

**Česká zemědělská univerzita v Praze**

**Fakulta agrobiologie, potravinových a přírodních zdrojů**

**Katedra genetiky a šlechtění**



**Včlenění skupin neznámých předků do předpovědi  
plemenných hodnot**

**Diplomová práce**

**Jana Grešová**

**Biotechnologie a šlechtění zvířat**

**doc. Ing. Luboš Vostrý, Ph.D.**

**© 2019 ČZU v Praze**

### **Čestné prohlášení**

Prohlašuji, že svou diplomovou práci "Včlenění skupin neznámých předků do předpovědi plemenných hodnot" jsem vypracovala samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autorka uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 12.4.2019

---

## **Poděkování**

Děkuji svým blízkým a všem přátelům, kteří mi při psaní práce a celém studiu byli oporou. Dále děkuji panu profesoru Ing. Josefu Příbylovi, DrSc., který mě uvedl do světa šlechtění zvířat a vedoucímu práce panu docentu Ing. Luboši Vostrému, Ph.D.

# Včlenění skupin neznámých předků do předpovědi plemenných hodnot

## Souhrn

V rámci této práce byly prostudovány postupy, které by mohly vést k lepším odhadům plemenných hodnot. Předpokladem bylo, že způsob sestavení genetických skupin může značně ovlivňovat výsledné odhady plemenných hodnot. Genetické skupiny jsou vytvořené skupiny zvířat, které seskupují koncová zvířata v rodokmenu, jejichž rodiče jsou neznámí. K těmto neúplným rodokmenům může docházet díky chybějícím údajům, importu zvířat či využití zahraničního plemenného materiálu, ke kterému nejsou v České republice dostatečné podklady.

Cílem této práce bylo odhadnout plemenné hodnoty s využitím různých možností sestavení genetických skupin a porovnání těchto nových odhadů s odhady bez genetických skupin. K dispozici byly údaje celkem od 8459584 jedinců, které sloužily jako podklady pro stanovení skupin neznámých rodičů. Metody začlenění genetických skupin byly zkoumány na souboru odregresovaných plemenných pro dlouhověkost s celkem 6061 záznamy užitkovostí. Byly zkoumány dva modely možného sestavení genetických skupin neznámých rodičů. Tyto skupiny sestaveny podle tří kritérií. Prvním kritériem bylo plemeno posledního známého jedince, dále jeho pohlaví, a nakonec rok narození. Na základě těchto údajů bylo sestaveno různé množství genetických skupin. Pro první model to bylo celkem šest skupin o velkém počtu jedinců v každé této skupině. Pro druhý model bylo sestaveno celkem 104 genetických skupin. Genetické skupiny ve druhém modelu se výrazně lišily počtem jedinců v nich zahrnutých. Sestavené genetické skupiny byly následně zahrnuty do rodokmenů a použity k výpočtům odhadů plemenných hodnot. Plemenné hodnoty byly odhadnuty metodou BLUP AM pomocí programu blupf90. Jako kontrolní model byl použit model s rodokmenem bez genetických skupin.

Pro porovnání výsledných odhadů plemenných hodnot byla použita metoda korelace. Výsledné korelační koeficienty se lišily v závislosti na tom, zdali byly porovnávány celkové soubory se všemi odhady plemenných hodnot či pouze odhady pro jedince se známou užitkovostí. Dále byla testována vhodnost použitých modelů Akaikovým kritériem. Jako nejvhodnější model byl dle tohoto kritéria stanoven model dvě, který obsahoval celkem 104 skupin a byl sestaven na základě plemenné příslušnosti, pohlaví a roku narození posledního známého předka.

**Klíčová slova:** plemenná hodnota, genetické skupiny, skupiny neznámých rodičů, fantom rodiče, genomické hodnocení, metafounders

# Incorporating groups of unknown ancestors into the predictions of breeding values

## Summary

In the framework of this work, there have been studied model which could lead to better estimates of breeding values. The assumption was that the method of assembling genetic groups could greatly influence the resulting estimates of breeding values. Genetic groups are created from animals whithe pedigree and whose parents are unknown. These incomplete pedigrees may be due to missing data, import of animals or use of foreign breeding material, for which there are insufficient documents in the Czech Republic.

The aim of this work was to estimate breeding values using various possibilities of assembling genetic groups and comparing these new estimates with estimates without genetic groups. Data were available from 8459584 individuals, which served as a basis for identifying groups of unknown parents. Methods of incorporating genetic groups have been investigated on a set of deregressed breeding values for longevity with a total of 6061 estimates. Two models of creating genetic groups of unknown parents were investigated. These groups are based on three criteria. The first criterion was the breed of the last known individual, then his sex and finally the year of birth. Based on these data, a variety of genetic groups were assembled. For the first model, there were a total of six groups of a large number of individuals in each group. A total of 104 genetic groups were compiled for the second model. Genetic groups in the second model differed significantly in the number of individuals included. The assembled genetic groups were subsequently included in the pedigrees and used for calculation of the estimates of breeding value s. Breeding values were estimated by BLUP AM using blupf90. As a control model, a pedigree model without genetic groups was used.

The correlation method was used to compare estimated breeding values. The resulting correlation coefficients differed depending on whether the total sets which were compared included all animals or only individuals with known performance. Furthermore, the suitability of the models used by the Akaik criteria was tested. The most suitable model according to this criterion was the model two, which contained a total of 104 groups and was based on breeding, gender and birth year of the last known ancestor.

**Keywords:** breeding value, genetics groups, unknown parent groups, fantom parents, genomic evaluation, metafounders

## Obsah

1	Úvod .....	7
2	Vědecká hypotéza a cíle práce .....	9
3	Literární rešerše .....	10
3.1	Šlechtění.....	10
3.2	Plemenná hodnota .....	11
3.3	Lineární modely .....	13
3.4	Genetické skupiny .....	17
3.5	GEPH – Genomická plemenná hodnota .....	20
3.6	Zakomponování skupin neznámých rodičů do ssGBLUP.....	23
3.7	Matafounders – zakladatelé.....	25
4	Metodika.....	31
4.1	Vstupní data a jejich úprava .....	31
4.2	Rozčlenění do genetických skupin.....	31
4.3	Modelová rovnice .....	33
4.4	Tvorba parametrových souborů .....	33
5	Výsledky.....	35
5.1	Popisné statistiky údajů .....	35
5.2	Výsledky metody BLUP .....	39
6	Diskuze.....	42
7	Závěr .....	45
8	Literatura .....	46
9	Seznam použitých zkratk a symbolů.....	50
10	Samostatné přílohy .....	I

# 1 Úvod

Využití genetiky ve šlechtění hospodářských zvířat má velký význam. V současné době se ve šlechtění využívá stanovené plemenné hodnoty zvířat za účelem odhadu jejich genetického založení. Tento odhad je pak využit pro výběr zvířat do další plemenitby s cílem navýšovat genetický zisk. Genetický zisk je výsledkem odezvy na selekci a označuje také jako selekční pokrok. Pro co nejpřesnější odhady plemenných hodnot je důležité zahrnutí velkého množství dat o fenotypových vlastnostech do kontroly užítkovosti. Dnes je plemenná hodnota odhadována především pomocí BLUP Animal modelu, a to díky nástupu výkonnější výpočetní techniky, ke které došlo v osmdesátých a devadesátých letech.

Výhodou metody BLUP je, že pomocí lineárních modelů se smíšenými efekty lze zahrnout jak efekty pevné, tak efekty náhodné. B jako best znamená vytváření nejvyšší možné korelace mezi skutečnou a odhadovanou plemennou hodnotou nebo nejnižší možné předpovědi rozptylu chyby, L – lineární znamená, že odhady jsou lineární funkcí pozorování, U - unbiased vyjádřuje nevychýlený neboli objektivní odhad z hodnot pro pevné a náhodné efekty a P jako prediction neboli předpověď skutečné plemenné hodnoty (Mrode et Thompson, 2005).

Animal model (AM) neboli individuální model jedince odhaduje plemennou hodnotu pro každého jedince samostatně jako první jej popsal Henderson (1984). Odhad PH pro každé zvíře se určuje s ohledem na jeho vlastní užítkovost a současně na užítkovost jeho příbuzných anebo skupin příbuzných jedinců jako jsou rodiče, prarodiče, sourozenci, polosourozenci a potomci. Tyto příbuzenské vztahy jsou do modelu zaneseny skrz aditivně genetickou matici příbuznosti a pro každé zvíře je sestavena rovnice zvlášť. Pro takovéto vyhodnocování je třeba znalosti rodokmenů velkého množství zvířat.

Pro spolehlivé předpovědi PH je nutné znát příbuzenské vazby zvířat, nicméně toto není vždy možné. Každý zaznamenaný rodokmen totiž někde končí. Podle tradičního vyhodnocování jsou tito koncoví rodiče bráni jako nepřibuzní – to znamená, že jejich koeficient inbreedingu je roven nule. Toto ale nemusí odpovídat skutečnosti. A to zvláště pokud vezmeme v úvahu větší množství takovýchto „koncových“ zvířat v rodokmenu. V nyní využívaných odhadech PH jsou všichni tito „koncoví“ jedinci uvažováni jako na sebe nepřibuzní. To by znamenalo, že každý z těchto jedinců pochází z unikátní, nepřibuzné populace. Pokud uvažujeme aditivní účinek genů, pak se předpokládá nulová příbuznost koncových zvířat v rodokmenu. K těmto situacím může dojít například v případě ne zcela úplně zaznamenaných rodokmenů nebo chybějících či ztracených dat.

Díky současné situaci na trhu, kdy je možné poměrně snadno získat plemenný materiál ze zahraničí, pak může dojít k situacím, kdy od těchto jedinců využitých v plemenitbě nejsou dostupné rodokmenové údaje. Stejný problém nastává v případě importů zvířat.

Takovéto neúplné či chybějící rodokmeny pak mohou způsobovat zkreslení v předpovědi plemenných hodnot. Abychom předešli takovému zkreslení, tam, kde není znám genetický vztah rodičů v rodokmenu zvířete, jsou tito rodiče slučováni do takzvaných genetických skupin. Celková genetická hodnota pro každé zvíře pak zahrnuje funkci těchto genetických skupin. Tato funkce genetických skupin je specifická pro každého jedince a závisí na tom, kolik generací je tato genetická skupina vzdálena.

V současné době a s rozvojem nových technologií se do hodnocení zvířat zapojuje i genomika. Mluví se o takzvaném "genetickém hodnocení" a stanovuje se genomická plemenná hodnota (GEPH). K tomu se využívá velkého počtu genetických SNP markerů. Tyto metody jsou vylepšením dřívějších postupů. Nicméně i pro výpočty odhadů plemenných hodnot se zapojením genomických údajů je třeba znát příbuzenské vazby zvířat v rodokmenu. A tak i v tomto případě, kdy díky údajům z takzvaných čipů dochází při výpočtech k doplnění o genomické vztahy zvířat, mohou neúplné rodokmeny zkreslovat odhady PH (GEPH).



## **2 Vědecká hypotéza a cíle práce**

Cílem práce je na základě údajů získaných z kontroly užítkovosti skotu navrhnout vhodný postup pro tvorbu a následné včlenění skupin neznámých předků do modelové rovnice pro odhad plemenné hodnoty.

Vědecká hypotéza: Vhodným modelováním a začlenění skupiny neznámých předků do modelové rovnice pro odhad plemenné hodnoty dojde k přesnějšímu odhadu plemenné hodnoty jedince, která se projeví snížením reziduální chyby modelu.

## 3 Literární rešerše

### 3.1 Šlechtění

Podstatou šlechtění je vybrat pro další reprodukci rodiče s větší genetickou hodnotou, než je průměr této výchozí populace rodičů. Základem šlechtění je populační genetika a matematicko-statistické metody. Předmětem šlechtění není jedinec, ale celá populace. Hlavní užitkové vlastnosti jsou podmíněny velkým počtem genů malého účinku. Ty se nazývají polygeny a jsou základem kvantitativní genetiky.

Šlechtitelský program si lze představit jako nikdy nekončící proces, který se skládá z kontroly užitkovosti (vlastnosti, kterou chceme šlechtit), kontroly dědivosti a předpovědi plemenných hodnot (PH). Na základě těchto hodnot pak probíhá výběr dalších jedinců do plemenitby a tvorba další generace. U nové generace je opět potřeba provést kontrolu užitkovosti, stanovit plemenné hodnoty a celý proces zopakovat. Aby bylo šlechtění úspěšné, je předně třeba správně organizovat kontrolu užitkovosti, mít dostatek podkladových dat a co nejlépe předpovídat plemenné hodnoty (Mrode, 2005). Cílem tedy není jednorázové zvýšení genetické úrovně, ale její postupné a trvalé zvyšování.

