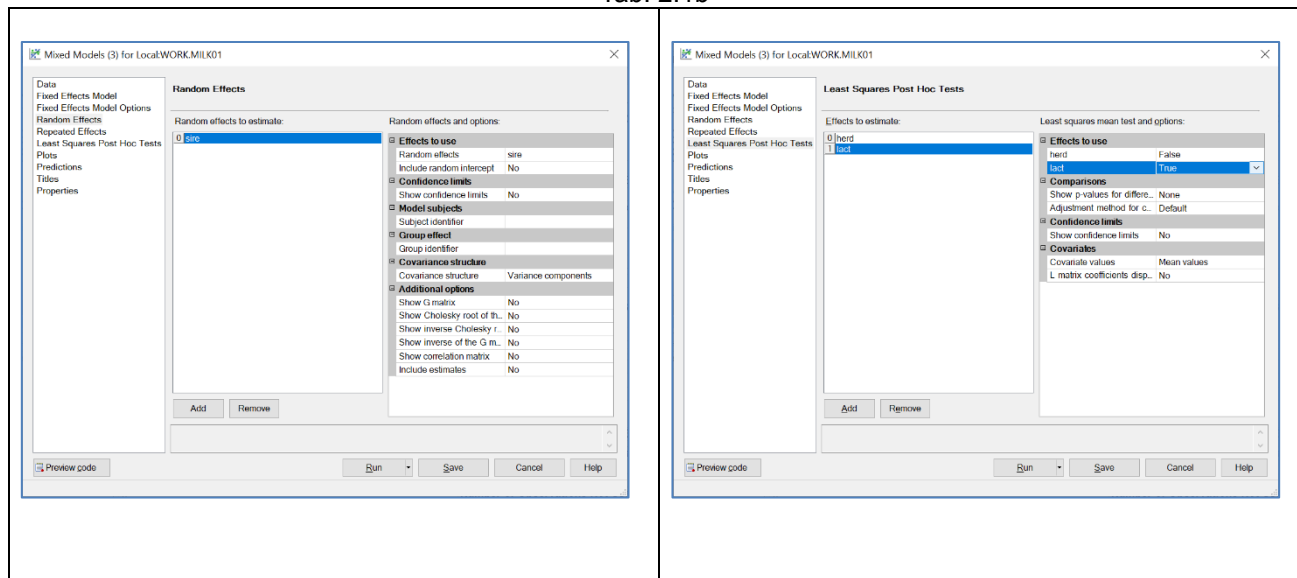
Zmiešaný lineárny model má nasledovné základné nastavenia a parametre pre pevné a náhodné efekty:

Tab. 2.1a



Náhodný efekt je definovaný s možnosťou odhadu riešení pre všetky jeho úrovne. Uvedené v skutočnosti predstavuje odhad riešenia produkcie mlieka pre jednotlivých plemenníkov (otcov) hodnotených kráv, ktorý môže byť považovaný za jednoduchú východiskovú formu ich plemennej hodnoty.

Tab. 2.1b



Model môžeme zapísať aj priamo pomocou zdrojového kódu, ktorý sa môže veľmi jednoducho modifikovať pri zostavovaní rôznych typov zmiešaných modelov (Tabuľka 2.2c).

Tab. 2.1c

```
ODS GRAPHICS ON;
PROC MIXED DATA = "c:\Biometrics\Data\milk01.sas7bdat"
  PLOTS (ONLY)=ALL
  METHOD=REML;
CLASS herd lact sire;
MODEL milk=herd lact;
RANDOM sire / CL
  ALPHA=0.05 TYPE=VC SOLUTION;
LSMEANS herd /;
LSMEANS lact /;
RUN;
ODS GRAPHICS OFF;
```

Výsledky úlohy 1

V tabuľkách 2.2a, 2.2.b a 2.2c uvádzame základné informácie o použitom zmiešanom modeli spolu s definovanou kovariančnou štruktúrou a samotnou metódou odhadu. Zobrazené sú jednotlivé úrovne definovaných pevných a náhodných efektov a celkové rozmery odhadovaných parametrov.

Tab. 2.2a

Model Information	
Dependent Variable	milk
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Tab. 2.2b

Class Level Information		
Class	Levels	Values
herd	57	1 100 101 103 104 105 106 107 108 109 110 13 14 17 18 2 22 23 26 27 29 30 34 36 37 38 4 41 45 46 48 49 5 52 53 54 55 59 6 60 62 64 66 68 69 7 70 75 77 8 89 90 95 96 97 98 99
lact	5	1 2 3 4 5
sire	38	1 2 3 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 4

Tab. 2.2c

Dimensions	
Covariance Parameters	2
Columns in X	63
Columns in Z	38
Subjects	1
Max Obs per Subject	3397

Na základe odhadnutých kovariačných parametrov je možné vypočítať koeficient heritability pre hodnotený ukazovateľ produkcia mlieka (kg).

Tab. 2.2d

Covariance Parameter Estimates	
Cov Parm	Estimate

sire	441019
Residual	15257503

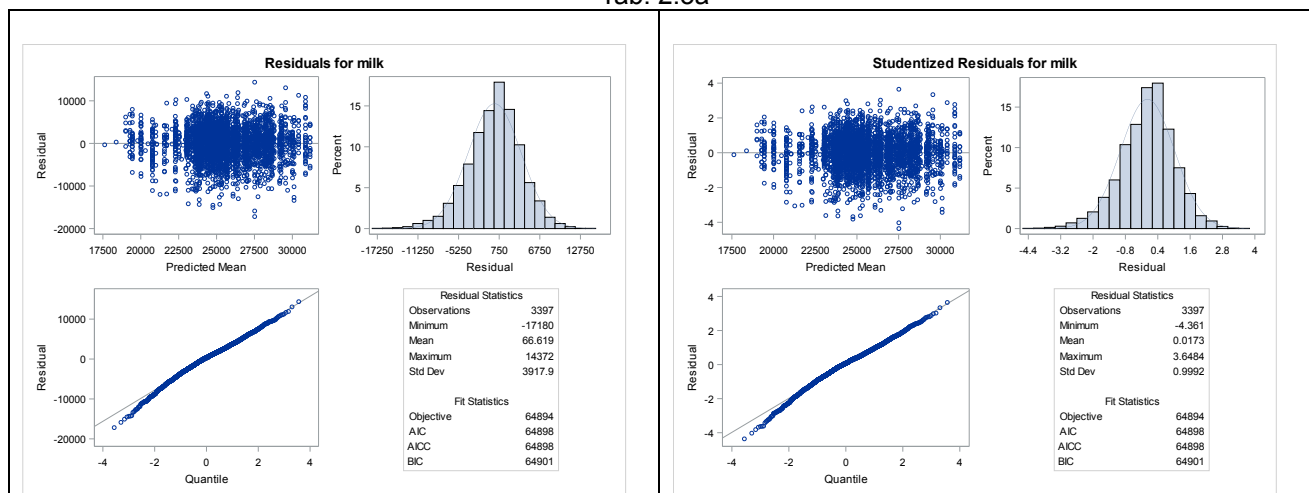
$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} = 4 * \sigma_S^2 / (\sigma_S^2 + \sigma_e^2)$$

σ_A^2 – rozptyl aditívny, σ_P^2 – rozptyl fenotypový, σ_S^2 – rozptyl plemenníci, σ_e^2 – rozptyl reziduálny

