Genetika ve šlechtění zvířat

část 6. (rough draft version)

7. Biometrické metody v genetice – lineární modely

Cílem:

1. popsat genetickou strukturu populace
2. popsat změny genetické výstavby populací

Možnosti biometrických metod

A. odhady výkonnosti populací – čistokrevné i hybridní
B. odhady genetických parametrů – h^2, r_{op}, r_G, ...
C. odhady plemenné hodnoty (PH) – rozdíly mezi jedincem a vrstevníky, očištěný od negenetických vlivů
D. stanovení selekčního (genetického) zisku
E. optimalizace selekčních a hybridizačních programů

Uplatnění poznatků: molekulární a biochemické genetiky, cytogenetiky, imunogenetiky a genové manipulace (transgenose, ...). Tyto poznatky budou využity GENETIKOU POPULACÍ, která musí posoudit jejich význam ve šlechtění hospodářských zvířat.

Kvantitativní genetika – hodnocení pomocí modelů

Biometrika v genetice (≈ *kvantitativní genetika*) vychází z počtu pravděpodobnosti (jedná se o hromadné náhodné jevy), protože jednotlivé geny a jejich efekty prozatím nelze studovat přímo. Jejich analýza je na základě teorie statistiky a pravděpodobnosti. Společné efekty více genů vytváří proměnlivost, většinou s normálním rozdělením, kterou lze analyzovat matematicko-statistickými operacemi. Účinek polygenů se sleduje na základě počtu pravděpodobnosti.

- využití:

A. volit výběrové metody – např. jak velký soubor je nutné sledovat,
B. použití odhadu – bodový (normální křivka) a intervalový (stupeň volnosti, intervaly),
   - rizika odhadu: $\alpha = 0,05$ a $\alpha = 0,01$,
C. testování hypotéz: H_0 – nulová hypotéza, H_1 – alternativní hypotéza – homogenity rozptylu,
   - průkaznost mezi průměry (t-test),
   - analýza variance (ANOVA),
   - efekty pevné, náhodné, smíšené,
D. stanovení stupně genetické podobnosti,
E. analýza společných genových efektů.

Genetika kvantitativních znaků má základ v možnosti zjištění:

- četnost různých alel a genotypů,
- efektu produktů alel jednotlivých genů.

Genetika kvantitativních znaků je založena na hodnocení účinku POLYGENŮ na základě normálního rozdělení fenotypových hodnot (tj. mají plynulou proměnlivost). Přenos genetické informace (GI) u kvantitativních vlastností je **POLYGENNÍ** (**velký počet lokusů** s mendelistickým přenosem + větší či menší vliv **prostředí – vnitřní** a vnější).

Operační metody pro analýzu přenosu této GI: **biometrické**.

Pravidla přenosu:

GI (mendelisticky), genová vazba, interakce, vazba na pohlaví, …
Prostředí – vnější, vnitřní

Jakési hřiště s pravidly.
Hra polygennů – je analyzována biometrickými pravidly; matematické a statistické operace.

Analýza variance (ANOVA)

Jedná se o statistickou metodu, jejíž principy zavedl již R. Fisher na počátku 20. století. Tato technika umožňuje rozčlenit celkové variance do složek (komponent) podle jednotlivých příčinných zdrojů (např. genetické a prostřeďové).

$$\sigma^2_{celková} = \sigma^2_1 + \sigma^2_2 + \ldots + \sigma^2_k$$

Celková variance je dána součtem jednotlivých variancí dílčích zdrojů (faktorů) variability.

Funkce ANOVA:

1. odhad pevných efektů,
2. odhad komponent (složek) variance – podíl jednotlivých variancí, např. varianci genotypovou,
3. testování hypotéz o příčinách variance modelem (jak vznikla, velikost vlivu faktorů).

Metody výpočtů:

ANOVA
NEBALANCOVANÉ METODY

speciální případ nebalancovaných metod

BALANCOVANÉ METODY
- výjimečné
- speciální případ nebalancované metody

ANOVA – nebalancované metody:

1. velké systém rovnic s využitím matic
2. nelze realizovat podle plánu – náhodný efekt (využití u zvířat)
3. hodnotí se chovy, šlechtění (software: Harvey, SAS, BMPD) – ***metody nejmenších čtverců, maximální věrohodnosti***

1. otec má 100 potomků, 2. jich má 50 a 3. 10 – to je nebalancované

Balancované metody:

1. přesnější
2. plánované pokusy (u zvířat toho nelze dosáhnout)

- stejný počet pozorování ve všech podtřídách

Měřítka asociací

- LINEÁRNÍ VZTAH: regresní koeficient, korelační koeficient
- NELINEÁRNÍ VZTAH: transformace jedné nebo obou proměnných na lineární vztah

Interklasní korelace: směr a síla oboustranné závislosti proměnlivosti vlastností X a Y.

Intraklasní korelace: při měření stejné vlastnosti u příbuzných jedinců. Koeficient opakovatelnosti. Tento je velmi důležitý pro kvantitativní genetiku.