#### Kvantitativní vlastnosti

I pro genetiku kvantitativních vlastností je základem mendelistická teorie dědičnosti. V tomto případě ale nepracuje přímo s jednotlivými geny, nýbrž s celými skupinami genů. U kvantitativních vlastností nejsou jednotlivé geny ani zjištěné ani měřitelné. Nicméně dohromady efekty takovýchto genů poskytují variabilitu, kterou lze vyjádřit normálním rozdělením. Na rozdíl od kvalitativních vlastností, u kvantitativních nelze přímo z fenotypu usuzovat na genotyp, ale je třeba použít řadu genetických parametrů a biometrické metody. Kvantitativní vlastnosti jsou totiž podmíněny i působením efektů prostředí (Falconer, 1960). Celková fenotypová proměnlivost se tedy dělí na proměnlivost podmíněnou genetickým založením a proměnlivost podmíněnou prostředím a je hodnocena pomocí celkového fenotypového rozptylu  $\sigma^2_P$ :

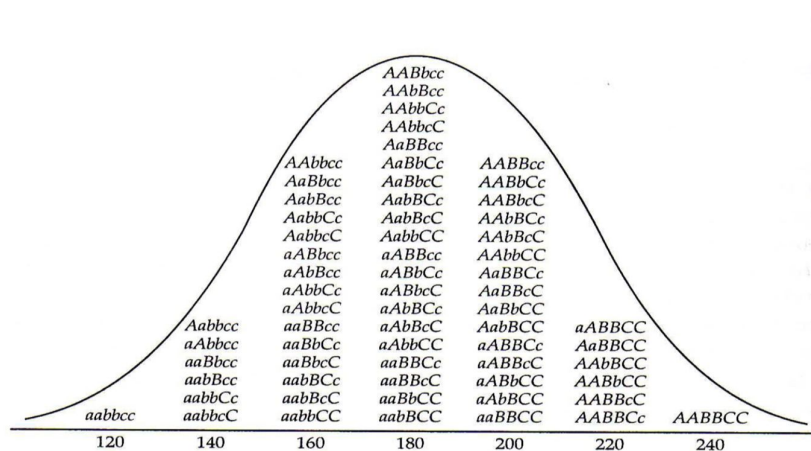
$$\sigma^2_P = \sigma^2_G + \sigma^2_E. \quad (3.1.1)$$

kde je  $\sigma^2_G$  genetický rozptyl a  $\sigma^2_E$  je rozptyl reziduální (Mrode, 2005). Genetický rozptyl lze pak dále ještě rozložit na rozptyl podmíněný aditivním (sčítavým) působením genů, rozptyl podmíněný dominancí genů a rozptyl podmíněný interakcí mezi geny.

### 3.2 Plemenná hodnota

#### Genotypová hodnota

Z rodičů na potomky se skrz generace nepřenášejí jejich celé genotypy, ale geny. Při pohlavním rozmnožování dochází k rozkladu rodičovských genotypů na gamety. Z alel, které jsou na těchto gametách umístěny, se pak tvoří nové genotypy potomků (Falconer, 1960). U vlastností s jednoduchou mendelistickou dědičností jsme schopni předpovědět pravděpodobnost narození potomků s konkrétním genotypem. Nicméně u vlastností ovlivněných kvantitativní dědičností nejsme schopni určit jednotlivé geny, které za tuto vlastnost zodpovídají. Proto u takovýchto genů odhadujeme určité měřítko, které se nazývá průměrný efekt genu. Pokud předpokládáme, že jeden rodič přispívá svým genem do genotypu potomka a příspěvek genu druhého rodiče pochází náhodně z populace, pak můžeme odhadnout průměrnou odchylku od populačního průměru jedinců. Tato odchylka se nazývá průměrný efekt genu. Součty průměrných účinků genů na všech párech alel všech lokusů konkrétní vlastnosti pak odpovídají aditivní genotypové hodnotě jedince pro tuto vlastnost (Falconer, 1960). Aditivní efekt genů se předává z generace na generaci, a tudíž je základním kamenem šlechtění.



Obr. 1: Příklad normálního rozdělení průměrného efektu genů

Skrze tyto průměrné účinky rodičovských genů lze určit průměrnou genotypovou hodnotu potomků (Falconer, 1960). Genotyp jedince se do jeho užitkovosti promítne pouze z deseti procent.

Genotyp odhadujeme na základě populačních a genetických parametrů jako jsou fenotypová odchylka, genetická odchylka, odchylka způsobená prostředím, dědivost a genetické korelace.

## Plemenná hodnota

Na rozdíl od průměrného efektu genu, který je vlastností genů a populace, plemenná hodnota je vlastností jedince a populace. Z tohoto vyplývá, že pokud odhadujeme plemennou hodnotu, je nutné specifikovat populaci, ze které tento jedinec pochází.

Předpokladem pro odhad plemenné hodnoty je kontrola užítkovosti. Její spolehlivý odhad je možný jen na základě velkého množství podkladových dat z různých zdrojů informací (Mrode, 2005). Za takovéto zdroje můžeme považovat hlavně potomky zvířat, u kterých plemennou hodnotu odhadujeme, ale také jejich rodiče, prarodiče, sourozence či polosourozence. Tyto vztahy mezi jedincem a jeho příbuznými můžeme stanovit podle úsekových koeficientů.

Pokud předpokládáme, že veškeré systematické efekty byly eliminovány, fenotypová hodnota se pak rovná  $P = G + E$ , kde  $G$  jsou genetické efekty a  $E$  jsou náhodné prostředové efekty. Genotypovou hodnotu nelze přímo odhadnout, ale můžeme odhadovat plemennou hodnotu na základě aditivně genetického efektu.  $G$  se potom rovná  $A + D + I$ . Kde  $A$  je právě aditivní efekt genů,  $D$  je dominantní efekt a  $I$  je pro interakce mezi geny (Fisher, 1918).

PH lze také použít pro předpovědi užítkovosti. Každý jedinec získává polovinu svého genetického založení od matky a druhou od otce. Na základě PH rodičů lze tedy předpovědět užítkovost budoucího potomstva, která se rovná průměru PH rodičů. Budoucí PH jednoho potomka však nelze předpovědět. A to z důvodu mendelistického výběru (MV). MV představuje náhodné kombinování genů při procesu meiózy a je definován jako odchylka průměru aditivních genových efektů jedince získaných od obou rodičů od průměrných aditivních efektů, které jsou společné všem potomkům stejných rodičů. Díky náhodnému kombinování genů do rodičovských gamet proto nelze předpovědět PH pro jediného potomka. Genetická proměnlivost potomků stejných rodičů je tedy rozdílná, protože všichni potomci nezdědí přesně ty samé geny. Avšak pokud bychom měli velký počet potomků, pak jejich průměrná hodnota MV bude rovna nule (Mrode, 2005). Z výše uvedeného tedy vyplývá, že díky zákonitostem MV je možné při zjednodušeném příkladu uvažovat příbuznost 100 % pro dva sourozence a 0 % pro jejich další dva sourozence. V průměru si tito sourozenci budou příbuzní na 50 % (Fisher, 1918).

Plemenná hodnota je základním parametrem šlechtění. U kvantitativních vlastností nelze přímo určit genotypovou hodnotu. Pomocí PH však můžeme odhadnout jeho genetické založení. Vyjadřuje se odchylkou užítkové vlastnosti od průměru populace. Tato odchylka je vyjádřena jako relativní číslo a platí pouze pro populaci, pro kterou byly dané plemenné hodnoty předpovězeny. Cílem PH je očistit výsledky užítkovosti matematickými postupy od vlivů prostředí a získat co nejlepší odhad genetické založení jedince (Mrode, 2005).

### **Systematické vlivy prostředí**

Tyto vlivy můžeme rozdělit na systematické prostředkové efekty, což jsou například pohlaví jedince, jeho věk, věk matky, pořadí vrhu, počet sourozenců ve vrhu, rok narození, roční období a podobně. Systematické efekty působí na celou skupinu zvířat a při výpočetních postupech jsme je schopni očistit (Mrode, 2005). Dále můžeme tyto efekty rozdělit ještě na efekty vnější, kam patří země, oblast a čas. Ten je velmi důležitý a do genetických hodnocení se zařazuje jako rok a roční období. Velkou roli hraje zejména v případě, že máme od jednoho zvířete více záznamů právě z rozdílných časových úseků. Efekt času v sobě může zahrnovat například i vliv kvality a druhu krmiva, který patří mezi nesystematické efekty, proto může mít stanovení tohoto efektu při hodnocení velký vliv. Mezi vnitřní prostředkové efekty patří věk jedince, pohlaví, věk matky či četnost vrhu.

### **Nesystematické vlivy prostředí**

Pokud bychom chovali všechna zvířata jedné populace v naprosto totožných podmínkách, pak by jejich fenotypovou hodnotu ovlivňovaly pouze náhodné efekty prostředí. Takovéto efekty jsou často malé a může jich být velké množství. Působí nejčastěji krátkodobě a nelze je tudíž zjistit ani kvantifikovat. Nesystematické efekty jsou součástí náhodné reziduální chyby a výpočty je nejsme schopni eliminovat (Mrode, 2005). Na každého jedince působí o neznámé velikosti a směru. Mezi tyto efekty patří například krátkodobá nemoc, říje, změna kvality krmiva a patří sem také efekt dominance genů a efekt interakce mezi geny.

Částečné eliminace těchto efektů je možno dosáhnout chovem zvířat v co nejpodobnějších podmínkách.

## **3.3 Lineární modely**

Jak již bylo řečeno, pro kvantitativní vlastnosti nejsou geny za ně zodpovědné zjistitelné ani měřitelné. Nicméně efekty většího počtu genů většinou poskytují variabilitu, kterou lze definovat normálním rozdělením.

Vychází se z předpokladu, že existuje lineární vztah mezi fenotypovou hodnotou a PH (Henderson, 1984). Pro tento předpoklad je nutné, aby fenotypové hodnoty pocházeli z normálního rozdělení.

Jelikož kvantitativní vlastnosti jsou vždy ovlivněny kombinací vlivu prostředí a genetického založení, může být toto vyjádřeno pomocí následující rovnice:

$$\text{Pozorovaný fenotyp} = \text{efekty prostředí} + \text{genetický efekt} + \text{reziduální efekt}$$

Modely využívané v kvantitativně genetických analýzách jsou ve většině případů lineární (Henderson, 1984). Skalární zápis lineární rovnice může vypadat následovně:

$$y_{ij} = \mu_i + g_i + e_{ij} \tag{3.3.1}$$

kde  $y$  vyjadřuje pozorovanou proměnnou  $j$ -tého jedince v  $i$ -té skupině,  $\mu$  je fixní efekt prostředí jako například rok narození nebo pohlaví,  $g$  součet složek genotypu zvířete, kam patří aditivní efekt genů, efekt dominance a efekt interakce mezi geny. Je vyjádřen jako náhodný efekt,  $e$  je náhodná reziduální chyba a je specifická pro  $j$ -tého jedince a  $i$ -tou skupinou (Mrode, 2005).

Jelikož efekt dominance ani efekt interakce mezi geny se nedědí do další generace, ale pokaždé jsou utvářeny znovu, při odhadování genetické hodnoty zvířete spoléháme na aditivní efekt genu. Aditivní efekt se předává z generace na generaci a tím pádem je jedinou složkou, na kterou můžeme vést selekci. Na základě tohoto pak  $g$  v rovnici výše vyjadřuje průměrný aditivní efekt genů, které jedinec získal od obou svých rodičů. Každý rodič předává na potomstvo polovinu svého genetického založení. Průměrný efekt této poloviny genetického založení, které rodič předává, pak odpovídá polovině genetické hodnoty (Fisher, 1918).

Jednoduché lineární modely můžeme rozdělit následovně:

– **Regresní modely**

Regresní modely jsou používány pro definování funkčních vztahů mezi proměnnými. Nejjednodušší regresní model se sestává ze dvou proměnných – závislé proměnné  $y$  a nezávislé proměnné  $x$ . Jednoduchá rovnice takového modelu je vyjádřena jako:

$$y = a + bx \tag{3.3.2}$$

kde  $a$  vyjadřuje konstantu a  $b$  regresní koeficient (Henderson, 1984).

– **Modely s fixními efekty**

Modely s fixními (pevnými) efekty jsou regresní modely, kde proměnná  $x$  je rovná nule nebo jedné. Rovnice takového modelu pak vypadá následovně:

$$y_{ijk} = \mu + a_i + b_j + c_k + e_{ijk} \tag{3.3.3}$$

kde  $\mu$  průměr populace,  $a$ ,  $b$ ,  $c$  jsou pevné efekty a dolní indexy těchto efektů se rovnají jejich úrovni,  $y$  je proměnná užitkovost a  $e$  je efekt reziduální chyby (Mrode, 2005).

– **Modely s náhodnými efekty**

Tyto modely jsou totožné, jako modely s pevnými efekty, pouze se liší definice proměnných v rovnici. Každá takováto proměnná pak efektem náhodného výběru ze sledované populace, jejíž průměr je roven nule. Při opakování výběru může tento náhodný výběr dosahovat jiných parametrů. U modelů s náhodnými efekty se také odhadují rozptyly,

jako například genetický rozptyl, který je základem pro odhady genetických parametrů dané populace.

- **Smišené modely pak obsahují jak efekty pevné, tak náhodné.**

Maticový zápis může vypadat například takto:

$$y = Xb + e \quad (3.3.4)$$

kde  $y$  je vektor pozorovaných hodnot – vektor dat,  $X$  je strukturní matice,  $b$  je vektor odhadovaných parametrů,  $e$  je vektor náhodných reziduálních chyb (Mrode, 2005).

$$\begin{bmatrix} 144 \\ 152 \\ 152 \\ 148 \\ 105 \\ 158 \\ 118 \\ 108 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ a_1 \\ a_2 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{131} \\ e_{221} \\ e_{222} \\ e_{223} \\ e_{231} \end{bmatrix}$$

Obr. 2: Příklad konkrétního jednoduchého maticového zápisu pro malé množství dat. V praxi mohou takovéto matice dosahovat obřích rozměrů.