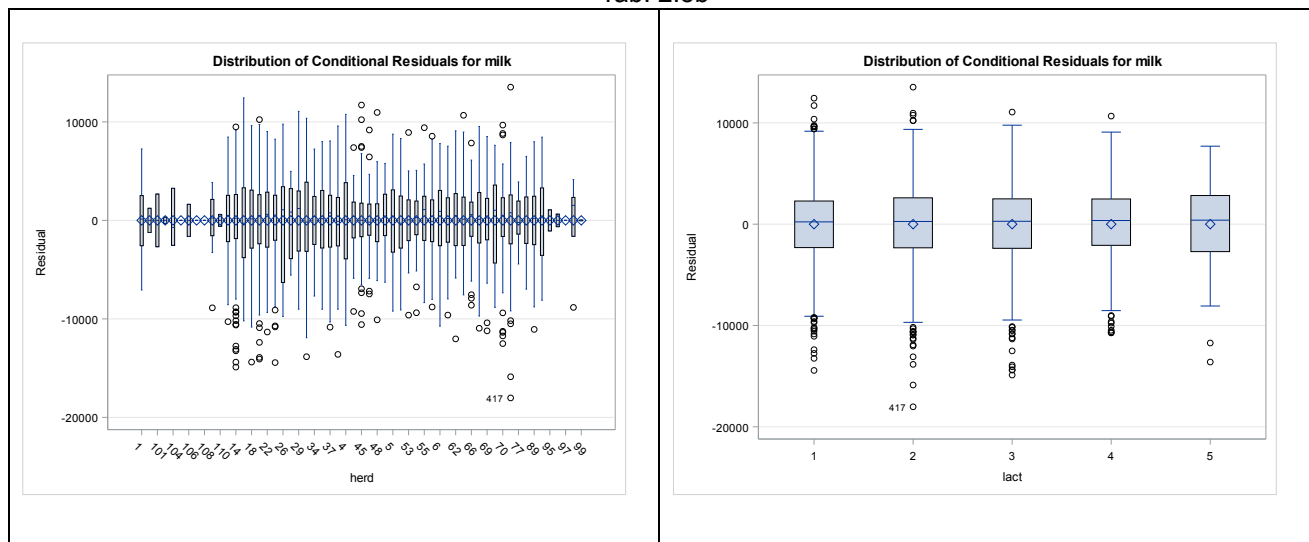
Odhadnutý koeficient heritability (h^2) dosiahol hodnotu 0,1124.

V tabuľkách 2.3a a 2.3b uvádzame celkovú analýzu rezíduí a jej distribúciu podľa použitých pevných efektov.

Tab. 2.3a



Tab. 2.3b



V tabuľke 2.4a uvádzame riešenia náhodného efektu sire (prvých 10 plemenníkov), ktoré môžu byť následne vyjadrené ako odhadnuté plemenné hodnoty konkrétnych jedincov.

Tab. 2.4a

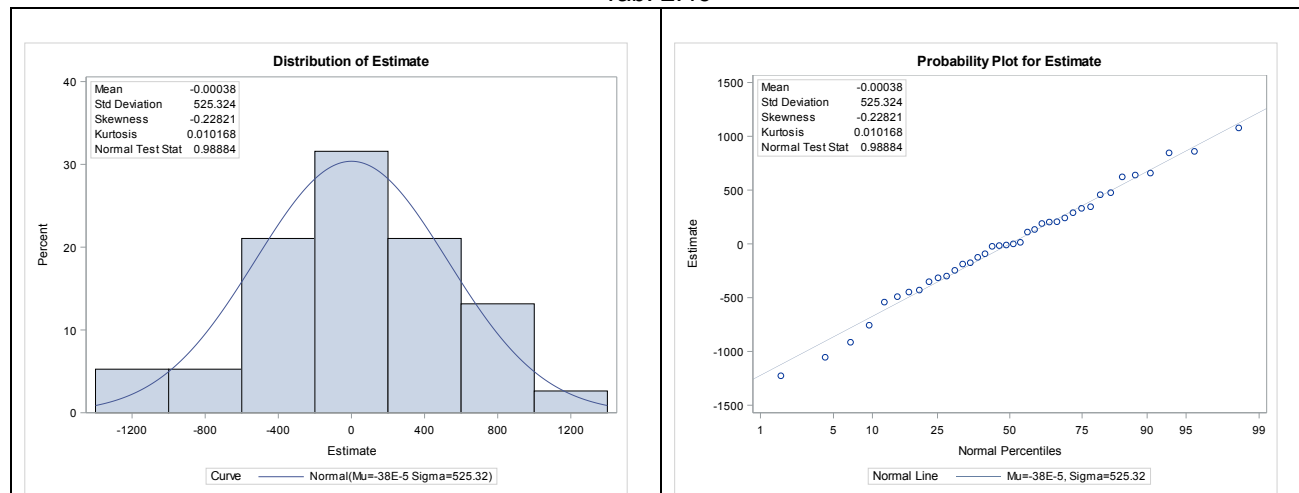
Solution for Random Effects									
Effect	sire	Estimate	Std Err Pred	DF	t Value	Pr > t	Alpha	Lower	Upper
sire	1	-315.32	430.67	3300	-0.73	0.4641	0.05	-1159.73	529.10
sire	2	-193E-15	664.09	3300	-0.00	1.0000	0.05	-1302.07	1302.07
sire	3	189.37	343.31	3300	0.55	0.5813	0.05	-483.74	862.49
sire	319	846.07	378.64	3300	2.23	0.0255	0.05	103.68	1588.47
sire	320	658.41	404.29	3300	1.63	0.1035	0.05	-134.27	1451.09
sire	321	-10.5183	389.07	3300	-0.03	0.9784	0.05	-773.37	752.33
sire	322	-91.8342	382.03	3300	-0.24	0.8100	0.05	-840.87	657.21
sire	323	622.92	385.16	3300	1.62	0.1059	0.05	-132.26	1378.11
sire	324	-187.41	392.51	3300	-0.48	0.6331	0.05	-957.00	582.17
sire	325	14.4134	384.06	3300	0.04	0.9701	0.05	-738.60	767.43

Uskutočnená distribučná analýza odhadnutých riešení plemenníkov potvrdila ich ideálne normálne rozdelenie početnosti, čo je zárukou ich efektívneho využitia v rámci konkrétnej selekcie zvierat.

Tab. 2.4b

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.98884	Pr < W	0.9640
Kolmogorov-Smirnov	D	0.061679	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.018833	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.144531	Pr > A-Sq	>0.2500

Tab. 2.4c



Zdrojový kód zmiešaného modelu s náhodným efektom jedinca (id jedinca, krava) namiesto náhodného efektu sire (Tabuľka 2.5a).

Tab. 2.5a

```
Proc mixed data = "c:\Biometrics\Data\milk01.sas7bdat"
  plots(only)=all
  method=REML;
  class Herd Lact Id;
  model Milk=Herd Lact;
  random id / cl
    alpha=0.05 type=vc solution;
  lsmeans Herd /;
  lsmeans Lact /;
run;
```

Odhadnutý rozptyl náhodných efektov uvádzaný v tabuľke 2.5b je rozdielny oproti tomu, ako to bolo v prípade náhodného efektu plemenníka.

Tab. 2.5b

Covariance Parameter Estimates	
Cov Parm	Estimate
id	5493807
Residual	10402547

Odhadnutý koeficient heritability (h^2) dosiahol hodnotu 0,3456

Zdrojový kód zmiešaného modelu s dvoma náhodnými efektmi (sire a jedinec, tabuľka 2.5c).

Tab. 2.6a

```
Proc mixed data = "c:\Biometrics\Data\milk01.sas7bdat"
  plots(only)=all
  method=REML;
  class Herd Lact Id Sire;
  model Milk=Herd Lact;
  random Sire / cl
    alpha=0.05 type=vc solution;
  random id / cl
    alpha=0.05 type=vc solution;
  lsmeans Herd /;
  lsmeans Lact /;
run;
```

Tab. 2.6b

Covariance Parameter Estimates	
Cov Parm	Estimate
sire	301690
id	5278613
Residual	10397877

Odhadnutý koeficient heritability (h^2) dosiahol hodnotu 0,3841. Posledný použitý model bol na základe AIC (Akaike Information Criterion) zistený ako najpresnejší.

2. úloha (SAS)

Na základe lineárneho zmiešaného modelu (PROC MIXED) odhadnite plemenné hodnoty všetkých zvierat pre obe sledované vlastnosti (vzorové údaje 2.4). Využitie známej matice aditívnej genetickej podobnosti (matica G, 3 zvieratá x 2 vlastnosti, 6x6).