Biometrické modely – lineární

Kvalita statistické analýzy je nelépe posuzována modelem, u kterého se předpokládá, že popisuje data. Model musí adekvátně reprezentovat přirozenou povahu dat a vyjadřovat biologickou podstatu jevu. Existují tři úrovně modelů:

1. **Pravdivý (skutečný, teoretický) model** popisuje data přesně, bez reziduální nebo nevysvětlené variance. Pravdivý model není nikdy přesně znám.
2. **Ideální (praktický) model** je vytvořen výzkumníkem, který je tak blízký skutečnému modelu, jak jen to je možné. Takový model by se měl používat k analýzám, ale často není dostatek informací.
3. **Operační (pracovní, proveditelný) model** je zjednodušená forma ideálního modelu a je využíván výzkumníky v analýzách. Na této úrovni se vede široká diskuse o nejlepší operační model.

Dobrý operační model je odvozen od ideálního modelu. Z důvodu limitovaných zdrojů informací může být operační model odvozen vhodným zjednodušením ideálního modelu. Je-li provedeno více zjednodušení, nebo nějaké zjednodušení nemůže být využito, pak se může stát, že analýza bude znehodnocena. Výzkumník by měl přesně znát, které předpoklady je nutné využít pro tvorbu modelu.

POZOROVÁNÍ

Vektor pozorování obsahuje prvky vyplývající z měření vlastnosti v daných jednotkách. Předpokládá se, že vektor pozorování je náhodný vektor vybraný z nekonečně velké populace vektorů stejné délky. Prvky vektoru pozorování mají mlutivariátní distribuci, na které je založeno mnoho statistických metod. Tento předpoklad by měl být neustále testován. Měřené vlastnosti u hospodářských zvířat jsou většinou kontinuální proměnné a jsou blízké normální distribuci[1].

EFEKTY A PROMĚNNÉ

Efekty (faktory) se vztahují k proměnným, buď diskrétním, nebo kontinuálním, které mohou ovlivňovat nebo být ve vztahu k prvkům ve vektoru pozorování. Např. mléčná užitkovost u skotu je ovlivněna efekty např. věkem krávy, sezónou, jejím genetickým potenciálem, stádo a počtem dnů mezidobí. Model je pak sestaven z efektů, které jsou identifikovány, zaznamenávány.

Diskrétní efekty mají obvykle třídy nebo úrovně (např. laktace – 1., ..., n). Takže analýza dat by poskytla např. odhady rozdílů v mléčné užitkovosti za jednotlivé laktace.

Určité efekty jsou pro výzkumníka významnější (např. plemeno) a začleňuje je do modelu, aby redukoval reziduální varianci. Některé pro výzkumníka vedlejší efekty musí být také začleněny do modelu (např. věk, pohlaví, stádo, výživa, ...). Tyto efekty jsou někdy označovány jako „*obtěžující efekty*“, které však nemohou být opomenuty při analýze, neboť mohou mít významný vliv na interpretaci výsledků pro výzkumníkem sledované efekty.

[1] Jiná skupina vlastností, často měřené subjektivně, spadají do dvou kategorií. Např. věk v měsících, pohlaví, březost nebo nebřezost, stupeň obtížnosti telení, počet snesených vajec jsou kategorické vlastnosti, které není vhodné analyzovat metodami předpokládající kontinuální distribuci. Pokud se však takto analyzují, jsou výsledky téměř stejné jako při použití metod diskrétních dat.

PEVNÉ A NÁHODNÉ EFEKTY

V tradičním statistickém pojetí se efekty rozdělují na pevné a náhodné. V Bayesianské statistice takové rozdělení efektů není.

Pevné efekty (fixní) jsou ty, v kterých úrovně zahrnují všechny možné úrovně, které lze pozorovat. Např. pohlaví může být jen samčí nebo samičí, či kastráti. Je-li počet úrovní efektu malý, omezený a stálý v daném počtu, i když byl proveden výběr nesčetněkrát, jedná se o efekt pevný. Patří sem např. počet laktací, věk, systém managementu, počet klecí, plemeno, typ krmiva atd.

Náhodné efekty jsou efekty, jejichž úrovně jsou považovány za náhodně vybrané z nekonečně velké populace úrovní. Efekt prasete (jedince) je náhodný, protože populace prasat světa dostatečně velká na to, aby byla považována za nekonečně velkou. Skupina, která je zahrnuta do analýzy, je náhodným vzorkem z této populace. Ve skutečnosti jsou prasata v experimentu vybrána z relativně malé subpopulace prasat, ale stále se považují za náhodný efekt, protože kdyby byl experiment opakován, pravděpodobně by byla vybrána odlišná skupina jedinců.

1. Kolik úrovní má efekt v modelu? Jestliže málo, pak je to pravděpodobně pevný efekt, jestliže mnoho, pak se jedná o náhodný efekt.
2. Je počet úrovní efektu v populaci dost velký na to, aby mohla být považována za nekonečnou? Jestliže ano, pak je pravděpodobně efekt náhodný.
3. Budou použity opět stejné úrovně, jestliže by byl experiment opakován podruhé? Jestliže ano, pak se jedná pravděpodobně o pevný efekt.
4. Byly úrovně efektu určeny nenáhodným způsobem? Jestliže ano, pak by měl být efekt určen jako pevný.