Odhad efektů  $b$  je možný s použitím metody nejmenších čtverců, kdy součet čtverců odchylek pozorování od odhadnutých středních hodnot podskupin buněk ( $Xb$ ) je minimální. To lze vyjádřit jako že  $(y - XB)'(y - XB)$  se rovná minimální hodnotě. Tato metoda se nazývá BLUE - Best Linear Unbiased Estimation (Mrode, 2005).

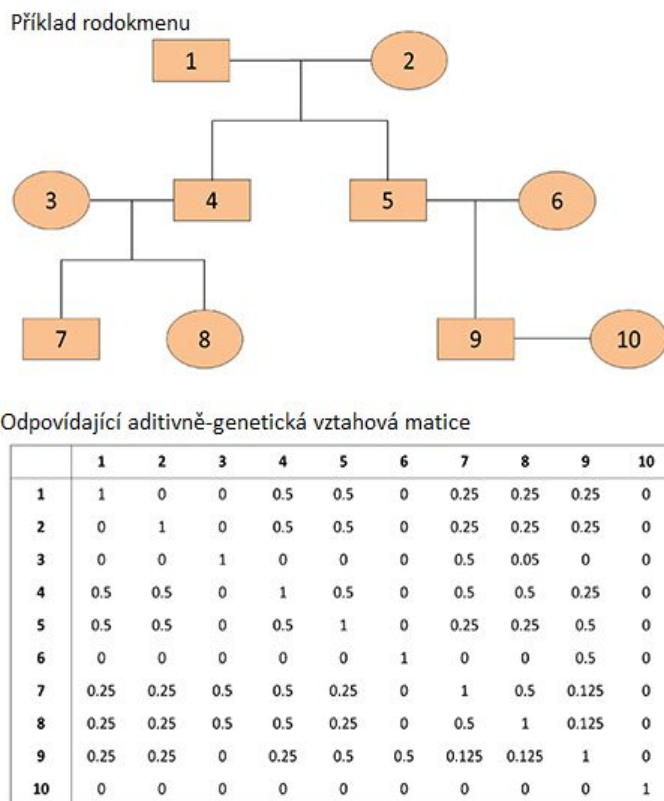
## BLUP

V praxi probíhá předpověď PH pomocí metody BLUP (Best Linear Unbiased Prediction) a to systémem rovnic lineárních modelů. Tyto rovnice jsou modifikované rovnice metody zobecněných nejmenších čtverců a momentálně se jedná o nejrošířenější metodu předpovědi PH.

## Animal model

Tento model nyní tvoří základ předpovědi plemenných hodnot. PH každého jedince se získává na základě informací o všech jeho příbuzných jedincích a stanovuje se pro celou hodnocenou populaci. Vzájemná příbuznost jedinců se do výpočtů zahrnuje pomocí aditivně-genetické matice příbuznosti  $A$ . Jedná se o čtvercovou matici jejíž velikost odpovídá počtu jedinců, kteří vstupují do hodnocení. Diagonální prvky této matice se  $1 + F_z$ , kde koeficient  $F_z$  vyjadřuje inbreeding  $i$ -tého jedince. Prvky mimo diagonálu vyjadřují stupně příbuznosti mezi  $i$ -tým a  $j$ -tým jedincem (Wright, 1922). Pokud je jedinec zcela nepříbuzný, pak se jeho  $F_z$  rovná nule a  $a_{ii}$  se rovná jedné. K takovéto situaci dojde i v případě, že rodiče daného jedince

nejsou známy. Takový jedinec se pak uvažuje jako zcela nepříbuzný. Stejně tomu tak je i pokud je neznámý pouze jeden rodič. Jedinci pocházející z takovýchto spojení jsou pak považováni za zcela nepříbuzné. Matice příbuznosti A má variančně-kovarianční strukturu mezi aditivně-genetickými hodnotami jedinců (Mrode, 2005).



Obr. 3: Příklad rodokmenu a k němu odpovídající aditivně-geentické vztahové matice (Sebastiani, 2014).

Matici A lze vyjádřit jako  $A = TDT'$  (Henderson, 1976; Thompson, 1986)

Kde D je diagonální matice a T je spodní triangulární matice.

Ze vztahové matice A je dále třeba vypočítat inverzní matici, jejíž prvky vyjadřují koeficient příbuznosti mezi danými jedinci a všechny prvky, které leží nad diagonálou, jsou rovny nule. Tato matice uvažuje pouze vztahy rodič – potomek. Můžeme tedy říci, že sleduje tok genů z jedné populace do druhé. Matice D je pak maticí pro mendelistický výběr.

Pro výpočty metody BLUP je potřeba sestavit inverzní matici A. Samotnou vztahovou matici tedy není třeba vytvářet, lze pracovat už rovnou s její inverzí. Postupy pro tvorbu inverzní matice  $A^{-1}$  vypracovali Henderson (1976), Quass (1976, 1984, 1955) a Meuwissen a Luo (1992).



### 3.4 Genetické skupiny

Westel (1988) použila pro zpřesnění genetického hodnocení zvířat skupinové genetické efekty, které představují průměry genetických příspěvků „fantom“ rodičů zvířete. Takzvaní fantom rodiče jsou vyjádřením neznámých rodičů nebo jen jednoho neznámého rodiče zvířete, které má zaznamenanou užitkovost a je součástí genetického hodnocení. Termín „fantom“ je použit pro vyjádření toho, že toto zvíře samo není cílem zájmu, ale je důležité ho použít pro výpočet. Neznámí rodiče ze stejného časového období pak vytvářejí genetické skupiny, jejichž průměr je příspěvkem do genetického hodnocení. Díky různým hlediskům pro selekci samců a samic se pak využívají různé postupy a čtyři paralelní soubory skupin mohou představovat otce synů, matky synů, matky dcer a otce dcer. Možností je také použít pouze dva soubory skupin – pro otce a matky, a to na základě pohlaví a generačního intervalu, neboť matka syna a matka dcery nemusí být nutně součástí stejné genetické skupiny i v případě, že tento syn a dcera se narodily ve stejném roce.

Neznámí rodiče se získají při vytváření rodokmenu tak, že jsou vyhledáni všichni otci a přidání do seznamu užitkovostí, stejně tak matky. Tento postup je pak opakován s přidáváním stále dokola dalších generací předků, až v rodokmenové složce nejsou další předci. A v tomto případě jsou tito neidentifikovaní rodiče považováni za neznámé, základní. Pokud rodič zvířete nemá přiřazenou užitkovost, je považován za zakládajícího i přesto, že je uveden v rodokmenové složce.

Quass (1988) navrhl vyjádření chybějících neznámých rodičů pomocí začlenění pevného efektu Q, kde využil Hendersenův algoritmus (Henderson, 1975) pro sestavování inverzní rodokmenové matice  $A^{-1}$  na základě toku genů z předků na potomky. Zvířatům přiřazoval buď číslo rodiče, nebo číslo genetické skupiny, ze které pochází tento neznámý rodič. Sestavené skupiny musí pro nezkreslené využití ve výpočtech obsahovat dostatečný počet jedinců. Rodiče jsou do skupin rozřazováni na základě vlastností svých potomků, jako je například datum narození tohoto potomka, stáj, země původu, rok započtení využití v chovu (Quass & Pollak, 1981) či podle časových období generačních intervalů (Westell, 1988).

Rovnice smíšeného lineárního modelu obsahující matici Q pro genetické skupiny neznámých rodičů, jak jej popsal Quass (1988):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{ZQg} + \mathbf{Zu}^* + \mathbf{e} \quad (3.4.1)$$

kde  $\mathbf{y}$  je vektor pozorování,  $\mathbf{X}$  a  $\mathbf{Z}$  jsou matice přiřazující pozorování pevným a náhodným efektům,  $\boldsymbol{\beta}$  je vektor pevných efektů,  $\mathbf{Q}$  matice přiřazující genetické skupiny,  $\mathbf{g}$  vektor pevného efektu skupin,  $\mathbf{u}^*$  je náhodný genetický efekt jedince a  $\mathbf{e}$  je vektor náhodných chyb.

Zkomponování matice Q do modelu smíšených rovnic:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}ZQ \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}ZQ \\ Q'Z'R^{-1}X & Q'Z'R^{-1}Z & Q'Z'R^{-1}ZQ \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u}^* \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ Q'Z'R^{-1}y \end{bmatrix}$$

Obr. 4: Rovnice smíšeného modelu s maticí Q pro pevné efekty genetických skupin (Quass, 1988)

$$[P_b : P] = \begin{matrix} & a & b & c & d & e & : & A & B & C & D \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} .5 & .5 & 0 & 0 & 0 & : & 0 & 0 & 0 & 0 \\ 0 & 0 & .5 & 0 & 0 & : & .5 & 0 & 0 & 0 \\ 0 & 0 & 0 & .5 & .5 & : & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & : & 0 & .5 & .5 & 0 \end{bmatrix} \end{matrix}$$

Obr. 5: Matice přiřazující nenulovou hodnotu (0,5) rodičům potomka. Tento rodič může být známý (P) i neznámý (P<sub>b</sub>). V případě neznámého rodiče je potomku přiřazena hodnota ve sloupci odpovídající dané genetické skupině, což je znázorněno malými písmeny.

V případě známého rodiče hodnota je přiřazena hodnota ve sloupci označenými velkými písmeny (Quass, 1988)

Při tomto způsobu vyhodnocení a začlenění skupin se předpokládá, že zakládající, neznámí rodiče jsou náhodným výběrem z velkých, náhodně se pářících populací. Neuvažuje se žádný inbreeding.

$$Q_b = \begin{matrix} & g_1 & g_2 \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \end{matrix}$$

Obr. 6: Matice  $Q_b$  je matice přiřazující zakládající, neznámá, zvířata k populačním průměrům svých genetických skupin. (Quass, 1988)

Matice  $Q$  je pak získána postmultiplikací  $(I - P) - 1P_b$  matice obsahující příspěvky očekávaných frakcí genů přenesených z předků na potomky a sumou příspěvků zvířat do jednotlivých genetických skupin.

$$Q = \begin{matrix} & g_1 & g_2 \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} .5 & .5 \\ .25 & .75 \\ .5 & .5 \\ .375 & .625 \end{bmatrix} \end{matrix}$$

Obr. 7: Matice  $Q$  vyjadřuje pro každé zvíře očekávané množství genů odvozené od základní populace (Quass, 1988)

Podobného způsobu vyjádření příspěvků základních populací se užívá při odhadování plemenných hodnot kříženců různých plemen (Dillard, 1980).

Algoritmus pro vytvoření  $A^*$  je odvozen od algoritmu pro vytvoření  $A^{-1}$  dle Hendersona. Každému zvířeti je přiřazeno buď číslo rodiče, nebo číslo genetické skupiny, ze které daný rodič pochází. Zařazení rodičů do skupin je na základě informací o potomcích a pohlaví neznámého rodiče. Daná genetická skupina musí obsahovat dostatečný počet jedinců, jinak jsou výsledky zkresleny. Příspěvek dané genetické skupiny se získává průměrováním genetických příspěvků neznámých rodičů (Westell, 1988).

### 3.5 GEPH – Genomická plemenná hodnota

Tradiční předpovědi plemenných hodnot kombinovaly pouze fenotypové hodnoty a pravděpodobnost, že geny jsou identické po předcích s použitím rodokmenových informací. S nástupem dokonalejších laboratorních metod se do klasického vyhodnocování předpovědí metodou BLUP zahrnují i genomické informace. K měření genetické podobnosti s použitím genetických molekulárních markerů typu SNP pro množství lokusů napříč genomem může být dosaženo větší přesnosti, než s pouhými informacemi na základě rodokmenu (VanRaden, 2008). Hlavní využití genomické informace v genetickém hodnocení je ve zpřesnění vzájemných vazeb mezi zvířaty. Zahrnuje také informace o náhodném Mendelistickém výběru (MV) (Hayes, 2009).

Cílem vyhodnocování je vysvětlit variabilitu studovaných charakteristik a proč se zvířata nebo skupiny od sebe odlišují (Příbyl, 2012).

Genomické vztahy lépe odhadují míru chromozomových segmentů, které jsou sdíleny mezi jedinci. Díky využití čipů s vysokou hustotou je možné identifikovat geny identické stavem (identity by state), které mohou být sdíleny skrz společného předka a nezaznamenány v rodokmenu.

Genomická selekce pracuje více se zdokonalováním vztahů mezi zvířaty, než se stanovováním efektů majorgenů (Mizstal, 2010). Pro začlenění genomických informací do předpovědí plemenných hodnot pomocí metody BLUP byla navržena metoda Multi-step. Multi-step metoda se sestává z těchto kroků – regulerní vyhodnocení pomocí animal modelu, stanovení genomických efektů pro relativně malý počet genotypovaných zvířat a stanovení genomických plemenných hodnot pomocí selekčního indexu (Mizstal, 2009). Nicméně tento postup je díky několika krokům více náchylný ke zkreslení a nepřesnostem. Proto byla navržena metoda Single-step (ssGBLUP), která je založena na lineárním smíšeném modelu. Používá vztahovou matici, kombinuje rodokmen s genomickými vztahy a eliminuje tím kroky komplexní metody multistepu, který je náchylný k chybám. (Mizstal, 2010).

Začlenění genomických koeficientů do čitatele matice příbuznosti umožňuje při stanovování odhadů plemenných hodnot použití současně fenotypových, rodokmenových i genomických hodnot. Při použití metody single-step je však nutné, aby informace na základě rodokmenu i genomické informace byly kompatibilní (Forni, 2011).