Zdrojový kód zmiešaného modelu so zohľadnením matice G:

Tab. 2.7a

```
Proc mixed data=h mmeq mmeqsol;
  class Trait Animal;
  model Y = Trait / noint s outp=predicted;
  random Trait*Animal / type=un gdata=g g gi s;
  repeated / type=un sub=Animal r ri;
  parms (4) (1) (5) / noiter;
run;
proc print data=predicted;
run;
```

Výsledky úlohy 2

V tabuľkách 2.8a, 2.8.b a 2.8c uvádzame základné súhrnné informácie výsledkov analýzy.

Tab. 2.8a

Model Information	
Dependent Variable	Y
Covariance Structure	Unstructured
Subject Effect	Animal
Estimation Method	REML
Residual Variance Method	None
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Tab. 2.8b

Class Level Information		
Class	Levels	Values
Trait	2	1 2
Animal	3	1 2 3

Tab. 2.8c

Dimensions	
Covariance Parameters	3
Columns in X	2
Columns in Z	6
Subjects	1
Max Obs per Subject	5

Analýzou sme získali riešenia pre pevné a náhodné efekty, ktoré sú v poslednom stĺpci (Tabuľka 2.9a).

Tab. 2.9a

Mixed Model Equations Solution												
Row	Effect	Trait	Animal	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	Trait	1		2.5508	1.5685	-1.3047	-1.1775	-1.1701	-1.3002	-1.1821	-1.1678	6.9909
2	Trait	2		1.5685	4.5539	-1.4112	-1.3534	-0.9410	-2.1592	-2.1055	-1.3149	6.9959
3	Trait*Animal	1	1	-1.3047	-1.4112	1.8282	1.0652	1.0206	1.8010	1.0925	1.0070	0.05450
4	Trait*Animal	1	2	-1.1775	-1.3534	1.0652	1.7589	0.7085	1.0900	1.7341	0.7209	-0.04955
5	Trait*Animal	1	3	-1.1701	-0.9410	1.0206	0.7085	1.7812	1.0095	0.7197	1.7756	0.02230
6	Trait*Animal	2	1	-1.3002	-2.1592	1.8010	1.0900	1.0095	2.7518	1.6392	1.4849	0.2651
7	Trait*Animal	2	2	-1.1821	-2.1055	1.0925	1.7341	0.7197	1.6392	2.6874	0.9930	-0.2601
8	Trait*Animal	2	3	-1.1678	-1.3149	1.0070	0.7209	1.7756	1.4849	0.9930	2.7645	0.1276

Podrobné výsledky riešení pevných a náhodných efektov, spolu so štandardnými chybami a štatistickými preukaznosťami odhadov, uvádzame v tabuľkách 2.9b a 2.9c.

Tab. 2.9b

Solution for Fixed Effects						
Effect	Trait	Estimate	Standard Error	DF	t Value	Pr > t
Trait	1	6.9909	1.5971	3	4.38	0.0221
Trait	2	6.9959	2.1340	3	3.28	0.0465

Odhady pre tieto dva znaky sú približne rovnaké, ale štandardná chyba druhého znaku je väčšia z dôvodu chýbajúceho jedného pozorovania.

Tab. 2.9c

Solution for Random Effects							
Effect	Trait	Animal	Estimate	Std Err Pred	DF	t Value	Pr > t
Trait*Animal	1	1	0.05450	1.3521	0	0.04	.
Trait*Animal	1	2	-0.04955	1.3262	0	-0.04	.
Trait*Animal	1	3	0.02230	1.3346	0	0.02	.
Trait*Animal	2	1	0.2651	1.6589	0	0.16	.
Trait*Animal	2	2	-0.2601	1.6393	0	-0.16	.
Trait*Animal	2	3	0.1276	1.6627	0	0.08	.

Pri riešení náhodných efektov sú uvádzané najlepšie lineárne nestranné predpovede (BLUP, plemenné hodnoty) oboch vlastností pre všetky tri zvieratá. Hodnoty parametrov spoľahlivosti (Pr) chýbajú, pretože štandardná metóda vedie k nulovým stupňom voľnosti pre testy parametrov náhodných efektov.

Na základe testu štatistickej preukaznosti dvoch hodnotených ukazovateľov (tabuľka 2.9d) bolo potvrdené, že dve analyzované vlastnosti sú rozdielne (potvrdená štatistická preukaznosť na hladine 0,05).

Tab. 2.9d

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Trait	2	3	10.59	0.0437

V tabuľke 2.9e uvádzame finálne odhadnuté predpovede (Pred) na základe odhadov jednotlivých vlastností a odhadov plemenných hodnôt. Odhadnuté predpovede môžu byť použité pri priamom hodnotení zvierat podľa jednotlivých znakov.