Výzkumník by měl být schopen si pomoci v tomto rozhodovacím procesu, jestliže ne, pak by měl vyhledat pomoc zkušeného statistika.

MODELY

Lineární modely obsahují řadu efektů (faktorů), které aditivně ovlivňují pozorování, ale proměnlivost v efektu může představovat např. i umocněný člen. Lineární modely jsou adekvátní u většiny biologických jevů. To však nenaznačuje, že nelineární modely nejsou důležité. Nelineární vztahy mohou být často aproximovány na lineární model. Dá-li nelineární model lepší ideální model než lineární, pak jeho použití je vhodnější.

V tradičním smyslu jsou lineární modely složeny ze tří částí:

1. Rovnice.
2. Matice očekávaných hodnot a variančně kovarianční matice náhodných proměnných.
3. Předpoklady a omezení.

ad 1. **Rovnice**

Rovnice modelu definuje efekty, které mohou mít vliv na pozorovanou vlastnost. Maticový zápis obecné modelové rovnice je: $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$

kde **y** je vektor pozorovaných hodnot vlastnosti,
b je vektor pevných efektů,
u je vektor náhodných efektů,
e je vektor náhodných reziduálních efektů,
X a **Z** jsou známé matice, matice pokusu, které popisují přesné vztahy mezi prvky **b** a **u** s prvky **y**.

Tato rovnice je smíšený model, která obsahuje fixní a náhodný efekt. Protože se však vektor **e** vyskytuje u všech lineárních modelů, lze říci, že jsou všechny smíšené.

Model fixních efektů je jen jeden: $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$.

Model náhodných efektů je jen jeden: $\mathbf{y} = \mathbf{Zu} + \mathbf{e}$., kdy $\mathbf{Xb} = 1\mu$.

ad 2. **Matice očekávaných hodnot a VCV**

Matice očekávaných hodnot jsou:

$$E\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{Xb} \\ 0 \\ 0 \end{bmatrix}$$

Matice varianćně kovarianční jsou:

$$V\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{bmatrix}$$

kde **G** a **R** jsou základní čtvercové matice s předpokladem nesingularity a pozitivní definovanosti a s prvky, které jsou známé. Takže: $\mathbf{V}(\mathbf{y}) = \mathbf{ZGZ}' + \mathbf{R}$.

ad 3. **Předpoklady a omezení**

Třetí část modelu začleňuje položky, které nejsou zjevné v bodech 1 a 2. Např. informace o datech nebo způsob jejich sběru, náhodnost výběru, podmínkách chovu apod. V této části modelu by měly být uvedeny rozdíly mezi operačním a ideálním modelem a měly by být možné efekty těchto rozdílů vysvětleny. Tato část však často bývá ignorována nebo přehlížena.

Lineární model není kompletní, dokud nejsou všechny tři části modelu vyřešeny.

Typy modelů

BIOMETRICKÉ LINEÁRNÍ MODELY – obecně

- model je definován rovnicí, která je lineární funkcí určitých parametrů a proměnných

$y_{ij} = \mu + a_i + e_{ij}$

μ – fixní populační průměr
a_i – efekt ošetření
e_{ij} – reziduum

1. REGRESNÍ MODELY – funkční vztahy

$y_i = a + bX_i + e_i$ a, b – pevné efekty (parametry)

2. MNOHONÁSOBNÉ REGRESNÍ VZTAHY

$y_i = a + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + e_i$

a, b_1, b_2, b_3 – konstanty
e_i – měřítko nejistoty, reziduum (náhodně proměnná)

3. MODELY S PEVNÝMI EFEKTY

(speciální případ regresního modelu, kde $x_i = 0$ nebo 1)

Pevný efekt je průměrná hodnota a další efekty, které chci vyhodnotit.

$y_{ijkl} = \mu + a_i + b_j + c_k + e_{ijk}$

-chceme vyhodnotit, kvantifikovat vliv pevných efektů: μ, a_i, b_j, c_k

- e_{ijkl} – náhodně proměnná N (0; 7)

Pevný efekt = záměrně vybraný (mám 3 býky a chci vědět, který je lepší; býci jsou zde jako pevné efekty; výživa; hnojení)

Cíl: přesná kvantifikace parametru.

Hierarchické (faktoriální) uspořádání: $y_{ijk} = \mu + a_i + b_{ij} + e_{ijk}$

Směs analyzovaných parametrů: $y_{ijkl} = \mu + b_i + b_{ij} + c_k + bX_{ijkl} + e_{ijkl}$

μ, a_i - pevný efekt
b_{ij} - hierarchické třídění
c_k - faktor konstantní
bX_{ijkl} – regresní faktor (sleduji selata, některá mají 30 kg, jiná více či méně – nepřesnosti a výkyvy odstraníme regresí)
e_{ijkl} – náhodný efekt

Čím více faktorů pokryjeme, tím je výpočet přesnější, tím více se blížíme k variabilitě způsobenou genotypem.