Mizstal (2009) navrhl úpravu čitatele vztahové matice  $A$ , která zahrnuje jak vztahy mezi zvířaty založené na rodokmenu, tak rozdíl mezi rodokmenově založenými a genomickými vztahy. Single step modifikuje čitatele vztahové matice  $A$  na matici  $H = A + A_{\Delta}$ , kde  $A_{\Delta}$ ,

zahrnuje odchylky od rodokmenových vztahů způsobené mendelistický výběr (Mizstal, 2009). Pracuje přímo s celým, neselektovaným základním souborem záznamů užitkovosti. Dokáže pracovat i se znaky s nižší dědivostí (Chen, 2011) i pro model smíšených rovnic (Aguilar, 2011). Vztah  $G - A$  vyjadřuje rozdíl mezi genomickými a rodokmenově založenými vztahy.

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{bmatrix} \quad (3.5.1)$$

kde dolní indexy vyjadřují negenotypovaná (1) a genotypovaná (2) zvířata,  $G$  je genomická vztahová matice (Aguilar, 2010).  $H$  je rodokmenově-genomická matice. V ní je genomická informace přenesená do kovariancí mezi všechna negenotypovaná zvířata (Legarra, 2009). Legarra také uvádí, že matice  $A$  nemusí být plně koherentní, protože matice  $G$  také zahrnuje informace o předcích genotypovaných zvířat. Například, genotypování jedinci  $M$  a  $B$  nemají v rodokmenové matici  $A$  žádný vztah, ale mohou si být příbuzní v genomické matici  $G$  díky společnému předku (který ale už není obsažen v matici  $A$ ). Nebo může posílit vztah mezi jedinci  $M$  a  $B$  (Legarra, 2009).

Pokud je matice  $G$  inkopatibilní s maticí  $A$ , váha rodokmenových a genomických informací bude nesprávná. Různé způsoby setavení matic  $G$  použitých v genomickém hodnocení Aguilarem (2010) vyústily v různá měřítka (váhy) a přesnosti plemenných hodnot. Odhady aditivních odchylek s použitím matice  $G$  mohou být daleko větší, než s použitím matice  $A$ . Různé  $G$  může vést k různým přesnostem plemenných hodnot (Aguilar, 2010). Tyto rozdíly mohou nastat díky nesprávnému rozvržení (měřítku) matice  $G$  vzhledem k matici  $A$ .

Pro zjednodušení výpočtů se využívá výpočtu inverzní matice  $H$ , jak popsal Aguilar (2010) či Christensen and Lund (2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}, \quad (3.5.2)$$

kde  $\mathbf{A}_{22}^{-1}$  je inverze rodokmenové matice pouze pro genotypovaná zvířata. Nicméně tento výraz je ještě nutné upravit, protože takto sestavená matice nemůže být vždy invertována, například:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \quad (3.5.3)$$

kde  $F$  je korekce genomických vztahů pomocí průměru matice  $G$  pouze pro genotypovaná zvířata,

$$F = ((1 - w)G + w A_{22})^{-1} - A_{22}^{-1} \quad (3.5.4)$$

kde  $w$  je váha rodokmenových vztahů od 0 do 1 (Christensen and Lund, 2010) a  $A_{22}$  je rodokmenová matice pouze pro genotypovaná zvířata.

Forni (2011) použil genomickou vztahovou matici takto:

$$G = \frac{(M - P)(M - P)'}{2 \sum_{j=1}^m P_j(1 - P_j)}, \quad (3.5.5)$$

kde  $M$  je matice výskytu alel SNP velikosti  $n, m$  pro  $n$  genotypovaných zvířat a  $m$  lokusů genetických markerů; elementy matice jsou čísla druhých alel daného zvířete na lokusu ( $j$ ) mající hodnoty 0, 1 nebo 2.  $P$  je matice průměrů SNP vyskytujících se napříč určitým lokusem se sloupcem  $j$  obsahující hodnoty  $2 \times q_j$ ; kde  $q_j$  je průměrná frekvence druhé alely pro SNP ve sloupci  $j$  v původní neselektované populaci. Nicméně tuto frekvenci je obvykle velmi obtížné stanovit, proto se často používá alelová frekvence současné populace nebo konstantní hodnota 0,5 pro všechny lokusy (Forni, 2011).

VanRaden (2008) uvádí (oproti použití v Quass, 1988), že alelové frekvence použité v  $P$  by měly být spíše z neselektované základní populace než z populací, které se objevují po selekci nebo inbreedingu. Použití dřívější nebo pozdější základní populace může vést k bližším nebo vzdálenějším příbuzenským vztahům a k většímu nebo menšímu inbreedingu. Odčítání  $P$  při výpočtu genomických vztahů zvýhodňuje více alely vzácnější než ty časté. Stejně tak genomický koeficient inbreedingu je větší, pokud je jedinec homozygotní ve vzácné alele, než pokud je homozygotní v alelách, které se vyskytují často.

Genomickou vztahovou matici lze sestavit pomocí matice mající genotypové informace pro každého jedince a markerů (VanRaden, 2008). Každý genotyp je odchylka od marker specifického populačního průměru, který je vypočítán použitím alelových frekvencí populace.

Obecná shoda napříč studiemi o genomické selekci je, že je tu zisk v předpovědi přesnosti pro mladé neproověřené býky, když je genomická informace zahrnuta, ve srovnání s tradičním vyhodnocením s použitím pouze rodokmenových informací (VanRaden, 2008 a

Miztal, 2009). Důvodem je pravděpodobně zdokonalení předpovědi odchylek mendelistického výběru mezi blízkými příbuznými v matici G.

Forni (2011) napsala, že zkreslení v genomických předpovědích může být spojeno s nesprávnými váhami polygenních a genomických komponentů. Kombinované rodokmenově genomické vztahové matice poskytují přirozenou cestu, jak vyvážit oba komponenty pro optimální předpovědi. Navíc single-step genomické předpovědi eliminují počet předpokladů a parametrů, které jsou vyžadované pro multi-step metody a možná poskytují přesnější předpovědi pro mladá zvířata. Single-step metoda může být jednoduše rozšířena na multi-trait analýzy a zvládne velké množství genomických informací. Rozšíření pro zahrnutí dalších distribucí markerů jako velké QTL nebo majorgeny je také možné. Nicméně výpočetní snahy mohou být dlouhodobý problém, protože genomické matice potřebují být vytvořeny a invertovány (Forni, 2011).

### 3.6 Zakomponování skupin neznámých rodičů do ssGBLUP

Při zahrnování skupin neznámých rodičů do single-step genomického BLUP hodnocení musí skupiny rodičů obsahovat i příspěvky z genomických vztahů, nikoliv jen z rodokmenové matice A, jinak dochází ke zkreslení předpovědi. Zkreslení se lze vyhnout několika způsoby, jako je modifikování smíšené modelové rovnice vhodnými změnami G a  $A_{22}$ , zahrnutím skupin neznámých rodičů jako samostatných efektů (ZQs, což by potenciálně značně zvyšovalo čas výpočtu, pokud by byl velký počet skupin), přidáním dalšího efektu jedince anebo dalšího polygenního efektu (Miztal, 2013).

Začleněním skupin neznámých rodičů do genomického hodnocení se zabýval Miztal (2013), který navrhl několik různých řešení tohoto výpočtu. Pokud jsou skupiny neznámých rodičů začleněny jako pevné efekty ZQs, rovnice smíšeného modelu pro ssGBLUP vypadá následovně:

$$\begin{bmatrix} \mathbf{Q}'\mathbf{Z}'\mathbf{Z}\mathbf{Q} & \mathbf{Q}'\mathbf{Z}\mathbf{Z} \\ \mathbf{Z}'\mathbf{Z}\mathbf{Q} & \mathbf{Z}'\mathbf{Z}+\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{s}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}'\mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}. \quad (3.6.1)$$

kde matice Q přiřazuje zvířata do genetických skupin, matice Z přiřazuje pozorování jedincům,  $\mathbf{H}^{-1}$  je inverzní rodokmenově-genomická matice.

A dále po QP transformaci:

$$\begin{bmatrix} \mathbf{Z}'\mathbf{Z}+\mathbf{H}^{-1} & -\mathbf{H}^{-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{H}^{-1} & \mathbf{Q}'\mathbf{H}^{-1}\mathbf{Q} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'\mathbf{y} \\ \mathbf{0} \end{bmatrix}. \quad (3.6.2)$$

Začlenění skupin neznámých rodičů jako pevného efektu by mohl být problém v případě, že by se jednalo o rozsáhlý soubor. V takovém případě by muselo být sestaveno velké množství skupin, což by znamenalo velký počet efektů a vyšší výpočetní čas. Problémem při použití multiple-trait modelu by také mohl být nedostatek údajů k některým vlastnostem, což by mohlo vést ke špatnému stanovení skupin rodičů a tím ke zkreslení celého hodnocení.

V případě velkých populací, kde je genotypovaná pouze malá část zvířat, je distribuce skupin v rámci a mimo tato genotypovaná zvířata jen minimálně odlišná, a proto je v takovémto případě možné ignorovat příspěvky skupin v genomickém vyhodnocení. Pro zjištění, zdali můžeme příspěvky ignorovat, se dá využít porovnání vyhodnocení pomocí regulerního BLUP a ssGBLUP modelu (Mizstal, 2013). Rovněž je vhodné ignorovat příspěvky genetických skupin v případě, že hodnocení, které zahrnuje genetické skupiny, snižuje spolehlivost odhadovaných plemenných hodnot (Phocas a Laloe, 2004). Příspěvky mohou být ignorovány i v situaci, kdy modifikujeme matici  $A_{22}$ ,  $G$  nebo obě tak, že jejich příspěvky do genetických skupin se vyruší (Mizstal, 2013).

Mizstal (2014) rozdělil matici  $Q$  do dvou částí –  $Q_1$  a  $Q_2$ . Matice přiřazuje vztahy mezi negenotypovanými zvířaty a neznámými rodiči do genetických skupin na základě rodokmenové matice  $A$ . Matice přiřazuje vztahy do genetických skupin jen pro genotypovaná zvířata na základě genomické matice  $G$ .

Možností je také zahrnutí skupin neznámých rodičů za předpokladu dvou aditivních efektů:

$$\mathbf{y} = \mathbf{Zu}_p + \mathbf{Za}_g + \mathbf{e}, \quad (3.6.3)$$

kde  $u_p = Qs + a_p$  je polygenní efekt, který zahrnuje efekt skupin neznámých rodičů,  $a_g$  je další genetický efekt, který představuje genetickou informaci, ale ignoruje efekt skupin. Pokud může být model přizpůsoben softwarem, přidání dalšího aditivního efektu by neměl být problém (Mizstal, 2013).

V případě ovcí nebo hovězího dobytka může nastat případ, že rodokmeny zvířat jsou nekompletní (někteří rodiče u jedinců nejsou přiřazeni), i když jsou tyto rodiče v rodokmenu. Místo těchto rodičů je pak použito efektů skupin neznámých rodičů, které jsou utvořeny na základě očekávaného čísla generace nebo jsou funkcí roku narození jedince. V tomto případě by, na rozdíl od rodokmenové matice  $A$ , matice  $G$  byla zcela kompletní a popisovala by správně vztahy mezi zvířaty (VanRaden, 2008). Matici  $A$  by pak na základě vztahů v matici  $G$  bylo možné zrekonstruovat, ačkoliv velmi složitě a pouze pokud bude mít jedinec dostatečný počet genotypovaných příbuzných jedinců. Další možností je zahrnout chybějící rodiče jako příbuzné a to tak, že všechna zvířata tvořící konkrétní skupinu neznámých rodičů jsou



předpokládána jako příbuzná, sdílející společný průměrný vztah, který je vyjádřen jako průměrný inbreeding rodičů ze stejného časového období. Jedinci narození ve stejném čase jsou spíše příbuzní na populaci podobnou než na celou základní populaci, která se náhodně páří. Průměrný inbreeding potomstva, který se rovná polovině průměru vztahu rodičů, je potom stejný pro zvířata s kompletními i nekompletními rodokmeny (VanRaden, 1992). VanRaden také navrhl algoritmus pro stanovení většiny chybějícího inbreedingu (VanRaden, 1992) nebo může být inbreeding odvozen z matice G (VanRaden, 2008).

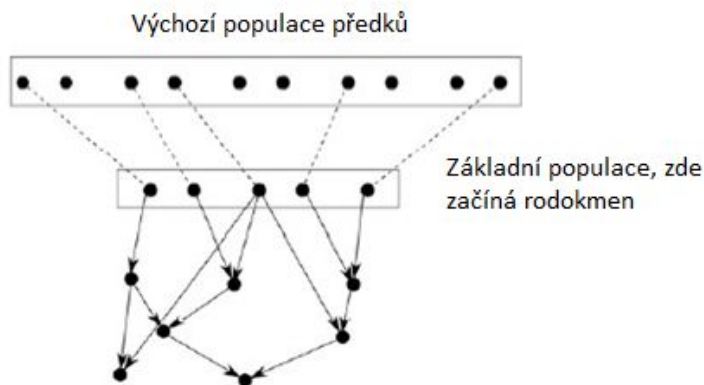
Skupiny neznámých rodičů je třeba stanovit i ve vyhodnocování plemenných kříženců, kde musí reflektovat jak rozdíly mezi plemeny, tak chybějící rodokmeny. Jedinci z různých plemen jsou považováni za vzájemně nepříbuzné (Harris et Johnson, 2010).