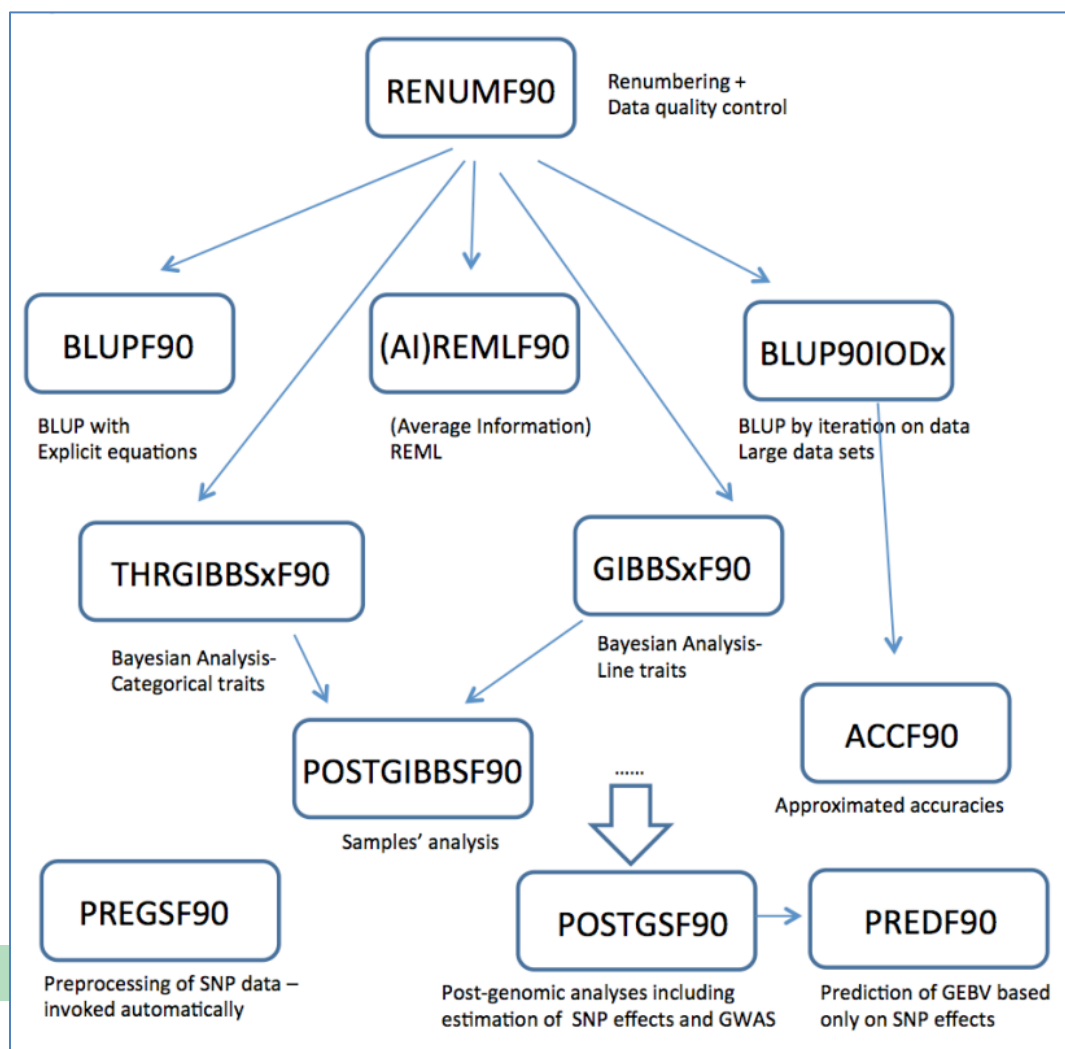
Tab. 2.9e

Obs	Trait	Animal	Y	Pred	StdErrPred	DF	Alpha	Lower	Upper	Resid
1	1	1	6	7.04542	1.33027	0	0.05	.	.	-1.04542
2	1	2	8	6.94137	1.39806	0	0.05	.	.	1.05863
3	1	3	7	7.01321	1.41129	0	0.05	.	.	-0.01321
4	2	1	9	7.26094	1.72839	0	0.05	.	.	1.73906
5	2	2	5	6.73576	1.74077	0	0.05	.	.	-1.73576
6	2	3	.	7.12015	2.99088	0	0.05	.	.	.

Predpovedané hodnoty ale nie sú predikciami budúcich záznamov, pretože neobsahujú zložku zodpovedajúcu budúcej chybe pozorovaní.

Aplikácia programov BLUPF90

BLUPF90 predstavuje komplexnú skupinu programov pre výpočty a analýzy založené na princípe zmiešaných lineárnych modelov so zameraním na genetické hodnotenie a šľachtenie hospodárskych zvierat. Programy využívajú jednotnú formu vstupných údajov, ktorá je využiteľná vo všetkých programoch. Využívajú niekoľko metód analýz rozptylu a metód odhadu genetických parametrov a plemenných hodnôt, ktoré je možné aplikovať aj na pomerne rozsiahlych údajových a rodokmeňových súboroch. Samostatnú skupinu tvoria tzv. prahové (thresholds) modely využívané v rámci analýz kategoriálnych údajov. Pomocou programov môžeme odhadovať presnosti a spoľahlivosti jednotlivých výpočtov. Programy umožňujú pracovať, okrem klasických rodokmeňových informácií, aj s genomickými informáciami (SNP informácie). Ich využitie je preto možné aj pri celogenómových asociačných štúdiách (GWAS). Programy sú voľne využiteľné v rámci vzdelávania a výskumu. V nasledujúcom obrázku uvádzame nadväznosť jednotlivých vybraných programov (Misztal et al, 2018).



Obr. 1 (<http://nce.ads.uga.edu/wiki/doku.php>)

Praktické použitie programov BLUPF90

Programy využívajú dva základné vstupné textové súbory: údajový súbor a rodokmeňový súbor. Jednotlivé položky v súboroch sú oddelené minimálne jednou medzerou. V programoch BLUPF90 dátové a rodokmeňové súbory môžu obsahovať iba číselné hodnoty (celé čísla, alebo reálne čísla). Označenie skupiny efektov musí byť tiež celé číslo začínajúce hodnotou 1. Súbory originálnych údajov zvyčajne obsahujú znaky abecedy, alebo symboly. Typické to je pre označenie zvierat, alebo označenie skupín zvierat. Z uvedeného dôvodu je nutné uskutočniť prekódovanie (prečíslovanie) originálnych údajov pomocou špeciálneho programu, ktorý je na tieto účely vytvorený (program RENUMF90). Program RENUMF90 podľa zvoleného nastavenia modelu prečísľuje údajové a rodokmeňové súbory, uskutoční prvotnú analýzu údajov a vytvorí základný parametrický súbor podľa navrhnutého modelu hodnotenia.

Tab. 2.10a Příklad originálních souborů (Masuda, 2019)

Údajový súbor:	Rodokmeňový súbor:
ID006 A 1 1.0 3.0	ID001 0 0
ID009 A 2 1.0 2.0	ID002 0 0
ID012 A 1 2.0 4.0	ID003 0 0
ID007 B 2 2.0 6.0	ID004 0 0
ID010 B 1 1.0 3.0	ID005 0 0
ID013 B 2 2.0 6.0	ID006 0 0
ID008 C 1 2.0 6.0	ID007 ID002 ID005
ID011 C 2 1.0 6.0	ID008 ID001 ID004
ID014 C 1 1.0 8.0	ID009 ID002 ID003
ID015 C 2 2.0 4.0	ID010 ID007 ID006
	ID011 ID007 ID004
	ID012 ID011 ID008
	ID013 ID011 ID010
	ID014 ID009 ID013
	ID015 ID011 ID010

Tab. 2.10b Příklad prekódovaných souborů (Masuda, 2019)

Údajový súbor (data.txt):	Rodokmeňový súbor (pedigree.txt):
3.0 1 1 1.0 6	1 7 5 1 0 2 1 0 0 ID015
2.0 1 2 1.0 9	13 0 0 3 0 0 0 0 2 ID004
4.0 1 1 2.0 12	11 0 0 3 0 0 0 0 1 ID005
6.0 2 2 2.0 7	2 0 0 3 0 0 1 0 1 ID006
3.0 2 1 1.0 10	3 12 11 1 0 2 1 2 0 ID007
6.0 2 2 2.0 13	4 14 13 1 0 2 1 0 1 ID008
6.0 3 1 2.0 8	5 3 2 1 0 2 1 0 2 ID010
6.0 3 2 1.0 11	6 12 15 1 0 2 1 1 0 ID009
8.0 3 1 1.0 14	7 3 13 1 0 2 1 3 0 ID011
4.0 3 2 2.0 15	8 7 4 1 0 2 1 0 0 ID012

```

14 0 0 3 0 0 0 1 0 ID001
9 7 5 1 0 2 1 0 1 ID013
12 0 0 3 0 0 0 2 0 ID002
10 6 9 1 0 2 1 0 0 ID014
15 0 0 3 0 0 0 0 1 ID003

```

3. úloha (BLUPF90)

Jednoduchý zmiešaný model

Modelová rovnica: $y_{ijk} = A_i + S_j + \beta x_{ijk} + e_{ijk}$,

kde y_{ijk} je hodnotený ukazovateľ, A_i je pevný efekt, S_j je náhodný efekt, βx_{ijk} je pevný regresný efekt a e_{ijk} je náhodný reziduálny efekt.

Animal model

Modelová rovnica: $y_{ijk} = A_i + S_j + \beta x_k + u_k + e_{ijk}$,

kde y_{ijk} je hodnotený ukazovateľ, A_i je pevný efekt, S_j je pevný efekt, βx_k je pevný regresný efekt, u_k je náhodný aditívny genetický efekt a e_{ijk} je náhodný reziduálny efekt.

Tab. 2.10c Parametrické súbory

Jednoduchý zmiešaný model	Animal model
DATAFILE data.txt	DATAFILE data.txt
NUMBER_OF_TRAITS	NUMBER_OF_TRAITS
1	1
NUMBER_OF_EFFECTS	NUMBER_OF_EFFECTS
3	4
OBSERVATION(S)	OBSERVATION(S)
1	1
WEIGHT(S)	WEIGHT(S)
EFFECTS:	EFFECTS:
2 3 cross	2 3 cross
3 2 cross	3 2 cross
4 1 cov	4 1 cov
RANDOM_RESIDUAL VALUES	5 15 cross
2.0	RANDOM_RESIDUAL VALUES
RANDOM_GROUP	2.0
2	RANDOM_GROUP
RANDOM_TYPE	4
diagonal	RANDOM_TYPE
FILE	add_animal
(CO)VARIANCES	FILE
1.0	pedigree.txt
	(CO)VARIANCES
	0.5

Výsledky úlohy 3

4. úloha (BLUPF90)

Výsledky úlohy 4

Odhad rozptylov a kovariancií

1 Premenné a distribúcie

Náhodné premenné sú vzorky z populácie s daným súborom parametrov populácie.