4. MODELY S NÁHODNÝMI EFEKTY

$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ijkl}$

μ - fixní parametr

α_i, β_j, γ_k, e_{ijk} – náhodné parametry (chov, otec, ...) – jejich průměr je vždy roven 0.

Náhodný efekt – jiný vzorek jedinců = jiný výsledek (náhodný výběr zvířat)

- vybírám skupiny polosourozenců, ale nemáme je všechny, proto nemůže být pevným efektem
- chci zjistit odhady <u>variancí</u> – „*varianční komponenty*“ (mezi polosourozenci) pro výpočet genetických parametrů (h^2)

5. MODELY SE SMÍŠENÝMI EFEKTY (výše uvedené modely jsou všechny smíšené!)

$y_{ijk} = \mu + a_i + \beta_j + e_{ijk}$

μ, a_i – pevné efekty

β_j, e_{ijk} – náhodný efekt

- fixní efekty komplikují odhad variance náhodných efektů
- náhodné efekty komplikují odhad variance pevných efektů
- **smíšené modely se používají k odhadu PH**

Vyjádření modelů maticovým zápisem

- Skalární zápis modelu s pevnými efekty:

  $$y_{ijk} = \mu + a_i + b_j + e_{ijk}$$

  -jedna pozorovaná hodnota (zastupuje všechny pozorované hodnoty) je symbolicky znázorněna
- Maticový model s pevnými efekty, kde jsou vyjádřeny všechny pozorované hodnoty:

  $$\mathbf{Y} = \mathbf{Xb} + \mathbf{e}$$

  Y – vektor pozorování

  X – incidenční matice (designová, strukturní matice) – uvádí, které pevné efekty jsou obsaženy v Y)

  b – vektor odhadovaných parametrů

  e – vektor náhodných efektů: $\mathbf{e} \sim N(0, I\sigma_e^2)$

Vše co zjistím o jednom zvířeti, zapíši do jednoho řádku matice.

Analýza množství tuku u osmi dojnic s vlivem efektů stáda a věku:

a_i – stádo (i = 1, 2); b_j – věk (j = 1, 2, 3)

		věk		
		b_1	b_2	b_3
stádo	a_1	165 154	136	161
	a_2		115 142 186	112

Vyjádření dat ve vektorovém a maticovém formátu:

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{131} \\ y_{221} \\ y_{222} \\ y_{223} \\ y_{231} \end{bmatrix} = \begin{bmatrix} 165 \\ 154 \\ 136 \\ 161 \\ 115 \\ 142 \\ 186 \\ 112 \end{bmatrix} = \begin{bmatrix} \mu + a_1 + b_1 + e_{111} \\ \mu + a_1 + b_1 + e_{112} \\ \mu + a_1 + b_2 + e_{121} \\ \mu + a_1 + b_3 + e_{131} \\ \mu + a_2 + b_2 + e_{221} \\ \mu + a_2 + b_2 + e_{222} \\ \mu + a_2 + b_2 + e_{223} \\ \mu + a_2 + b_3 + e_{231} \end{bmatrix} = \begin{bmatrix} \mu & a_1 & a_2 & b_1 & b_2 & b_3 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} \mu \\ a_1 \\ a_2 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{131} \\ e_{221} \\ e_{222} \\ e_{223} \\ e_{231} \end{bmatrix}$$

$$\mathbf{y} = \quad \mathbf{y} = \quad \mathbf{Xb} + \mathbf{e} \quad = \quad \mathbf{X} \quad \mathbf{b} \quad + \quad \mathbf{e}$$

Maticový zápis je méně názorný než data zapsaná v tabulce, ale je kratší a úplnější než model skalární. Musí se však matice **X** definovat. Ta však při větším objemu dat může nabývat velikých rozměrů – nutná výkonná výpočetní technika a softwarové zázemí.

Je také třeba definovat **disperzní** (**variančně kovarianční, VCV**) **matici** pozorování. Můžeme vycházet z předpokladu, že každý náhodný efekt e_{ijk} je vybrán ze základního souboru s nulovým průměrem a variancí např. 30 kg.

$$\mathbf{V_e} = \sigma^2 \mathbf{I} = \begin{bmatrix} \sigma^2_{e1} & \sigma_{e1e2} & \sigma_{e1e3} & . & . & . \\ \sigma_{e2e1} & \sigma^2_{e2} & \sigma_{e2e3} & . & . & . \\ \sigma_{e3e1} & \sigma_{e3e2} & \sigma^2_{e3} & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \end{bmatrix} = \begin{bmatrix} 30 & & & & & & & \\ & 30 & & & & & & \\ & & 30 & & & & & \\ & & & 30 & & & & \\ & & & & 30 & & & \\ & & & & & 30 & & \\ & & & & & & 30 & \\ & & & & & & & 30 \end{bmatrix}$$

$$= 30 \begin{bmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ & & & & 1 & & & \\ & & & & & 1 & & \\ & & & & & & 1 & \\ & & & & & & & 1 \end{bmatrix} = 30\,\mathbf{I}_8 = 30\mathbf{I} = \sigma^2\mathbf{I}$$