Nekompletní rodokmeny také mohou způsobit to, že pokud máme u některých zvířat dlouhé a plné rodokmeny a u jiných někteří předci chybí, budou hodnoty v matici G v průměru menší, než  $A_{22}$  pro zvířata s dlouhými rodokmeny a v průměru větší, než  $A_{22}$  pro zvířata s krátkými (nekompletními) rodokmeny. Odhady plemenných hodnot pak budou zkresleny nahoru pro krátké rodokmeny nebo dolů pro dlouhé rodokmeny (Chen, 2011). Například pro zvířata s dlouhými rodokmeny pak budou diagonální elementy  $A_{22}^{-1}$  negativní, protože  $(G^{-1} - A_{22}^{-1})$ .

### 3.7 Matafounders – zakladatelé

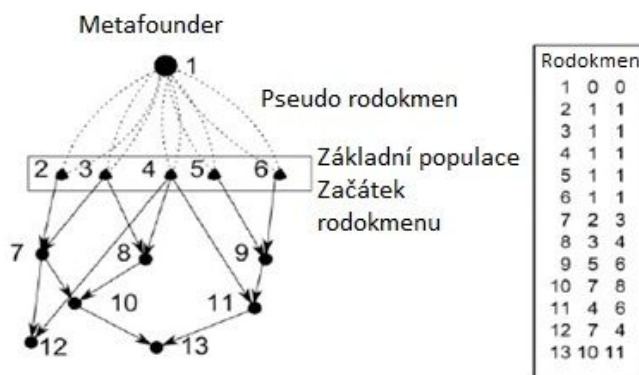
V současné době praxe genetického vyhodnocování s použitím genetických skupin nebo skupin neznámých rodičů předpokládá, že jedinci v různých základních populacích mají různé průměrné hodnoty a tyto hodnoty jsou stanoveny jako pevné efekty. Tyto zahrnují rok narození, zemi původu, pohlaví a způsob selekce (Legarra, 2015). Kennedy (1991) však poukázal na to, že genetické skupiny nejsou předpokládány zcela přesně, že k sobě mohou být příbuzné a že by měla být započítávána i redukce rozptylu díky genetickému griftu a selekci. Pomocí využití molekulárních markerů se ukazuje stále více případů nalezených vztahů tam, kde se původně dle rodokmenu mělo jednat o nepříbuzná zvířata (VanRaden, 2011). Genomické vztahy nejsou závislé na znalosti rodokmenu a jsou více přesné, neboť zvažují realizované, nikoliv očekávané vztahy (VanRaden, 2008).

Hypotéza, že v základní populaci jsou zakladatelé nepříbuzní a tato populace je odvozena z velmi velké populace předků, není pravděpodobně správný předpoklad. Rozporují to také informace na základě markerů, kdy zvířata, která se zdají nepříbuzná, mají shodné alely na markerech (Legarra, 2015).



Obr. 8: Základní populace na základě rodokmenu a odhadovaná populace předků, ze kterých je tato základní populace odvozena (Legarra, 2015).

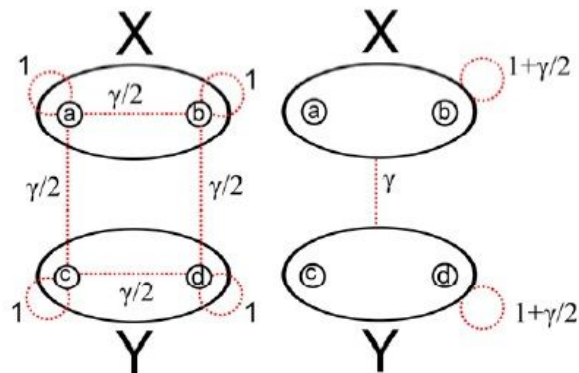
Legarra (2015) představil koncept metafounderů, kde metafounder je pseudo jedinec, který „nahrazuje“ generaci předků či genetické skupiny a skupiny neznámých rodičů. Tento pseudo jedinec může být považován zároveň za matku i otce.



Obr. 9: Metafounder se základní populací a rodokmenem (Legarra, 2015).

Metafounder představuje konečný zdroj gamet – při náhodném výběru dvou gamet, mají tyto gamety vztah  $\gamma/2$ , který se rovná  $1/2 \cdot N_e$ , kde  $N_e$  je efektivní velikost populace. Metafounder má vlastní vztah  $\alpha_{11} = \gamma$  a koeficient inbreedingu  $F_i = \alpha_{11} - 1$ , tudíž  $F_i = \gamma - 1$ , který bude obvykle negativní. To znamená odklon od Hardy-Weibergovy rovnováhy. Negativní koeficient inbreedingu vyjadřuje přebytek heterozygotů, tedy že ve většině případů, kdy vybíráme dvě náhodné gamety, budou tyto gamety rozdílné. Předpokládáme tedy obrovskou velikost zdroje gamet. Pokud by například  $\gamma = 0$ , pak by  $F_i = -1$  a dvě náhodně vybrané gamety by byly vždy různé a nepříbuzné, což vyjadřuje nekonečnou velikost poolu gamet. Pokud by naopak například  $\gamma = 2$ , pak by  $F_i = 1$  a dvě náhodně vytažené gamety byly

vždy identické, pool sestává z jedné gamety a je plně homozygotní. Populace by v takovémto případě byla identická a kompletně inbrední (Chirstensen, 2012).



Obr. 10: Červeně jsou naznačené vztahy v základní populaci dvou jedinců X a Y. Na levo vyjádřené jako gamety, napravo jako jedinci. Jedinci obsahují gamety (a, b, c, d). Vztahy v rámci jedné gamety jsou rovny jedné, mezi dvěma gametami pak  $\gamma/2$ . Vztah jedince se sebou samým se rovná  $1 + \gamma/2$  a mezi dvěma jedinci  $\gamma$  (Legarra, 2015).

S použitím systému metafounderů se obvyklá pravidla pro výpočet příbuzenských vztahů a inbreedingu mění pouze trošku. Emik a Terrill (1949) navrhli aditivní vztahový koeficient, kdy všechna zvířata v základní populaci mají vlastní vztah roven jedné. Pak:

$$a_{ij} = 0,5 (a_{\alpha j} + a_{s j}) \text{ a } a_{ii} = 1 + 0,5 (a_{sd}) . \quad (3.7.1)$$

následně se zahrnutím metafoundera ( $\alpha_{11} = \gamma$ ):

$$a_{22} = 1 + 0,5 a_{11} = 1 + \gamma/2 . \quad (3.7.2)$$

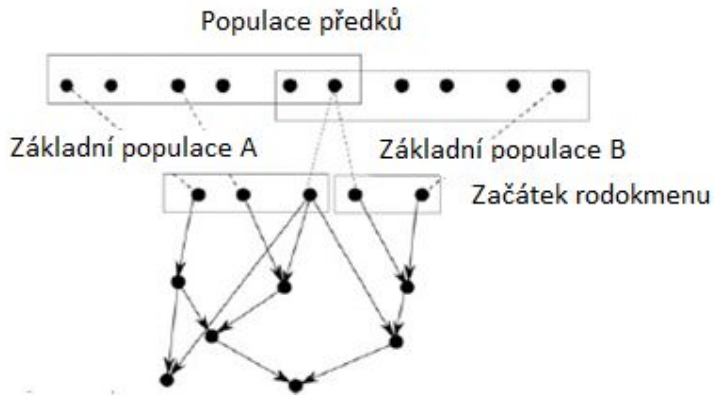
Pro jedince 1 a 2:

$$a_{12} = 0,5 (a_{11} + a_{11}) = \gamma . \quad (3.7.3)$$

Pro jedince 2 a 3:

$$a_{23} = 0,5 (a_{12} + a_{12}) = \gamma . \quad (3.7.4)$$

Z metafoundera pak pochází základní populace – např. A a B, a ty se mohou (ale nemusí) překrývat.



Obr. 11: Vyjádření dvou překrývajících se základních populací (Legarra, 2015).

Vyjádření vztahů napříč populacemi:

$$\Gamma = \begin{pmatrix} \gamma^A & \gamma^{A,B} \\ \text{symm} & \gamma^B \\ & \dots \end{pmatrix}.$$

(3.7.5)

kde  $\gamma^{A,B}$  značí překrytí populací. Populace A je z  $n_A$  gamet a populace B z  $n_B$  gamet.

$$\gamma^A = 1/n_A,$$

$$\gamma^B = 1/n_B,$$

$$\gamma^{A,B} = n_{AB} / n_A n_B,$$

(3.7.6)

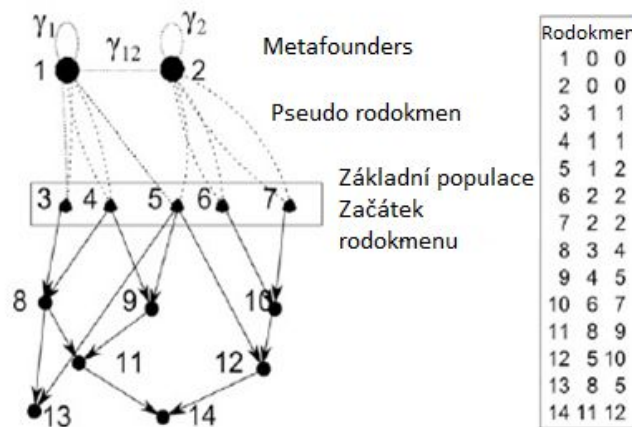
pokud se populace vlivem selekce odlišily, pak  $\gamma^{A,B}$  může být menší než nula, to znamená, že nedochází k překrytí populací. Nicméně matice by měla být pozitivní.

Za předpokladu, že každou populaci předků považujeme za metafoundera, příbuzenské vztahy mezi metafoundery mohou být vyjádřeny následovně:

$$\Gamma = \begin{pmatrix} \gamma^A & \gamma^{A,B} \\ & \gamma^B \\ & & \dots \end{pmatrix}$$

(3.7.7)

Vzájemné vztahy mezi metafoundery lze zjistit z historie nebo odvodit z genomických vztahů. Rodokmen s několika metafoundery utváří matici  $\Gamma$ . Legarra (2015) popisuje postup, kde se nejdříve invertuje  $\Gamma$ , čímž vznikne submatice  $A \Gamma^{-1}$  a následně použije Hendersonovo pravidlo s elementem  $D_{ij}$ .



Obr. 12: Vyjádření populace se dvěma vzájemně příbuznými metafoundery. Vlastní vztahové koeficienty jsou  $\gamma_1$  a  $\gamma_2$ , vzájemný vztah metafounderů je vyjádřen jako  $\gamma_{12}$  (Legarra, 2015).

Metafoundeři jsou zobecněním skupin neznámých rodičů, protože generalizovaná inverze  $\Gamma = 0$  je rovna nule a jinak pravidla pro inverzi a hodnoty  $D_{ij}$  jsou stejné. Pro  $\Gamma = 0$  dostaneme stejné  $A \Gamma^{-1}$  jako pro skupiny neznámých rodičů jak publikoval Quass (1988).

Pokud uvažujeme skupiny neznámých rodičů, dosazujeme zvířata (jako kdyby) z další základní populace, neboť tato zvířata nemají žádné příbuzenské vztahy se zbytkem zvířat. Nicméně u těchto základních populací se postupně zvyšuje inbreeding a každá takováto populace pochází z předchozí, starší populace. Například pokud je skupina neznámých rodičů z roku 2010, bude tato skupina více příbuzná na skupinu z roku 2006, než na skupinu z roku 1990. Legarra (2015) navrhuje řešit tento problém s použitím metafounderů.

Když uvažujeme metafoundery, inbreeding automaticky stoupá. Pokud k inbreedingu přidáme konstantu ( $\gamma/2$ ), odhady se nezmění a inbreeding by měl být lépe a snadněji předpověditelný. Použití metafounderů předpokládá společný genetický základ napříč všemi

základními populacemi. Metadoundeři představují limitovaný pool gamet, ze kterého jsou vytaženi zakladatelé. Vztahy mezi těmito metafoundery lze odvodit z dat markerů.

Konečnou velikostí populace se zabýval už Jacquard (1969, 1974), nicméně jeho studie nebyly brány v potaz. VanRaden (1992) použil koeficient inbreedingu pro genetické skupiny na základě koeficientu jejich vrstevníků. Namísto toho navrhuje Legarra (2015) použití genomické informace. Použití metafounderů s maticí  $\Gamma$  dovoluje spojení rodokmenu, genomických vztahů a inbreedingu (Vitezica, 2011).

## 4 Metodika

Pro výpočet a začlenění genetických skupin do předpovědí plemenných hodnot byla použita metoda sestavení genetických skupin a jejich začlenění pomocí výpočetního softwaru Blupf90 (Misztal et al., 2002).

Postup diplomové práce byl uspořádán do několika kroků. V prvním kroku byly nahrány soubory s užitkovostmi zvířat, načtení souborů s rodokmeny a byla provedena jejich úprava. Ve druhém kroku byly záznamy ke každému zvířeti rozšířeny o doplňující informace, na základě kterých se poté tvořily skupiny neznámých předků přiřazené k jedincům s neznámými rodiči. Tyto genetické skupiny byly následně v rodokmenovém souboru přečíslovány tak, aby bylo možné dále použít tyto soubory ve vyhodnocení pomocí metody GBLUP. V dalším kroku byly předpovězeny plemenné hodnoty bez genetických skupin a s genetickými skupinami. Cílem bylo zjistit vliv začlenění genetických skupin a otestovat různých způsob jejich sestavení. Práce s datovými soubory a úprava dat byla provedena v programu SAS University Edition (Statistical Analysis System) prostřednictvím Oracle VM VirtualBox a programem SAS (Statistical Analysis System) verze 9.4 (SAS, 2013). Pro výpočty plemenných hodnot byly provedeny pomocí softwaru Blupf90 (Misztal et al., 2002).