Náhodné premenné môžu byť diskrétné, s obmedzeným počtom odlišných možných hodnôt, príp nepretržitý.

2 spojité náhodné premenné

Kumulatívna distribučná funkcia náhodnej premennej i

Teória predpovedí

1. Úvod

Best Linear Unbiased Prediction (BLUP) bola vyvinutá pre chov zvierat Dr. Charlesom Roy Henderson okolo roku 1949. Metódy boli prvýkrát aplikované na genetické hodnotenie dojníc na severovýchode Spojených štátov amerických v roku 1970. BLUP sa používa takmer vo všetkých krajinách a u všetkých hospodárskych zvierat druhov na genetické hodnotenie jednotlivých zvierat.

DEFINÍCIA: Predikcia je odhad realizovanej hodnoty náhodnej premennej (od údaje), ktoré boli odobraté z populácie so známou štruktúrou rozptylu a kovariancie.

2 Všeobecný lineárny zmiešaný model

Písanie lineárneho modelu

1 Časti modelu

Lineárny model sa v tradičnom zmysle skladá z troch častí:

1. Rovnica.

2. Očakávania a matice rozptylu-kovariancie náhodných premenných.

3. Predpoklady, obmedzenia a obmedzenia.

1.1 Rovnica

Rovnica modelu obsahuje pozorovací vektor pre danú vlastnosť(y) záujmu, faktory ktoré vysvetľujú, ako k pozorovaniu došlo, a zvyškový efekt, ktorý zahŕňa všetko, čo nie vysvetliteľné.

Maticová formulácia všeobecnej modelovej rovnice je:

Parametre pevných efektov sú spojené so známymi vysvetľujúcimi premennými, ako v štandardnom lineárnom modeli. Tieto premenné môžu byť buď kvalitatívne (ako pri tradičnej analýze rozptylu) alebo kvantitatívne (ako pri štandardnej lineárnej regresii). Avšak kovariančné parametre sú to, čo odlišuje zmiešaný lineárny model od štandardného lineárneho modelu.

Potreba parametrov kovariancie vzniká v aplikáciách pomerne často, pričom tieto dva najtypickejšie scenáre:

Experimentálne jednotky, na ktorých sa merajú údaje, môžu byť zoskupené do zhlukov a údaje zo spoločného zhluku sú korelované.

Opakované merania sa vykonávajú na rovnakej experimentálnej jednotke a tieto opakované merania sú korelované alebo vykazujú variabilitu, ktorá sa mení.

Prvý scenár možno zovšeobecniť tak, aby zahŕňal jednu sadu klastrov vnorených do inej. Napríklad, ak sú študenti experimentálnou jednotkou, môžu byť zoskupení do tried, ktoré môžu byť zase zoskupené do škôl. Každá úroveň tejto hierarchie môže predstavovať ďalší zdroj variability a korelácie. Druhý scenár sa vyskytuje v longitudinálnych štúdiách, kde sa v priebehu času vykonávajú opakované merania. Alternatívne môžu mať opakované merania priestorový alebo mnohorozmerný charakter.

PROC MIXED prispôsobí štruktúru, ktorú vyberiete, k údajom pomocou metódy obmedzenej maximálnej pravdepodobnosti (REML), ktorá je tiež známa ako zvyšková maximálna pravdepodobnosť. Práve tu sa využíva Gaussov predpoklad pre údaje. K dispozícii sú aj iné metódy odhadu, vrátane maximálnej pravdepodobnosti a MIVQUE0. Podrobnosti o týchto metódach odhadu sú diskutované v nasledujúcich častiach.

Keď sa model prispôsobí vašim údajom, môžete ho použiť na vyvodenie štatistických záverov prostredníctvom parametrov s pevnými efektmi a kovariancie. PROC MIXED počíta niekoľko rôznych štatistík vhodných na generovanie testov hypotéz a intervalov spoľahlivosti. Platnosť týchto štatistík závisí od zvoleného priemeru a modelu rozptylu a kovariancie, preto je dôležité vybrať model opatrne. Niektoré výstupy z PROC MIXED vám pomôžu posúdiť váš model a porovnať ho s ostatnými.

Táto časť predstavuje matematickú notáciu používanú v tejto kapitole na popis zmiešaného lineárneho modelu. Mali by ste byť oboznámení so základnou maticovou algebrou (pozri Searle 1982). Podrobnejší popis zmiešaného modelu je obsiahnutý v časti „Teória zmiešaných modelov“.

Štatistický model je matematický popis toho, ako sa generujú údaje. Štandardný lineárny model, ktorý sa používa v postupe GLM, je jedným z najbežnejších štatistických modelov:

1.1.2 Faktory

Diskrétne alebo spojité Spojitý faktor je faktor, ktorý má nekonečný rozsah možných hodnôt. Napríklad, ak je pozorovanie vzdialenosť, na ktorú môže byť kameň vrhnutý, potom a spojitým faktorom by bola hmotnosť horniny. Ak je pozorovaním rýchlosť rastu, potom by nepretržitým faktorom bolo množstvo zjedeného krmiva.

Diskrétne faktory majú zvyčajne triedy alebo úrovne, napríklad vek pri otelení môže mať štyri úrovne

(napr. 20 až 24 mesiacov, 25 až 28 mesiacov, 29 až 32 mesiacov a 33 mesiacov alebo viac). An

analýza dojivosti kráv bude závisieť od vekových úrovní kráv.

Fixné faktory V tradičnom „frekvencistickom“ prístupe musia byť fixné a náhodné faktory vyznamenali.

Ak je počet úrovní faktora malý alebo obmedzený na pevné číslo, potom tento faktor je zvyčajne fixný.

Ak sa závery o faktore obmedzia na tento súbor úrovní a na žiadne iné, potom je tento faktor zvyčajne pevný.

Ak sa urobila nová vzorka pozorovaní (nový experiment) a rovnaké úrovne a faktor sú v oboch vzorkách, potom je faktor zvyčajne fixný.

Ak boli hladiny faktora určené ako výsledok výberu medzi možnými dostupnými úrovne, potom by tento faktor mal byť pravdepodobne pevným faktorom.

Regresie kontinuálneho faktora sú zvyčajne fixným faktorom (ale nie vždy).

Náhodné faktory Ak je počet úrovní faktora veľký, potom tento faktor môže byť náhodný faktor.

Ak sa závery o faktore budú robiť pre celú populáciu pojmov úrovné, potom môže byť tento faktor náhodným faktorom.

Ak sú úrovné faktora vzorkou z nekonečne veľkej populácie, potom tento faktor je zvyčajne náhodný.

Ak sa urobila nová vzorka pozorovaní (nový experiment) a hladiny boli medzi týmito dvoma vzorkami úplne odlišné, potom sú faktory zvyčajne náhodné.

Príklady pevných a náhodných faktorov.

Pevné náhodné

Diéta Zvieratá

Súčasné skupiny plemien

Sex Stádo-Rok-Ročné obdobia

Vekové úrovné Trvalé prostredie

Klietky, nádrže Materské účinky

Ročné vrhy

Maternálne modely

U cicavčích druhov hospodárskych zvierat, ako je hovädzí dobytok, ovce alebo ošípané, poskytuje samica

prostredie pre prežitie a rast svojich potomkov z hľadiska ochrany a výživy.

Samice sa líšia v schopnosti poskytnúť vhodné prostredie pre svoje potomstvo, a to variabilita má genetický základ. Potomstvo priamo zdedí schopnosť rásť (alebo prežiť).

obaja rodičia a z hľadiska životného prostredia sa majú lepšie alebo horšie v závislosti od genetického pôvodu matky

schopnosť. Materská schopnosť je genetická vlastnosť, ktorú matka vyjadruje vo výkonnosti potomstva,

a prenáša sa, ako všetky genetické vlastnosti, od oboch rodičov. Materská schopnosť je len vyjadrená

samicami, keď majú potomstvo (t. j. podobne ako dojivosť u dojnic).