Výraz $\sigma^2\mathbf{I}$ je skalární diagonální maticí, kdy jednotková matice je vynásobena skalárem (zde variancí). Pozorování kromě **e** neobsahují žádné jiné náhodné efekty. VCV matice pozorování **y** je stejná s VCV pro **e**, má stejné variance a kovariance:

$$\mathbf{V_y} = \mathbf{V_e} = \sigma^2\mathbf{I}$$

Řešení nejmenších čtverců pro zobecněný lineární model

Řešení zobecněného lineárního modelu $\mathbf{Y} = \mathbf{Xb} + \mathbf{e}$ je odvozeno tímto způsobem, kdy reziduální součet čtverců v maticovém zápisu je vypočítán následovně:

$$(\mathbf{y} - \mathbf{Xb})` \ (\mathbf{y} - \mathbf{Xb}) = \mathbf{e}`\mathbf{e}$$

$$\mathbf{y}`\mathbf{y} - 2(\mathbf{Xb})`\mathbf{y} + (\mathbf{Xb})`\mathbf{Xb} = \mathbf{e}`\mathbf{e}$$

Provede-li se derivace s ohledem, že $\mathbf{b}$ se rovná 0, získáme:

$$\mathbf{X}`\mathbf{X}\,\mathbf{b} = \mathbf{X}`\mathbf{y}$$
$$\mathbf{b} = (\mathbf{X}`\mathbf{X})^{-1}\,(\mathbf{X}`\mathbf{y})$$

Této rovnici se říká *„normální rovnice“* a je často používána v moderní statistice. Jsou-li pozorování korelovaná a nemají-li stejné variance nebo oboje, pak normální rovnice mohou být modifikovány:

$$\mathbf{X}`\mathbf{V}^{-1}\mathbf{X}\,\mathbf{b} = \mathbf{X}`\mathbf{V}^{-1}\mathbf{y}$$
$$\mathbf{b} = (\mathbf{X}`\mathbf{V}^{-1}\mathbf{X})^{-1}\,(\mathbf{X}`\mathbf{V}^{-1}\mathbf{y})$$

kde $\mathbf{V}$ je variančně kovarianční matice mezi pozorováními, která je diagonální s počtem řádků a sloupců rovno počtu pozorování. Řešení těchto rovnic je obtížné, protože vyžadují inverzi matice $\mathbf{V}$, kterou je obtížné vypočítat ve velkých souborech dat.

Řešení poslední rovnice se nazývá řešení *„zobecněných nejmenších čtverců“*, které minimalizuje $\mathbf{e}`\mathbf{e}$.

Příklady lineárních modelů

I. MODEL S JEDNÍM FIXNÍM EFEKTEM

- Pro analýzu např. hmotnosti těla býčků mezi 190. a 210. dnem života lze použít tento model: $y_{ij} = \mu + a_i + e_{ij}$, kde

y_{ij} - hmotnost j-tého býčka i-té skupiny; μ - je obecný průměr; a_i – efekt i-té skupiny; e_{ij} – reziduální efekt

Věková skupina	Hmotnost býčků v kg
1	198 204 201
2	203 206 210
3	205 212 216
4	225 220

Čtyři věkové skupiny byly tvořeny v pětidenních intervalech a byly určeny jako fixní.

- Očekávané hodnoty a variance jsou:

$$E(y_{ij}) = \mu + a_i \qquad E(e_{ij}) = 0$$

a $$V(y_{ij}) = V(e_{ij}) = \sigma_i^2$$

Takže reziduální variance je odlišná pro každou věkovou skupinu.

$$\sigma_1^2 = 0{,}9\,\sigma^2 \qquad \sigma_2^2 = 0{,}97\,\sigma^2 \qquad \sigma_3^2 = 1{,}03\,\sigma^2 \qquad \sigma_4^2 = 1{,}08\,\sigma^2$$

Maticový zápis:

$$\mathbf{y} = \begin{bmatrix} 198 \\ 204 \\ 201 \\ 203 \\ 206 \\ 210 \\ 205 \\ 212 \\ 216 \\ 225 \\ 220 \end{bmatrix}, \qquad \mathbf{Xb} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

Matice **Zu** v tomto příkladě neexistuje. Takže, V(**e**) = **R**, kde **R** je diagonální matice řádu 11 s diagonálními prvky:

R = diag(0,9 0,9 0,9 0,97 0,97 0,97 1,03 1,03 1,03 1,08 1,08)

- Třetí část modelu zahrnuje:
  1. Všichni jedinci jsou stejného plemene.
  2. Věk matek nemá vliv na hmotnost v 200 dnech.
  3. Všichni jedinci byli stejného plemene.
  4. Všichni jedinci jsou navzájem nepříbuzní.
  5. Všichni jedinci byli chování ve stejných podmínkách.

Všechna tato tvrzení by měla být doložena – vlastní evidence, citace, …

II. MODEL S JEDNÍM NÁHODNÝM EFEKTEM

Máme záznamy o imunitní odpovědi u myší na injekci nějaké látky.