### 4.1 Vstupní data a jejich úprava

Vstupní údaje byly poskytnuty ve formě několika textových souborů. Prvními soubory byly oddělené soubory býků a krav obsahující údaje pro tvorbu rodokmenu jako číslo zvířete, číslo otce, číslo matky, datum narození a také zaznamenané plemeno daného jedince. V souboru býků byly záznamy pro celkem 84490 jedinců a v souboru krav celkem 8375094 jedinců. Dále byl nahrán soubor užitkovostí se záznamy pro celkem 6061 jedinců. Tento soubor obsahoval přečíslované číslo jedince, sro a relativní plemennou hodnotu pro dlouhověkost.

Nejprve bylo potřeba vytvořit souhrnný soubor krav a býků a přiřadit každému jedinci jeho číslo dle číselníkového souboru. Pomocí čísla jedince z číselníku pak bylo možno doplnit do rodokmenového souboru údaje o plemeni a datu narození, které byly potřeba pro tvorbu genetických skupin. Úpravy dat s popisem prováděných operací jsou uvedeny v příloze 1.

### 4.2 Rozčlenění do genetických skupin

Na základě rodokmenového souboru a údajů k jednotlivým zvířatům byly sestaveny genetické skupiny. Pokud byl v rodokmenovém souboru předek neznámý – údaj o otci, matce nebo obou se rovnal nule, byl vygenerován předek, který se rovnal číslu 2000000.

K takovému předku se následně přičetly kódy dle pohlaví a plemene posledního známého jedince. Takto bylo vytvořeno celkem šest genetických skupin o velkém počtu jedinců. Ve druhém případě sestavení genetických skupin bylo číslo uměle vygenerovaného předka kromě kódu pohlaví a plemene rozšířeno ještě o rok narození posledního známého jedince.

Toto sestavování genetických skupin bylo zvoleno na základě předpokladu, že hlavní vliv na genetickou odlišnost hodnocených zvířat hraje jejich plemenná příslušnost. V hodnoceném souboru se nacházela hlavně tři plemena, další plemena tvořila jen malou část zástupců anebo bylo plemeno jedinců neznámé. Jedinci těchto málo zastoupených nebo neznámých plemen byli sloučeni.

Na základě publikace Westell (1988). Způsob selekce a výběru jednotlivých pohlaví do další plemenitby je odlišný, a proto byly sestaveny skupiny neznámých rodičů zvlášť pro krávy a zvlášť pro býky.

Dalším použitým parametrem pro sestavení genetických skupin byl rok narození. Ten byl zvolen proto, protože se předpokládá, že koeficient příbuzenské plemenitby zvířat i jejich genetická hodnota je v každém období různá. Na základě toho předpokladu pak můžeme uvažovat, že jedinci narození přibližně ve stejném časovém období jsou si více podobní než tito jedinci s jedinci z jiného časového období.

Skupiny dle roku narození zvířat byly sestaveny pro každé čtyři roky pro neznámé otce a každé dva roky pro neznámé matky, jako vyjádření generačních intervalů pro každé pohlaví. Tento postup vedl k velkému množství genetických skupin, z nichž některé nedosahovaly doporučeného minimálního počtu jedinců tvořících skupinu neznámých předků dle Interbullu (2001). Proto byly v tomto případě skupiny s nižším počtem jedinců sloučeny se skupinami v podobném časovém trendu. Takovýmto postupem bylo sestaveno, vygenerováno a zařazeno do rodokmenu celkem 104 genetických skupin.

Následně bylo potřeba sestavené genetické skupiny přečíslovat tak, aby jejich číselný kód navazoval na jedince v rodokmenu. Byl tak vytvořen rodokmen, kdy měl každý jedinec známého rodiče anebo byl jeho chybějící rodič uvažován jako součást genetické skupiny podle plemene, pohlaví a ve druhém případě ještě dle roku narození. Číselné kódy skupin neznámých rodičů byly vyšší než nejvyšší číslo zvířete. Tyto skupiny neznámých předků tak byly na úplném konci rodokmenu.

Aby bylo možné určit, který předek je neznámý, do výpočetního souboru byla přidána položka koeficientu. Tento koeficient byl vytvořen dle následujícího klíče:

- koeficient = 1 oba rodiče jsou známí
- koeficient = 2 jeden rodič je neznámý
- koeficient = 3 oba rodiče jsou neznámí



Do softwaru pro výpočet plemenných hodnot dle metody BLUP bylo potřeba nahrávat rodokmenový soubor ve formátu číslo jedince, číslo otce, číslo matky. Koeficient počtu neznámých rodičů byl dosazen do čtvrtého sloupce. Plemenné hodnoty byly odhadnuty metodou BLUP AM pomocí programu blupf90 (Misztal 2002, Misztal 2009). Příklad rodokmenového souboru je uveden v tabulce 1.

Tab. 1: Příklad rodokmenového souboru pro vstup do genetického hodnocení.

6492	5895	5787	1
6493	3912	5822	1
6494	2618	5784	1
6495	4012	5788	1
6496	3957	46199	2
6497	46199	46199	3
6498	5880	6144	1
6499	46199	46199	3
6500	46200	46200	3
6501	46199	46199	3
6502	46199	46199	3
6503	3096	5796	1

### 4.3 Modelová rovnice

Modelová rovnice pro výpočet plemenných hodnot byla ve formátu:

$$y = sro + animal + e;$$

kde:  $y$  = vypočtená odregresovaná plemenná hodnota pro dlouhověkost

$sro$  = vektor jedniček jako pevný efekt

$animal$  = náhodný efekt jedince

$e$  = zbytková chyba

### 4.4 Tvorba parametrových souborů

Metoda BLUP byla spuštěna v rámci programu blupf90 (Misztal, 2009). Program blupf90 byl spuštěn na domácím počítači i na počítači v rámci Výzkumného ústavu živočišné výroby v Uhřetěvsi.

Bylo nastaveno celkem šest parametrových souborů, které vstupovaly do programu blupf90 (Misztal, 2009) pro výpočet plemenných hodnot. Parametrový soubor programu popisuje vstupní soubor se závislou a nezávislou proměnnou. Dále soubor obsahuje údaje o pořadí jednotlivých proměnných v rámci vstupního souboru s údaji a hodnoty variancí a

kovariancí pro náhodný efekt a informaci o druhu jednotlivých efektů (spojité, ve třídách). Dále je v něm uvedena hodnota konvergence, které je třeba dosáhnout pro ukončení výpočtu.

Výpočty plemenných hodnot byly spouštěny jako single trait model bez a s použitím vytvořených genetických skupin neznámých rodičů ve dvojím provedení sestavení těchto skupin.

Počty úrovní u náhodného efektu jedince se i jednotlivých modelů měnily a jsou uvedeny v tabulce č. 2 níže.

Tab. 2: Počty úrovní efektu jedince pro jednotlivé modely

Model	počet úrovní
bez UPG	46196
UPG1	46202
UPG2	46300

Tabulka 2 popisuje úrovně náhodného efektu jedince, které byly použity pro výpočet metody BLUP. Model bez genetických skupin, v tabulce ozančený jako „bez UPG“ měl 46196 úrovní, model s malým počtem genetických skupin (UPG1) měl 46202 úrovní a model s velkým počtem genetických skupin (UPG2) 46300 úrovní efektu. Celý parametrový soubor programu blupf90 je uveden v příloze1 - Parametrový soubor programu blupf90. Popisné statistiky byly vypočteny pomocí procedur means a freq počítačového programu SAS.

## 5 Výsledky

### 5.1 Popisné statistiky údajů

V souboru s užitkovostmi bylo celkem 6061 záznamů. Příložená tabulka 3 znázorňuje výsledky popisné statistiky upraveného souboru užitkovostí.

Tab. 3: Popisné charakteristiky souboru užitkovostí

proměnná	celkový počet	průměr	odchylka	minimum	maximum
sro	6061	1.0000000	0	1.0000000	1.0000000
jedinec	6061	23928.37	8595.85	2376.00	40814.00
drp	6061	137.8423113	72.2832963	54.3311000	513.0229000
váha	6061	37.0862403	47.8624900	1.1792000	170.2394000

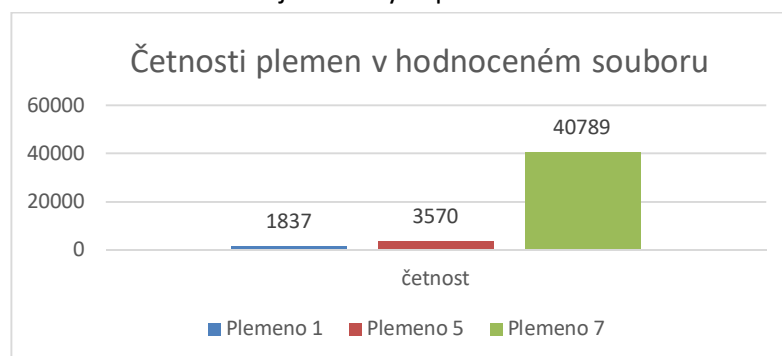
V tabulce můžeme vidět, že odregresivní plemenná hodnota byla k dispozici u 6061 jedinců. Odchylka této hodnoty byla 72,28 s minimální zaznamenanou hodnotou 54,33 a nevyšší hodnotou 513,02. Pro každou odregresovanou hodnotu byla uvedena váha tohoto údaje. Váha odregresovaných hodnot, které byly do modelové rovnice použity jako vektor pozorování, byla pro výpočet také zohledněna a to tak, že byla zadána do parametřového souboru, který vstupoval do programu pro výpočet PH.

#### Genetické skupiny

Genetické skupiny byly sestaveny pro dva modely, a to na základě tří proměnných. První proměnnou bylo pohlaví koncových jedinců v rodokmenu. Dále byla všem jedincům na základě dvou poskytnutých souborů přiřazena plemenná příslušnost a datum narození. Tyto údaje pak byly použity pro rozřazení posledních známých předků do několika skupin, viz metodika.

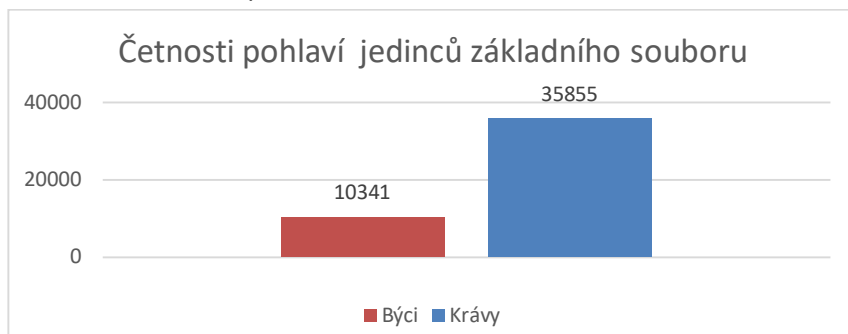
Četnosti jednotlivých plemen po úpravě vstupních dat jsou uvedeny v grafu 1, četnosti dle pohlaví v grafu 2 a četnosti dle roku narození v příloze 2.

Graf 1: Četnosti jednotlivých plemen



Z grafu je patrné, že rozdělení četností pro jednotlivé proměnné není napříč hodnoceným souborem rovnoměrné. Jedinci plemen označených jako 1 a 5 byly zastoupeni pouze z 3,98 a 7,73 % oproti jedincům plemene označeného jako 7, které bylo v souboru přítomno z 88,29 %.

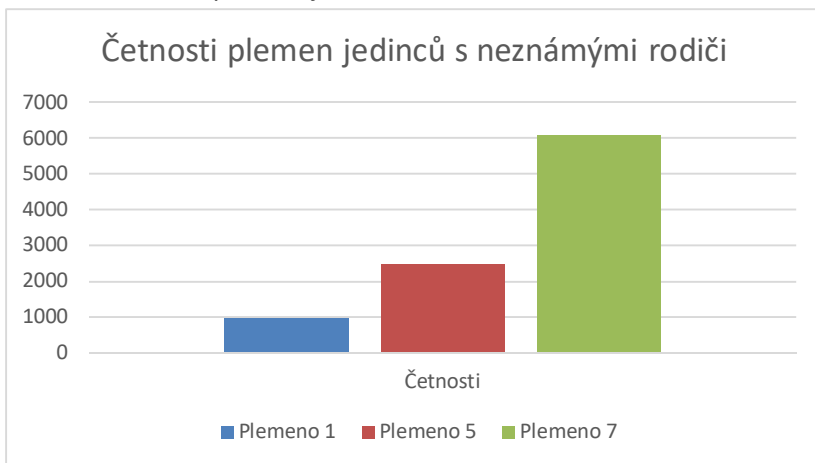
Graf 2: Četnosti dle pohlaví



Jak je znázorněno v grafu 2, většina zvířat celkového hodnoceného souboru pocházela ze souboru krav. Poměr pohlaví byl tedy 1:3 pro býky ke kravám.