A model to account for maternal ability is $y = Xb + Z1a + Z2m + Z3p + e$,

Prahové modely

Prahové modely

1 Kategórické údaje

Ak sú pozorovania vyjadrené fenotypovo ako člen jednej z m kategórií, potom

údaje sú kategórické, diskkrétne a nesúvislé. Pri dvoch kategóriách, $m = 2$, potom táto vlastnosť je

„všetko alebo žiadne“ alebo binárne. Väčšina znakov choroby je binomická, pričom zviera je buď choré

alebo nie je chorý. Ľahkosť otelení alebo znaky konformácie majú viac ako dve kategórie usporiadané

ohraničená sekvencia od jedného extrémneho výrazu k opačnému extrémnemu výrazu. Telenie

ľahkosť má napríklad štyri kategórie, a to (1) telenie bez pomoci, (2) mierna pomoc, (3)

tvrdý ťah alebo (4) je potrebný cisársky rez. Napríklad uhol zadku má deväť kategórií, kde

kolíky sa pohybujú od príliš nízkych po príliš vysoké, pričom požadovaná kategória je v strede.

Kategórické znaky môžu byť dedené polygénnym spôsobom. Základná náchylnosť na

chorobný znak, ľahkosť otelenia alebo typový znak môžu byť v skutočnosti nepretržité a môžu nasledovať normálne

distribúcia. Základná spojitá stupnica je známa ako stupnica zodpovednosti. O zodpovednosti

stupnica je prahový bod (alebo body), kde nad týmto prahom zviera prejavuje chorobu

fenotyp a pod prahovým bodom zviera ochorenie neprejavuje. Zodpovednosť

mierka je len koncepcná a nedá sa dodržať.

6.24 Prahový model

6.24.1 Model

Niektoré vlastnosti vykazujú kategorické reakcie, ktoré môžu byť potenciálne zoradené podľa stupňa, napr. ťažkosti s otelením

(ľahké, mierne ťažké alebo ťažké) a choroba (nie je potrebná žiadna liečba, liečený alebo zomrel).

Pre takéto vlastnosti,

hoci fenotyp nie je súvislý, stále môžeme predpokladať, že genetické aj environmentálne faktory ovplyvňujú vlastnosti a dochádza k polygénemu efektu. V tomto prípade ide o rozumný model, ktorý má každé zviera

nepozorovateľná, hypotetická náhodná premenná, ako je fenotyp v spojitom znaku a taká premenná

prispieva ku kategorickej reakcii. V tomto modeli môžu k podkladu prispieť rôzne faktory premenlivý. Keď premenná prekročí určitú úroveň, vyjadruje inú kategóriu odzvy. Toto

je základná myšlienka prahového modelu a premenná sa nazýva zodpovednosť a úroveň je známa ako

prah. Zvyčajne sa predpokladá, že záväzok je normálne rozdelený v základnom rozsahu. Môžeme aplikovať lineárny model na základnú premennú. V tejto teórii je počet prahov ekvivalentný k počtu kategórií mínus 1.

Autor cituje dobre známy súbor údajov, ktorý pôvodne poskytli Gianola a Foulley (1983). Vlastnosť je

ľahkosť otelenia s 3 kategóriami. Vzor pre zodpovednosť je nasledovný:

$li\ jkl = Ahoj + Cj + sk + ei\ jkl$

kde $li\ jkl$ je záväzok, Hi a Cj sú pevné účinky, sk

je náhodný sire efekt a $ei\ jkl$ je

zvyškový. Zložky rozptylu sú σ

2

$s = 1/19 = 0,0526$ a σ

2

$e = 1,0$.

Rodina BLUPF90 mala predtým niekoľko softvérov na zvládnutie prahu. Teraz THRGIBBS1F90

je jediný verejne dostupný program. Toto je Gibbs sampler, takže potrebujeme veľa vzoriek na odhad

parametre polohy ako zadné prostriedky.

Viacznakové (multitrait) modely

Zvieratá sú bežne pozorované pre viac ako jednu vlastnosť, pretože veľa vlastností celkovo ovplyvňuje

ziskovosť. Model viacerých znakov (MT) je model, v ktorom sa analyzujú dva alebo viac znakov súčasne, aby sa využili genetické a environmentálne korelácie medzi nimi vlastnosti.

Znaky nízkej dedičnosti: Modely MT sú užitočné pre vlastnosti, pri ktorých je rozdiel medzi genetické a reziduálne korelácie sú veľké (napr. rozdiel väčší ako 0,5) alebo kde jeden znak má oveľa vyššiu dedičnosť ako iná vlastnosť. EBV pre znaky s nízkou dedičnosťou majú tendenciu získať väčšiu presnosť ako EBV pre vlastnosti s vysokou dedičnosťou, hoci všetky vlastnosti profitujú

do určitej miery zo simultánnej analýzy.

Culling: Ďalšie použitie modelov MT je pre vlastnosti, ktoré sa vyskytujú v rôznych obdobiach života zviera, takže utratenie zvierat vedie k menšiemu počtu pozorovaní na zvieratách, pokiaľ ide o znaky, ktoré sa vyskytujú

neskôr v živote v porovnaní s tými na začiatku. V dôsledku toho zvieratá, ktoré majú pozorovania neskôr

v živote majú tendenciu byť vyberaní na základe ich výkonnosti pre skoršie črty. Takže analýza znaky neskoršieho života by samy o sebe mohli trpieť účinkami zaujatosti utratenia a výsledného EBV

môže viesť k chybám pri výbere budúcich rodičov. MT analýza, ktorá zahŕňa všetky pozorovania na zvierati, na ktorom sa zakladali rozhodnutia o utratení, sa niektorým ukázalo, že to berie do úvahy

stupňa, za výber, ktorý sa uskutočnil. Ak je percento pozorovaní chýbajúcich vlastností vysoká, potom by sa do EBV pre túto vlastnosť a možno aj iné mohla vkradnúť zaujatosť.

Modely MT neponúkajú veľké zvýšenie presnosti v prípadoch dedičnosti vlastností sú podobné vo veľkosti, kde genetické aj reziduálne korelácie sú relatívne rovnaké veľkosť a znamenie, alebo kde je každé zviera merané pre všetky vlastnosti. Ak však utratenie zaujatosti

môže existovať, potom by sa mala vykonať analýza MT, aj keď sú parametre podobné. MT analýza sa spolieha na presnosť predpokladaných genetických a reziduálnych korelácií.

Ak sa odhady parametrov výrazne líšia od základných neznámych skutočných hodnôt, potom by analýza MT mohla spôsobiť toľko škody, ako aj úžitku.

Animal model

Pri genetickom hodnotení zvierat sa východiskové merania vykonávajú na jednotlivých zvieratách, a preto model analýzy nazývaný animal model zahŕňa aditívny genetický efekt zvierat. Zostávajúce položky v modeli sú rôznorodé faktory, ktoré majú vplyv na hodnotený znak. Predpokladom použitia animal modelu je aby zvieratá, ktoré sú hodnotené predstavovali náhodnú vzorku, alebo úplný súbor potomstva chovných samcov a samic v populácii. To znamená, že jednotlivé merania by sa nemali získavať iba od selektovaného potomstva, najlepšieho potomstva, alebo najhoršieho potomstva, ale mala by to byť skutočná náhodná vzorka potomstva.

Jednoduchý animal model s jedným záznamom na zviera môžeme zapísať:

$$y_i = \mu + a_i + e_i,$$

kde y_i je fenotypový záznam alebo pozorovanie, μ je celkový priemer pozorovaní, a_i sú aditívne genetické hodnoty zvierat a e_i sú súvisiace reziduálne alebo environmentálne účinky pozorovaní.

Základné predpoklady animal modelu sú:

1. Populácia je veľká (nekonečná).
2. Predpokladá sa, že asociácia alel od dvoch rodičov je náhodná.
3. Na jednotlivé záznamy nebol aplikovaný žiadny výber.
4. Existujú iba aditívne genetické účinky.
5. Absentuje vplyv iných faktorov na pozorovania.