Myš	Imunitní odpověď	
1	15	27
2	32	28
3	45	50
4	11	17
5	42	35

Bylo sledováno 5 myší, u každé byly provedeny 2 měření. Dvě myši byly ze stejného vrhu (1. a 2.), ale ostatní byli nepříbuzné.

- Model by mohl být tento:

$$y_{ij} = \mu + m_i + e_{ij} \quad \text{, kde}$$

y_{ij} – odpověď na látku; μ - je obecný průměr; m_i – efekt i-té myši; e_{ij} – reziduální efekt

- Očekáváné hodnoty a variance jsou:

$$E(y_{ij}) = \mu \qquad E(m_{ij}) = 0 \qquad E(e_{ij}) = 0$$

a

$$V(y_{ij}) = V(m_{ij} + e_{ij}) = V(m_{ij}) + V(e_{ij}) + 2\text{cov}(m_{ij}, e_{ij}) = \sigma^2_m + \sigma^2_e + 0$$

Maticový zápis:

$$\mathbf{y} = \begin{bmatrix} 15 \\ 27 \\ 32 \\ 28 \\ 45 \\ 50 \\ 11 \\ 17 \\ 42 \\ 35 \end{bmatrix}, \qquad \mathbf{Xb} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \end{bmatrix}$$

$$\mathbf{Xb} = 1\mu \text{ a } V(\mathbf{m}) = \mathbf{G} = \begin{bmatrix} 1 & 0{,}5 & 0 & 0 & 0 \\ 0{,}5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \sigma^2_m \text{ ; } V(\mathbf{e}) = \mathbf{R} = \mathbf{I}\sigma^2_e, \text{ a } V(\mathbf{y}) = \mathbf{ZGZ}' + \mathbf{R},$$

Vlastní sourozenci mají aditivní genetickou příbuznost 0,5. Musí se také specifikovat hodnoty σ^2_m a σ^2_e nebo nejméně poměr jejich hodnot. V tomto příkladě předpokládáme

$$\frac{\sigma^2_e}{\sigma^2_m} = 0{,}5$$

- Třetí část modelu může zahrnovat následující tvrzení:
  1. Nejsou rozdíly mezi prvním a druhým měřením imunitní odpovědi v důsledku jiné doby měření nebo jiných faktorů.
  2. Všechny myši mají stejnou výživu a chovné podmínky.
  3. Odběr krve neovlivňuje imunitní odpověď myší.
  4. Věk myší také neovlivňuje jejich imunitní odpověď nebo byly všechny myši stejného věku.

Cílem této analýzy byl odhad průměrné odpovědi myší na látku.

III. MODEL SMÍŠENÝ

Byla sledována hmotnost telat masného skotu v 200 dnech věku.

jalovice	býčci
198	187
211	194
220	202
-	185

- Rovnice modelu může být: $y_{ij} = s_i + t_j + e_{ij}$, kde

  y_{ij} – hmotnost v 200 dnech; s_i – efekt i-tého pohlaví telete (pevný efekt); t_j – efekt j-tého telete (náhodný efekt); e_{ij} – reziduální efekt

- Očekávané hodnoty a variance jsou:

$$E(t_j) = 0 \qquad E(e_{ij}) = 0$$

$$\text{a} \qquad V(t_j) = \sigma^2_t \qquad V(e_{ij}) = \sigma^2_{ei}$$

Dále $\mathrm{cov}(t_{j'}, t_{j''}) = 0$, což znamená, že všechna telata jsou na sobě nezávislá. σ^2_{ei} znamená, že reziduální variance je různá pro každé pohlaví i-tého telete. $\mathrm{cov}(e_{ij'}, e_{ij''}) = 0$ a $\mathrm{cov}(e_{ij'}, e_{i'j''}) = 0$ říká, že všechny reziduální efekty jsou nezávislé navzájem na sobě, v a mezi pohlavími.

- Předpoklady a omezení:
  1) Všechna telata jsou stejného plemene.
  2) Všechna telata jsou chována ve stejných podmínkách ve stejné době.
  3) Všechna telata jsou potomky matek stejného věku (např. 3 roky staré)
  4) Maternální efekt neovlivňuje hmotnost telat v 200 dnech věku.
  5) Efekt telete obsahuje všechny genetické efekty.
  6) Všechny hmotnosti byly měřeny a zapsány přesně (bez odhadu)

Předpoklad o maternálním efektu nemusí být správný, ale bez rodokmenové informace není možné začlenit maternální efekty do modelu. Efekt věku matek je většinou znám a neměl by být ignorován. Jestliže jeden nebo více předpokladů jsou známé, pak by měl být model přepsán, tak aby bylo co nejméně předpokladů.