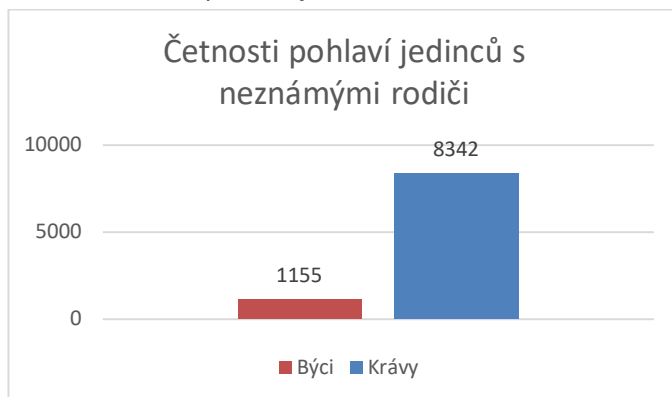
Dále byly hodnoceny statistiky jedinců, jejichž rodiče byly neznámí. Těchto jedinců bylo celkem 9497, což je 20,55 % z celkového počtu jedinců uvedených v celkovém souboru.

Graf 3: Četnosti plemen jedinců



V grafu je znázorněno rozdělení jednotlivých četností plemen pro jedince s neznámými rodiči. Stejně jako v celkovém souboru jsou nejvíce zastoupeni jedinci plemene 7 a to ze 63,9 %. Nejméně jedinců bylo příslušníky plemene 1.

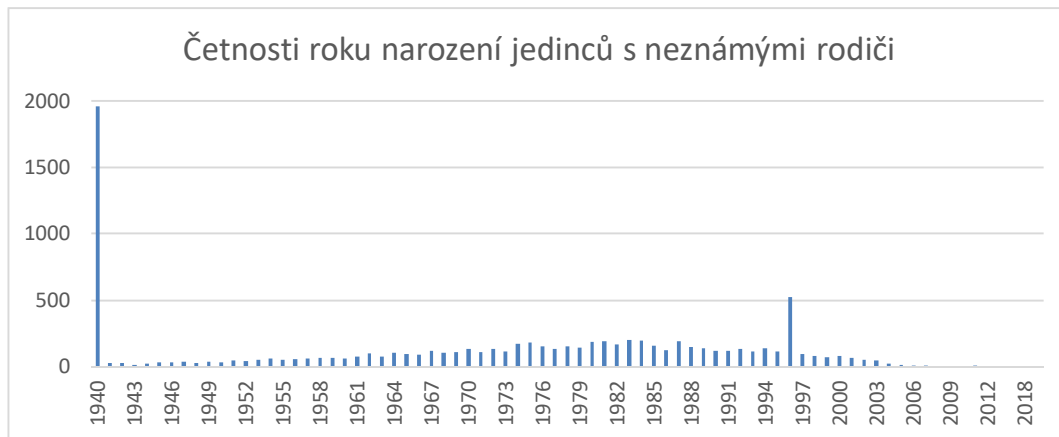
Graf 4: Četnosti pohlaví jedinců



Graf znázorňuje, že jedinci, jejichž rodiče byly neznámí, byly převážně pohlaví 2, které vyjadřovalo krávy.

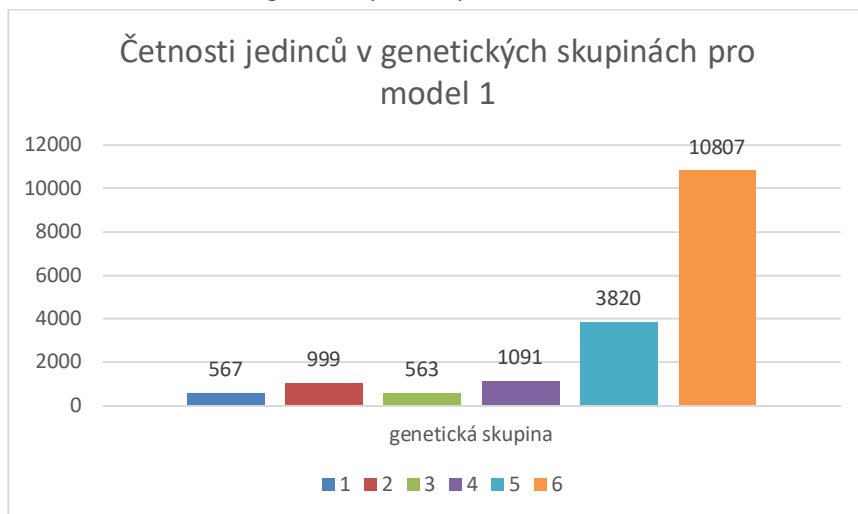
Od roku 1950 můžeme zaznamenat zvýšený počet chybějících rodičů. Jejich počet opět postupně klesá od roku 2000. Toto lze vysvětlit zaznamenáváním kontroly užitkovosti a k ní potřebných dat. Jedinci s neznámými rodiči, jejichž rok narození byl nižší, než rok 1940, byli sloučeni do jedné velké skupiny, jelikož jejich četnost byla pro každý zaznamenaný rok velmi nízká, a to v řádu jednotlivců.

Graf 5: Četnosti roku narození jedinců



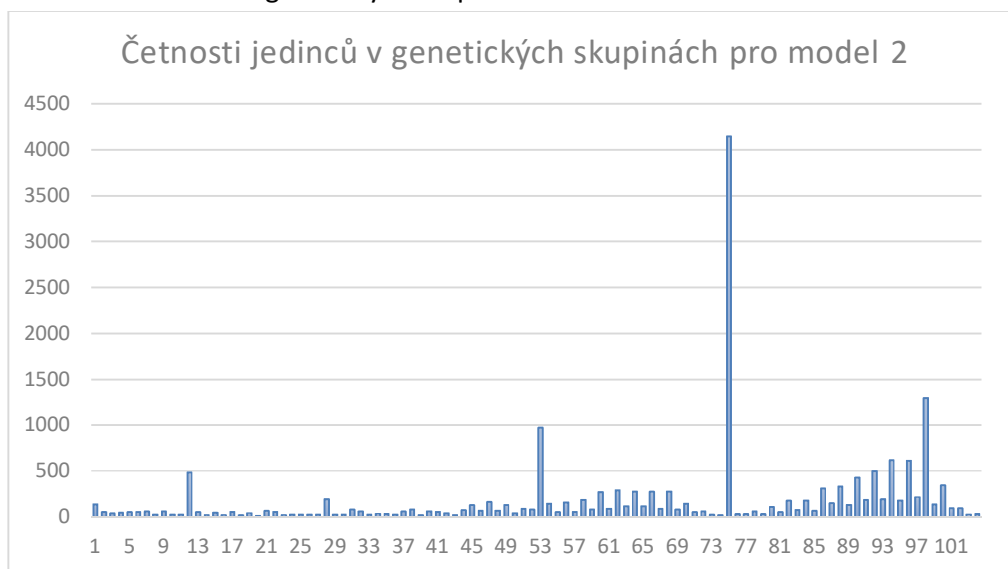
Vytvořené genetické skupiny pak pro model jedna obsahovaly jedince dle jejich pohlaví a plemenné příslušnosti. Ve druhém modelu byl zohledněn i rok narození poslejšího známého jedince. Počet skupin a počet zvířat v každé genetické skupině jsou vyjádřeny v grafu 6 pro model 1 a v grafu 7 pro model 2. Genetické skupiny byly sestavovány celkem pro 17847 neznámých rodičů.

Graf 6: Počet zvířat v genetických skupinách dle modelu 1



Pro model jedna bylo sestaveno celkem šest genetických skupin pro 17847 neznámých rodičů. Největší genetická skupina obsahovala 10807 a byla vytvořena pro krávy plemene 7. Tato skupina tvořila 60% podíl všech sestavených genetických skupin. Druhé dvě největší skupiny tvořily 21,4 % a 11,19 % z celkového podílu genetických skupin a rovněž byla sestavena pro krávy. Genetické skupiny býků s neznámými rodiči tvořily pouze 3 %, 3 % a 5 %. Nejvíce neznámých rodičů mělo u býků plemeno 5.

Graf 7: Počet zvířat v genetických skupinách dle modelu 2



Z grafu 7 je patrné, že přidáním roku narození pro sestavení genetických skupin značně narostl počet těchto skupin. Celkem bylo sestaveno 104 skupin. Největší zastoupení neznámých rodičů v počtu 4147 měla genetická skupina sestavená pro krávy plemene 7,

jejichž rok narození byl 1940 či menší. Tato skupina obsahovala 23 % ze všech posledních známých předků. Také genetická skupina sestavená pro rodiče krav plemene 7 a roku narození 1992 zaujímala větší podíl koncových zvířat v rodokmenu a to 1294, což znamenalo 7,25 % z cekového počtu.

Nejmladší genetická skupina byla sestavena pro neznámé rodiče krav plemene 7, jejichž rok narození byl 2008. Tato skupina čítala 30 jedinců. Nejméně početné genetické skupiny byly sestaveny pro neznámé rodiče býků.

## 5.2 Výsledky metody BLUP

Bylo získáno celkem 46196 údajů plemenných hodnot pro model s použitím tradiční metody BLUP, 46202 údajů plemenných hodnot pro metodu BLUP s použitím prvního modelu genetických skupin a 46300 údajů pro druhý model sestavení genetických skupin.

Tab. 4: Popisné charakteristiky odhadů plemenných hodnot pro všechna zvířata

proměnná	celkový počet	průměr	odchylka
ph	46197	-32.58439	30.00463
upg1	46203	48.63910	36.66038
upg2	46301	54.16241	170.21955

V tabulce 4 jsou porovnány odhady PH získané pro jedince celého rodokmenu včetně sestavených genetických skupin. Jako proměnná ph označen získaný odhad plemenných hodnot podle tradičních vyhodnocení bez využití genetických skupin. Jako proměnná upg1 je označen odhad PH pro první model sestavených genetických skupin a jako upg2 je označena proměnná pro odhady PH dle modelu dvě. Z tabulky je patrné, že největší odchylky odhadnutých PH dosahoval model 2, který obsahoval větší počet sestavených skupin. Model se šesti sestavenými genetickými skupinami (upg1) dosahoval odchylky 36,6 a průměru 48,6 oproti nízkému průměru tradičně užívaného modelu (ph). Odchylka tradičně užívaného modelu není výrazně odlišná od modelu zohledňující malý počet genetických skupin. Dále byly porovnávány pouze odhady PH pro zvířata, která měla záznam o užitkovosti. Charakteristiky tohoto souboru jsou uvedeny v tabulce 5.

Tab. 5: Popisné charakteristiky odhadů plemenných hodnot

proměnná	celkový počet	průměr	odchylka	Minimum	maximum
ph	6061	-48.62226	34.99850	-115.12707	94.59626
upg1	6061	32.16266	35.24571	-34.51617	178.80443
upg2	6061	29.01120	36.50697	-38.12203	281.99652

Jako proměnná ph označen získaný odhad plemenných hodnot podle tradičních vyhodnocení bez využití genetických skupin. Jako proměnná upg1 je označen odhad PH pro první model sestavených genetických skupin a jako upg2 je označena proměnná pro odhady

PH dle modelu dvě. Z hodnot uvedených v tabulce 10 lze vidět, že odchylky pro všechny hodnocené modely se vyrovnaly. Průměry odhadů pro modely s genetickými skupinami jsou si velmi blízké. Průměr odhadů dle modelu bez genetických skupin byl naopak nižší, než v tabulce 9, kdy byl zohledněn celý soubor jedinců v rodokmenu. Nejmenší hodnota odhadu byla nalezena ve výsledcích pro model bez genetických skupin. Naopak nejvyšší hodnoty dosahovala PH v souboru s modelem dvě, kde bylo sestaveno celkem 104 genetických skupin.

Dále byly dosažené odhady PH pro různé modely porovnány metodou korelace. Toto porovnání nalezneme v tabulce 6.

Tab. 6: Korelace odhadnutých plemenných hodnot pro všechna zvířata

	PH	PH - upg1	PH - upg2
PH	1.00000	0.81980	0.18134
PH - upg1	0.81980	1.00000	0.29948
PH - upg2	0.18134	0.29948	1.00000

V tabulce 6 jsou znázorněny korelace odhadnutých PH, kde ph označuje odhady dle tradičního vyhodnocení bez genetických skupin, upg1 znázorňuje odhady PH pro model se šesti genetickými skupinami a upg2 pro model s větším počtem genetických skupin. V těchto odhadech plemenných hodnot jsou zahrnuty veškeré PH všechně odhadnutých PH pro jednotlivé genetické skupiny. Nejmenší korelací odhadů dosahuje model 2 genetických skupin (upg2) s odhady PH bez genetických skupin (PH). Naopak korelace mezi tradičním modelem a modelem s menším počtem skupin je 0,82.

Dále byla PH znázorněna pouze pro zvířata, která měla zaznamenanou užitkovost. Korelace mezi těmito odhady PH je znázorněna v tabulce 7.

Tab. 7: Korelace odhadnutých plemenných hodnot - užitkovost

	PH	PH - upg1	PH - upg2
PH	1.00000	0.99917	0.98869
PH - upg1	0.99917	1.00000	0.99010
PH - upg2	0.98869	0.99010	1.00000

Tabulka 7 vyjadřuje korelace odhadů PH pouze pro jedince s užitkovostí. Hodnota PH znázorňuje odhady pro rodokmeny bez zapojení genetických skupin. Hodnota PH – upg1 vyjadřuje odhady pro model 1 se šesti genetickými skupinami a hodnota PH-upg2 pro model 2 s větším množstvím genetických skupin.



V následující tabulce jsou uvedeny odhady PH pro jednotlivé genetické skupiny modelu 1, které byly sestaveny na základě pohlaví a plemene posledního známého jedince.