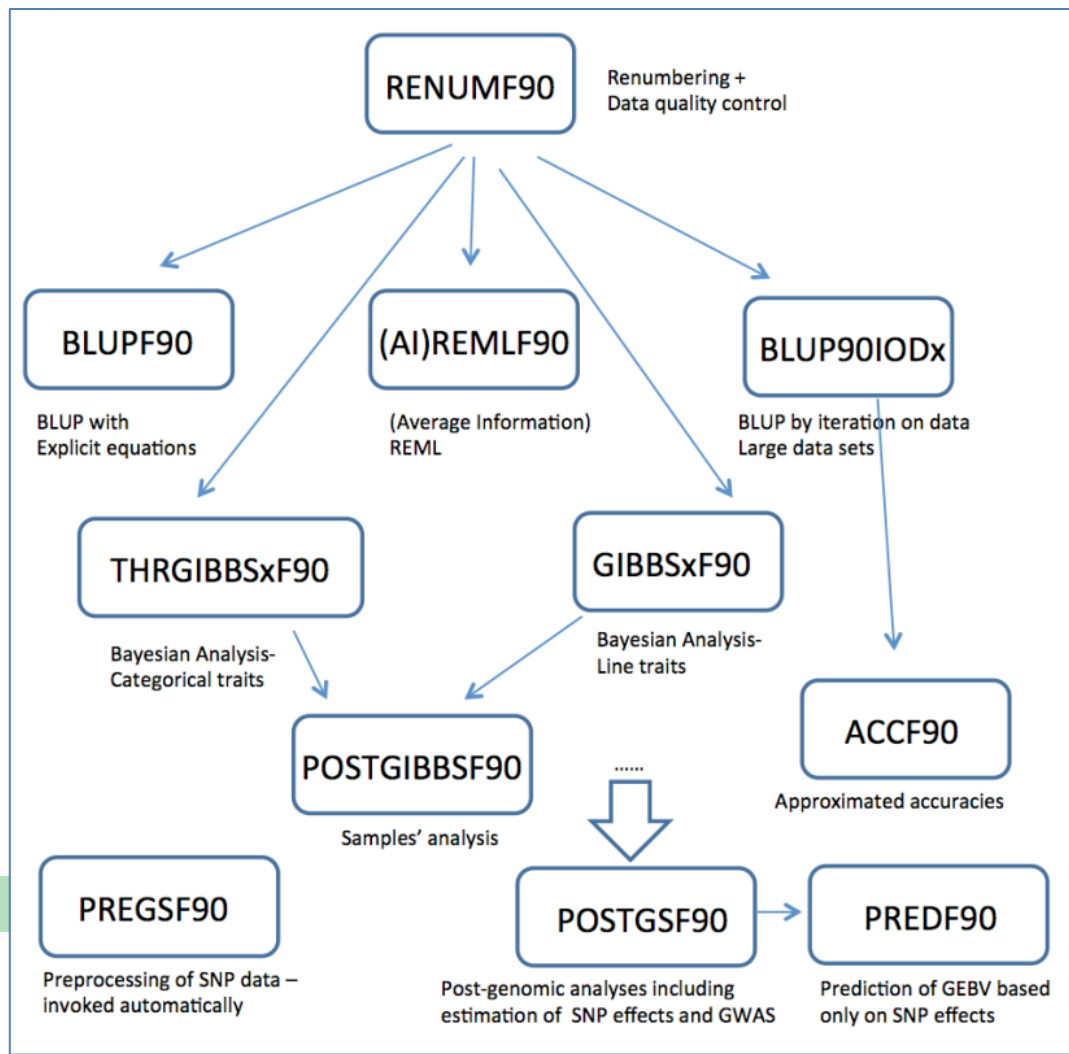
Náhodné regresné modely

V rámci analýzy biologických javov a procesov zohráva dôležitú úlohu analýza rastu a vývoja zvierat. Vlastnosti, ktoré sa merajú počas rôznych období ich života sa nazývajú longitudinálne údaje. Príkladmi sú telesná hmotnosť, dĺžka tela, produkcia mlieka, príjem krmiva, ukladanie tuku, produkcia vajec, alebo produkcia vlny. Na biologicko-genetickom základe sú ovplyvňované rôznymi génmi, ktoré sa zapínajú, alebo vypínajú, keď zviera starne, čo spôsobuje

zmeny vo fyziológii a v samotnej produkčnej schopnosti. Vek zvierat'a možno zaznamenávať v rokoch, mesiacoch, týždňoch, dňoch, hodinách, minútach, alebo sekundách, takže v skutočnosti môže existovať súvislý rozsah bodov v čase, keď bolo možné zviera pozorovať a hodnotiť nejakú vlastnosť. Ako príklad si vezmite telesnú hmotnosť prasničiek zo 60-dňového rastového testu.

Animal model s opakovanými záznamami

Opakovanie záznamov je typické pre veľa druhov hospodárskych zvierat v rámci jednotlivých skupín priamych produkčných a nepriamych sekundárnych vlastností a znakov. Ako príklad uvádzame: opakovaná mlieková úžitkovosť hovädzieho dobytku, oviec a kôz, produkcia vlny oviec, záznamy o narodených potomkoch plemenníc, veľkosť vrhu prasnice, veľkosť parožia jeleňa v rôznych ročných obdobiach, dostihové a parkúrové výsledky koní. Zvyčajne sa tieto vlastnosti považujú za vlastnosti korelované s vekom zvierat'a. Okrem aditívnej genetickej hodnoty zvierat'a pre danú vlastnosť, existuje účinok spoločného trvalého prostredia (PE), čo je v skutočnosti negenetický účinok spoločný pre všetky pozorovania na tom istom zvierati. Práve zohľadnenie účinku trvalého prostredia je charakteristické pre tento typ animal modelov s opakovanými záznamami.



zdroj: Misztal et al, 2018 (<http://nce.ads.uga.edu/wiki/doku.php>)

ASReml je štatistický balík, ktorý vyhovuje lineárnym zmiešaným modelom pomocou maximálnej reziduálnej pravdepodobnosti (REML). Modely lineárnych zmiešaných efektov poskytujú bohatý a flexibilný nástroj na analýzu mnohých súborov údajov, ktoré sa bežne vyskytujú v poľnohospodárskych, biologických, lekárskech a environmentálnych vedách. Typické aplikácie zahŕňajú analýzu (ne)vyvážených longitudinálnych údajov, analýzu opakovaných meraní, analýzu (ne)vyvážených navrhnutých experimentov, analýzu multi-environmentálnych skúšok, analýzu jednorozmerných a viacrozmerných šľachtiteľských a genetických údajov a analýzu pravidelných alebo nepravidelných priestorových údajov. ASReml poskytuje stabilnú platformu na poskytovanie osvedčených postupov a zároveň poskytuje súčasný výskum v aplikácii lineárnych zmiešaných modelov. Silnou stránkou ASReml je použitie algoritmu Average Information (AI) a metód riedkej matice na prispôbenie lineárneho zmiešaného modelu.

PEST je užívateľsky prívetivý Fortran program na predikciu plemenných hodnôt v praktickom chove zvierat. Zahŕňa pevné, náhodné a zmiešané modely. Súbor parametrov pozostáva z popisu vstupných údajov, štatistického modelu, výstupu, transformácií údajov atď.

Niektoré ďalšie funkcie PEST sú: Spracovanie chýbajúcich hodnôt; Rôzne matrice výskytu pre niektoré alebo všetky znaky; Zahŕnutie vzťahu medzi zvieratami; Liečba príbuzenského kríženia; Modely genetických skupín; Heterogénne odchýlky; Ošetrovanie modelov zvierat, samcov a matiek; Predikčné odchýlky chýb pre BLUP a štandardné chyby BLUE.