Maticový zápis:

$$\mathbf{y} = \begin{bmatrix} 198 \\ 211 \\ 220 \\ 187 \\ 194 \\ 202 \\ 185 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad \text{a } \mathbf{Z} = \mathbf{I} \text{ řádu 7.}$$

Takže, $\mathbf{G} = \mathbf{I}\sigma^2_t = \mathrm{diag}(\sigma^2_t)$

$$\mathbf{R} = \mathrm{diag}(\sigma^2_{e1}, \sigma^2_{e1}, \sigma^2_{e1}, \sigma^2_{e2}, \sigma^2_{e2}, \sigma^2_{e2}, \sigma^2_{e2})$$

IV. Model kovariancí

Chov lososa v koších v moři je velmi rozšířen. Genetické zlepšování ryb si nezískává moc pozornosti, protože jsou problémy s identifikací jedinců, zejména v mladém věku. Předkládaný model pro ohodnocení hodnoty rozdílných otců a matek je:

$$y_{ijkl} = \mu + C_i + bX_{ijkl} + S_j + D_{jk} + e_{ijkl} \quad \text{, kde}$$

y_{ijkl} – hmotnost při zabití
μ - je obecný průměr
C_i – efekt i-tého koše (skupiny)
X_{ijkl} – je délka těla ryby
b – regresní koeficient hmotnosti na délku těla
S_j – efekt j-tého otce
D_{jk} – efekt k-té matky v j-tém otci
e_{ij} – reziduální efekt

Pevné efekty modelu jsou μ, C_i a bX_{ijkl}. Efekty otce a matky by mohly být pevnými i náhodnými. Jestliže by byly náhodně vybírány skupiny otců a matek při opakovaném šlechtění, pak by byly náhodnými efekty. Předpokládejme, že jsou náhodnými efekty, pak:

Očekávané hodnoty a variance jsou:

$$E(y_{ijkl}) = \mu + C_i + bX_{ijkl} \qquad E(S_j) = 0 \quad E(D_{jk}) = 0 \qquad E(e_{ijkl}) = 0$$

a $$V(y_{ij}) = V(m_{ij} + e_{ij}) = V(m_{ij}) + V(e_{ij}) + 2\mathrm{cov}(m_{ij}, e_{ij}) = \sigma^2_m + \sigma^2_e + 0$$

$$\mathbf{V}\begin{bmatrix}\mathbf{s}\\ \mathbf{d}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G_s} & 0 & 0\\ 0 & \mathbf{G_d} & 0\\ 0 & 0 & \mathbf{R}\end{bmatrix}=\begin{bmatrix}\mathbf{I}\sigma_s^2 & 0 & 0\\ 0 & \mathbf{I}\sigma_d^2 & 0\\ 0 & 0 & \mathbf{I}\sigma_e^2\end{bmatrix} \quad \text{a nechť } \frac{\sigma_e^2}{\sigma_s^2}=15 \text{ a } \frac{\boldsymbol{\sigma}_e^2}{\boldsymbol{\sigma}_d^2}=10$$

Třetí část řešení modelu zahrnuje:

1. Otcové a matky nejsou geneticky příbuzní.
2. Variance matek zahrnuje složku genetickou a maternální.
3. Reziduální variance je stejná pro všechny skupiny (koše).
4. Není vliv různého období.

V. EKVIVALENTNÍ MODELY

Modely jsou ekvivalentní, když dávají stejné očekávané hodnoty **y** a stejnou matici variančně kovarianční **y**. Ekvivalentní modely jsou často používány pro zjednodušení výpočtů specifických analýz. Jako příkladem je redukovaný animal model (RAM).

Pro ilustraci si jej uvedeme v modelu: $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$

kde $\mathbf{Z}'=\begin{bmatrix}1 & 1 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 1\end{bmatrix}$, $\mathbf{V}(\mathbf{u}) = \mathbf{G_u} = \begin{bmatrix}5 & 1 & 2\\ 1 & 5 & 0\\ 2 & 0 & 5\end{bmatrix}$, a $\mathbf{V}(\mathbf{e}) = \mathbf{R_e} = 9\mathbf{I} + \mathbf{J}$, pak

$$\mathbf{V}(\mathbf{y}) = \mathbf{V} = \begin{bmatrix}15 & 6 & 6 & 2 & 3 & 3\\ 6 & 15 & 6 & 2 & 3 & 3\\ 6 & 6 & 15 & 2 & 3 & 3\\ 2 & 2 & 2 & 15 & 1 & 1\\ 3 & 3 & 3 & 1 & 15 & 6\\ 3 & 3 & 3 & 1 & 6 & 15\end{bmatrix}$$

Stejnou matici **V** bychom měli získat s modelem: $\mathbf{y} = \mathbf{Xb} + \mathbf{Zt} + \boldsymbol{\varepsilon}$

kde $\mathbf{V}(\mathbf{t}) = \mathbf{G_t} = \begin{bmatrix}6 & 2 & 3\\ 2 & 6 & 1\\ 3 & 1 & 6\end{bmatrix}$, a $\mathbf{V}(\boldsymbol{\varepsilon}) = \mathbf{R}_{\boldsymbol{\varepsilon}} = 9\mathbf{I}$,

pak $\mathbf{V} = \mathbf{ZG_uZ}' + \mathbf{R_e} = \mathbf{ZG_tZ}' + \mathbf{R_e}$.

Biometrické odhady genetických parametrů

Problémy aplikace kvantitativní genetiky na populace zvířat jsou ve skutečnosti problémy statistických odhadů. Šlechtění je založeno na znalosti genetické struktury populací, kterou zatím pro kvantitativní vlastnosti nelze určovat přímo (frekvence alel a genotypů genů), ale musíme analyzovat efekty, příčiny genetické a prostřeďové, které se podílejí na proměnlivosti a to na základě 2 parametrů ⇒ **VARIANCE** a **KOVARIANCE**.

Popis genetické struktury populace: var_x (σ_x^2), cov_{xy}, $\sigma_A^2; \sigma_D^2; \sigma_I^2; \text{cov}_{GE}$.

<u>Realizace</u> - odhad PH jedince (OPH) (*Estimate of Breeding Value* – EBV)
- odhad genotypových hodnot skupin jedinců

ODHADY (vlastnosti)

- který z odhadů je nejlepší odhad?!?

BLUE – ***B**est **L**inear **U**nbiased **E**stimators*

- nejlepší lineární nevychýlené odhady (nejmenších čtverců)

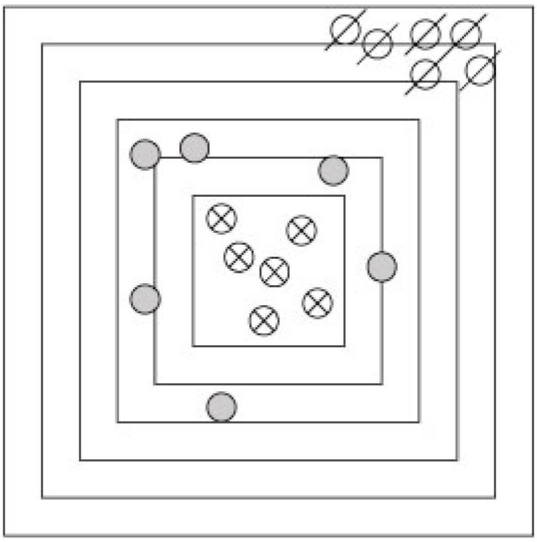
NEJLEPŠÍ – *BEST* (*PŘESNÉ*)

- nejlepší odhad průměru populace = náhodný vzorek (reprezentativní, dostatečný počet), pak je nejlepším odhadem
  - odhad $\hat{b}$ je nevychýleným parametrem b, když průměr všech možných hodnot se rovná parametru $E(\hat{b}) = b$
- nejlepší odhad je **odhad s minimální variancí** = metodou nejmenších čtverců (metoda odhadu), které minimalizují varianci
  - $E(\hat{u} - u)^2 = \min.$
- nejlepší odhad je odhad s maximální korelací mezi skutečnou a odhadovanou hodnotou
- nejlepší odhad PH – souhrnná PH = vložit do selekčního indexu, který hodnotí všechny PH pro všechny hodnocené vlastnosti; nejlepším odhadem je hodnota, která maximalizuje genetický zisk

Využíváme:

- **lineární modely** – každý odhad je počítán jako lineární kombinace pozorovaných hodnot
- **nevychýlený** (vyrovnaný) – při opakovaném odhadu je střední hodnota odhadu identická se skutečnými parametry

Př. Nevychýlenost (vyrovnanost) a přesnost – (nejlepší) – model terče



- nepřesná (vychýlená) s nízkou variabilitou

- přesná (nevychýlená) s velkou variabilitou

- přesná (nevychýlená) s nízkou variabilitou
- **nejlepší odhad**

Hledáme nejlepší a nevychýlený odhad – nejlepší a nejpřesnější pušku ze tří.

⇒ použít metodu **BLUE** – metoda odhadu nejmenších čtverců s <u>pevnými efekty</u>.

Odhady nejmenších čtverců jsou nejlepší také proto, že mají nejmenší varianci chyb. Tedy, že se v průměru u všech možných odhadů nejméně odchylují od skutečných hodnot.

Předpověď (*prediction*)

Skutečná plemenná hodnota se předpovídá, protože je predikcí náhodného efektu.

BLUP - ***B****est* ***L****inear* ***U****nbiased* ***P****rediction*

- **n**ejlepší **l**ineární **n**evychýlená **p**ředpověď **NLNP** (metoda nejmenších čtverců)
- metoda odhadu nejmenších čtverců náhodných nebo smíšených modelů

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

X, **Z** – incidenční matice, udávající, které efekty jsou obsaženy v pozorování

b – vektor obsahující všechny fixní efekty (fixní genetické rozdíly a systematické vlivy prostředí)

u – vektor všech náhodných efektů (stádo, rok, sezóna); obsahuje také OPH

e – náhodné nesystematické zbytkové efekty

- využití metody BLUP – pro OPH (např. metody „současné porovnání dcer s vrstevnicemi", „porovnání dcer se stádovým průměrem")

Způsob řešení pro výběr odhadců je mnoho. Ve šlechtění se v současné době využívá metoda nejmenších čtverců (*least square* – LS) nebo zobecněných nejmenších čtverců (*generalized least square* – GLM), metoda maximální věrohodnosti (*maximum likelihood* – ML) či její modifikovaná metoda restringované maximální věrohodnosti (REML).

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

únor '06