VCE je program na odhadovanie matíc kovariancie pomerne všeobecným spôsobom. Dostupné sú tri metódy: 1) REML využívajúce analytické gradienty, 2) Monte-Carlo EM, ktorý tiež robí REML a 3) Gibbsov odber vzoriek. VCE je kompletne prepísaný vo Fortran90. Dodáva sa ako spustiteľné súbory pripravené na spustenie a má veľmi flexibilné definície modelu.

R je 'GNU S', voľne dostupný jazyk a prostredie (softvér) pre štatistické výpočty a grafiku. Poskytuje širokú škálu štatistických a grafických techník: lineárne a nelineárne modelovanie, štatistické testy, zmiešané modely a komponenty rozptylu, analýzu časových radov, klasifikáciu, zhľukovanie atď. R možno rozšíriť prostredníctvom balíkov. Existuje asi osem balíkov dodávaných s distribúciou R a mnohé ďalšie sú dostupné prostredníctvom rodiny internetových stránok CRAN pokrývajúcich veľmi široký rozsah moderných štatistík.

Aplikácia programu R

Praktické použitie programu R (R Studio)

5. úloha (program R)

Pomocou lineárnych zmiešaných modelov odhadnite základné genetické parametre a plemenné hodnoty jedincov analyzovaného vzorového súboru kráv holštajnského plemena (Vazquez et al, 2010).

Lineárny animal model 1:

$$y_{ijk} = \beta_0 + \beta_1 L_i + \beta_2 \log(\text{DIM})_{ij} + c_j + h_k + e_{ijk},$$

kde y_{ijk} je štandardizovaná produkcia mlieka j-tej kravy na i-tej laktácii; β_0 je spoločný efekt pre všetky záznamy v rámci regresie; L_i je poradie laktácie ($i = 1, 2, \dots, 5$); $(\text{DIM})_{ij}$ je počet dní laktácie kravy; β_1 a β_2 sú fixné regresné koeficienty laktácie; c_j je náhodný aditívny efekt kravy j ($j = 1, 2, \dots, 1\,359$); h_k je náhodný efekt stáda k ($k = 1, 2, \dots, 57$); a e_{ijk} je náhodný efekt (rezíduum).

Lineárny animal model 2:

$$y_{ijk} = \beta_0 + \beta_1 L_i + \beta_2 \log(\text{DIM})_{ij} + c_j + pe_j + h_k + e_{ijk},$$

kde pe_j je permanentný náhodný efekt kravy.

Zdrojové kódy zmiešaných lineárnych modelov:

Tab.

```
#Lineárny animal model 1 (aditívny efekt)

milk <- within(milk, sdMilk <- milk / sd(milk))
system.time(
  fm1 <- pedigreemm(sdMilk ~ lact + log(dim) + (1|id) + (1|herd),
    data = milk, pedigree = list(id = pedCowsR))
```

```
summary(fm1)

# Lineárny animal model 1 (aditívny efekt + PE efekt)

milk <- within(milk, idPE <- id)
fm2 <- pedigreemm(sdMilk ~ lact + (1|id) + log(dim) + (1|idPE) + (1|herd),
  data = milk, pedigree = list(id = pedCows))

summary(fm2)
```

Výsledky úlohy 5a (aditívny efekt)

Tab.

```
Linear mixed model fit by REML ['lmerpedigreemm']
Formula: sdMilk ~ lact + log(dim) + (1 | id) + (1 | herd)
Data: milk

REML criterion at convergence: 8384.8

Scaled residuals:
    Min       1Q   Median       3Q      Max
-4.2902 -0.5251  0.0179  0.5645  3.4840

Random effects:
Groups      Name          Variance Std.Dev.
id          (Intercept)  0.2780    0.5273

herd        (Intercept)  0.2078    0.4559
Residual                    0.4833    0.6952
Number of obs: 3397, groups: id, 1359; herd, 57

Fixed effects:
              Estimate Std. Error t value
(Intercept)  1.70371    0.26854   6.344
lact         -0.10779    0.01232  -8.747
log(dim)      0.73594    0.04380  16.804

Correlation of Fixed Effects:
          (Intr) lact
lact      -0.296
log(dim) -0.960  0.221
```

Výsledky úlohy 5b (aditívny efekt + permanentný efekt)

Tab.

```
Linear mixed model fit by REML ['lmerpedigreemm']
Formula: sdMilk ~ lact + (1 | id) + log(dim) + (1 | idPE) + (1 | herd)
Data: milk

REML criterion at convergence: 8364.4

Scaled residuals:
    Min       1Q   Median       3Q      Max
```

```
-4.2432 -0.5196 0.0118 0.5586 3.4682
```

Random effects:

Groups	Name	Variance	Std.Dev.
id	(Intercept)	0.06829	0.2613
idPE	(Intercept)	0.19880	0.4459
herd	(Intercept)	0.20403	0.4517
Residual		0.47749	0.6910

Number of obs: 3397, groups: id, 1359; idPE, 1359; herd, 57

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.70157	0.26960	6.311
lact	-0.10880	0.01229	-8.854
log(dim)	0.73233	0.04371	16.755

Correlation of Fixed Effects:

	(Intr)	lact
lact	-0.296	
log(dim)	-0.954	0.223

Súčasťou výsledkov oboch použitých modelov sú odhadnuté riešenia všetkých náhodných efektov, z ktorých odhadnuté riešenia kráv predstavujú v skutočnosti odhadnuté základné plemenné hodnoty. Základné plemenné hodnoty je následne možné upraviť a vyjadriť v bežne používaných formách.

6. úloha (program R)

Výsledky úlohy 6

Zdroje a zoznam použitej literatúry

- Caballero, A. and Toro, M. A. 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research* 75: 331-343.
- Colleau, J. J. 2002. An indirect approach to the extensive calculation of relationship coefficients. *Genetics Selection Evolution* 34: 409-421.
- David, P., Pujol, B., Viard, F., Castella, V. and Goudet, J. (2007), Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology*, 16: 2474
- Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. 4th ed. Harlow, UK: Longman.
- Harvey, W.R, 1977. Users Guide for LSML76. Mixed model least squares and maximum likelihood computer program. U.S. depart. of Agric., ARS.
- Henderson, C.R., 1953. Estimation of variance and covariance components *Biometrics* 9:226-252.
- Kempthorne, O. (1957). *An Introduction to Genetic Statistics*. New York: John Wiley & Sons.
- Masuda, Y. 2018. Introduction to BLUPF90 suite programs. University of Georgia. <http://nce.ads.uga.edu/wiki/doku.php?id=documentatio>
- Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, Z. Vitezica. 2018. Manual for BLUPF90 family programs.
- Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A.S., Lacey, R.C. & Dasmahapatra, K.K. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, 111: 3775-3780.
- Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32: 949-953.
- Sargolzaei, Mehdi & Iwaisaki, H. & Colleau, J.J.. (2006). CFC: A tool for monitoring genetic diversity. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production*. 27-28.
- Sargolzaei, M. and Colleau, J. J. 2006. Decomposing inbreeding and coancestry into ancestral components. *The 8th World Congress on Genetics Applied to Livestock Production*, Belo Horizonte, Brazil, August 13-18, 2006.
- Sargolzaei, M., Iwaisaki, H. and Colleau, J. J. 2005. A fast algorithm for computing inbreeding coefficients in large populations. *Journal of Animal Breeding and Genetics* 122: 325-331.
- Slate, J., David, P., Dodds, K. G., Veenvliet, B. A., Glass, B. C., Broad, T. E., & McEwan, J. C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity*, 93(3), 255-265.
- Sonesson, A. K. and Meuwissen, T. H. E. 2000. Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genetics Selection Evolution* 32: 231- 248.
- Szulkin, M., Bierne, N., & David, P. (2010). HETEROZYGOSITY-FITNESS CORRELATIONS: A TIME FOR REAPPRAISAL. *Evolution*, 64(5), 1202-1217.

- Wright, S. 1922. Coefficients of inbreeding and relationships. The American Naturalist 56: 330-338.