Tab. 8: Odhadnuté PH pro genetické skupiny modelu 1

Skupina	PH
2110000	76.46931542
2150000	22.02306106
2170000	87.00712931
2210000	-29.21013725
2250000	62.13213513
2270000	115.00086745

Plemenné hodnoty pro genetické skupiny byly odhadnuty na základě průměrů PH jedinců, kteří je tvoří. Z tabulky můžeme vyčíst, že odhady PH jsou velmi odlišné pro různé genetické skupiny. Nejvyšší hodnoty PH dosahuje genetická skupina pro krávy plemene 7. Nejmenší hodnotu naopak skupina pro krávy plemene 1. Odhady PH pro genetické skupiny sestavené podle modelu 2 vykazovaly mnohem výraznější variabilitu. Jednoduché porovnání odhadů PH genetických skupin těchto dvou modelů je znázorněno v tabulce 9.

Tab. 9: Základní charakteristiky odhadnutých PH pro genetické skupiny

	Počet skupin	Průměr	Odchylka	Minimum	Maximum
upg1	6	55.5703952	51.6008389	29.2101373	115.0008675
upg2	104	24.6061675	745.1478224	-5812.19	1467.90

V tabulce výše můžeme pozorovat, že směrodatná odchylka PH genetických skupin modelu 1 s šesti genetickými skupinami – v tabulce označena jako upg1, byla 51,6. Oproti tomu směrodatná odchylka odhadů PH pro celkem 104 byla velmi vysoká – a to 745,14.

Dále byly pro všechny tři modely vypočítány odhady relativní kvality statistického modelu, které se nazývá Akaikovo kritérium (AIC). Model bez genetických skupin měl AIC rovno hodnotě 59638,1400126061. Pro model 1 s menším počtem genetických skupin se AIC rovnalo 59548,470 a pro model s velkým počtem skupin bylo AIC rovno 58713,461. Nejnižší hodnota AIC tedy byla pro testovaný model 2, který je dle tohoto kritéria považován za optimální.

## 6 Diskuze

Cílem této práce bylo na základě světové literatury nastudovat a ověřit v praxi způsob začlenění skupin neznámých rodičů do předpovědi plemenných hodnot získaných z kontroly užitečnosti skotu navrhnout vhodný postup pro tvorbu a následné včlenění skupin neznámých předků do modelové rovnice pro odhad plemenné hodnoty.

Dle dostupných literárních zdrojů je předpoklad, že různé sestavení genetických skupin v odhadech PH může tyto odhady velkou měrou ovlivnit. Díky genetickým skupinám dochází k zohlednění možných příbuzenských vazeb mezi zvířaty, která jsou při jiném způsobu sestavení skupin považována za nepříbuzná.

Tyto genetické skupiny však mohou být vytvořeny různě. V této diplomové práci bylo použito sestavení genetických skupin na základě plemen, jež tvořily základní soubor. V tomto souboru dominovalo plemeno označené jako plemeno 7. Nicméně po analýze souboru, kde byly obsaženi pouze jedinci, kteří měli chybějícího jednoho nebo oba rodiče lze vidět, že poměr dalších dvou plemen vzrostl. Toto naznačuje, že ačkoliv v celkovém souboru byla většina jedinců považována za příslušníky plemene 7 a plemena 1 a 5 tvoří pouze minoritní podíl, při analýze koncových zvířat zjistíme, že tento podíl u nich není zachován. Je předpokládáno různé genetické založení zvířat odlišných plemen. Proto mohou mít příspěvky těchto zvířat velký vliv na odhady PH.

Dalším parametrem pro sestavování skupin bylo zvoleno pohlaví posledních zaznamenaných jedinců. Předpoklad rozdílných selekčních cest, a tedy rozdílného výběru jedinců do plemenitby, uvedla Westell (1988).

V prvním modelu této práce tedy byla pro tvorbu genetických skupin zohledněna kritéria plemene a pohlaví. Celkem bylo vytvořeno šest genetických skupin. Ve druhém modelu bylo k předchozím kritériím přidáno kritérium času v podobě roku narození. Vlivem šlechtění dochází ke zvyšování genetické úrovně zvířat i jejich koeficientu příbuznosti. Na základě této úvahy by pak zvířata s od sebe vzdálenými daty narození byla k sobě méně podobná než zvířata s blízkým rokem narození. V současné době stanovuje Holstein Association USA genetické skupiny podle roku narození zvířat a spojuje každé 4 roky pro neznámé otce a každé 2 roky pro neznámé matky. Dále zahrnuje pohlaví neznámých rodičů. Toto hodnocení se zohledněním geneických skupin je v USA stanovováno rutinně. Na základě tohoto modelu byl tedy k již existujícím kritériím, kterým byly zohledněny v modelu jedna, zahrnuty roky narození posledních známých jedinců. Rovněž bylo použito sestavení skupin dle generačního intervalu, a to každé čtyři roky pro neznámé otce a každé dva roky pro neznámé matky.

Tímto způsobem sestavení však došlo k vytvoření značného množství skupin. Některé skupiny seskupovaly pouze několik málo jedinců.

Interbull (2001) doporučuje sestavit genetické skupiny s ohledem na plemeno, zemi původu, způsob selekce a datum narození. Rovněž doporučuje stanovit minimální počet zvířat v genetické skupině na 20 jedinců. Toto doporučení bylo v této diplomové práci při výpočtech zohledněno a genetické skupiny, které dle prvních podmínek vykazovaly menší počet jedinců, byly sloučeny do skupin s blízkým rokem narození. Sloučení skupin bylo nutné u navrženého modelu 2 pro velký počet skupin, kdy bylo v konečné fázi sestaveno celkem 104 těchto genetických skupin. Nicméně některé genetické skupiny obsahovaly nejdříve pouze několik jedinců. Pokud by byly tyto skupiny ponechány bez sloučení, mohlo by dojít ke zkreslení odhadů, neboť příspěvky genetických skupin jsou tvořeny na základě průměru jedinců, jež jsou v nich zahrnuty.

V modelu s menším počtem skupin k těmto problémům nedošlo, neboť skupiny byly tvořeny velkým počtem jedinců viz. graf 5.

Rovněž byl vysloven předpoklad, že pro vlastnosti s nízkou dědivostí by bylo možná nutné sestavit skupiny s větším počtem jedinců. V případě nutnosti spojení skupin s cílem získání většího počtu jedinců by mělo být sloučení primárně těch skupin, které přispívají ke stanovení časového trendu. Nicméně je třeba opatrnosti, pokud existují velké rozdíly v průměrech genetických skupin s různými původy.

Tsuruta (2014) sestavoval genetické skupiny s ohledem na časové období, pohlaví a zemi původu. Země původu je pro sestavování skupin zohledňována z důvodu rozdílného způsobu šlechtění, který každý země využívá. Rovněž se předpokládá, že většina chybějících rodičů bude mít starší datum narození. Ve starších dobách ovšem ještě nebyl trh s plemenným materiálem tak otevřený a přístupný, jako je tomu nyní. Rovněž import a export zvířat neprobíhal v takové míře. Proto docházelo nutně k tomu, že zvířata ze stejné země si byla bližší než zvířata mezi zeměmi.

Sestavení skupin podle země však v této diplomové práci nebylo zohledněno z důvodu, že tento postup by vedl k vytvoření skupin, které by obsahovaly pouze malý počet jedinců, což by mohlo zkreslovat výsledné odhady PH.

V této diplomové práci byly tedy použity dva způsoby sestavení genetických skupin. První model zohledňoval pohlaví a plemeno posledního známého jedince, druhý model genetických skupin byl sestaven na základě pohlaví, plemene a roku narození posledního známého jedince. Takto nově vytvořené genetické skupiny byly zahrnuty do rodokmenů a ty byly použity pro nové odhady PH.

Vypočtené odhady PH byly mezi sebou porovnány pro model bez genetických skupin a modely s genetickými skupinami. K porovnání byla použita metoda korelace. Korelační koeficienty mezi jednotlivými odhady jsou uvedeny v tabulce 6. Korelační koeficient pro odhady PH bez genetických skupin a pro první model sestavení genetických skupin, na

základě odhadů pro celý soubor zvířat a genetických skupin zahrnutých do rodokmenu, udává hodnotu 0,82. Tuto hodnotu můžeme považovat uspokojivou. Oproti tomu hodnota korelačního koeficientu mezi odhady bez genetických skupin a odhadů na základě modelu dva je velmi nízká. Takto nízká hodnota byla způsobena pravděpodobně příspěvkem sestavených genetických skupin. U prvního korelačního koeficientu se toto neprojevilo. Jak je patrné v tabulce 6, kde byly porovnávány pouze odhady PH jednotlivých sestavených genetických skupin, odchylka pro druhý model je velmi vysoká.

Když byl porovnáván pouze soubor s 6061 jedinci, u kterých byla známa jejich užitkovost, odhady PH pro všechny tři modely si byly výrazně bližší. Toto můžeme vidět například v tabulce 7, kde jsou porovnány vzájemné korelace odhadů PH. Tyto korelace jsou velmi vysoké a blíží se jedné.

Rovněž bylo provedeno statistické hodnocení modelů na základě Akaikeho kritéria. S ohledem na získané hodnoty AIC byl statisticky nejvhodnějším modelem vybrán model 2, který byl sestaven na základě zohlednění plemene, pohlaví a roku narození posledních známých jedinců.

Z tohoto vyplývá, že model bez skupin neznámých předků je špatný, nejvhodnější je model upg2 a nízká korelace naznačuje neshodu

Další možností pro zlepšení sestavení skupin by mohlo být rovnoměrnější rodičů v genetických skupinách. V případě, že by bylo ve skupivě velké množství jedinců, bylo by možné tyto skupiny ještě dále dělit. Tím by došlo k rovnoměrnějšímu rozložení příspěvků a možná i k přesnějším odhadům.

Ke zlepšení přesnosti odhadů by také mohlo značně přispět využití začlenění genetických skupin do odhadů GEPH. Z toho důvodu bude plánovaný výzkum pokračovat s cílem o začlenění tohoto postupu do odhadů GEPH.

## 7 Závěr

- Cíl práce byl splněn. Byly provedeny dva způsoby sestavení genetických skupin. Rodokmeny sestavené na základě těchto skupin byly použity k odhadnutí plemenných hodnot.
- Metodou korelace byly porovnány odhady plemenných hodnot s genetickými skupinami a bez genetických skupin a bylo zjištěno, že na koeficienty korelace mezi odhady PH má velký vliv, zdali se je porovnává celý soubor pro všechny jedince zahrnuté do výpočtů, nebo pouze soubor s jedinci, kteří mají zaznamenanou užitkovost.
- Pomocí Akaikeova kritéria pro hodnocení spolehlivosti statistických modelů bylo zjištěno, že nevyšší spolehlivosti dosahuje druhý testovaný model sestavení genetických skupin, kdy bylo sestaveno větší množství skupin na základě třech kritérií. Tyto kritéria zahrnovala plemeno, pohlaví a rok narození posledního známého jedince.

## 8 Literatura

- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93:743–752.
- Aguilar I, Misztal I, Tsuruta S, Wiggans GR, Lawlor T J. 2011. Multiple trait genomic evaluation of conception rate in Holsteins. *Journal of Dairy Science* 94:2621–2624.
- Chen CY, Misztal I, Aguilar I, Tsuruta S, Meuwissen THE, Aggrey SE, Wing T, Muir WM. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. *Journal of Animal Science* 89:23–28.
- Christensen OF, Lund MS. 2010. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42:2.
- Christensen O, Madsen P, Nielsen B, Ostersen T, Su G. 2012. Single-step methods for genomic evaluation in pigs. *Animal* 6:1565–1571.
- Daetwyler HD, Villanueva B, Bijma P, Woolliams JA. 2007. Inbreeding in genomewideselection. *Journal of Animal Breeding and Genetics* 124:369-376.
- Dillard EU, Rodriguez O, Robison OW. 1980. Estimation of additive and nonadditive direct and maternal genetic effects from crossbreeding beef cattle. *Journal of Animal Science* 50:653.
- Emik LO, Terrill CE. 1949. Systematic procedures for calculating inbreeding coefficients. *Journal of Heredity* 40: 51–55.
- Falconer DS. 1960. *Introduction to Quantitative Genetics*. Oliver and Boyd Ltd., Edinburgh 1-140.
- Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Philosophical Transactions of the Royal Society* 52:399-433.
- Forni S, Aguilar I, Misztal I. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution* 43:1.
- Interbull. 2001. Interbull guidelines for national & international genetic evaluation systems in dairy cattle with focus on production traits. *Interbull Bull* 28.
- Harris BL, Johnson DL. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. *Journal of Dairy Science* 93 :1243–1252
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92:433–443.

- Henderson CR. 1975. Uses of relationships among sires to increase the accuracy of sire evaluation. *Journal of Dairy Science* 58:1731.
- Henderson CR. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69.
- HENDERSON CR. 1984. Applications of linear models in animal breeding. University of Guelph Press, Guelph, Canada.
- Kennedy B. 1991 CR Henderson: The unfinished legacy. *Journal of Dairy Science* 74: 4067–4081.
- Makgahlela ML, Strandén I, Nielsen US, Sillanpää MJ, Mäntysaari EA. 2013. The estimation of genomic relationships using breedwise allele frequencies among animals in multibreed populations. *Journal of Dairy Science* 96:5364–5375.
- Makgahlela ML, Strandén I, Nielsen US, Sillanpää MJ, Mäntysaari EA. 2014. Using the unified relationship matrix adjusted by breed-wise allele frequencies in genomic evaluation of a multibreed population. *Journal of Dairy Science* 97 :1117–1127.
- Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. 2002. Blupf90 and related programs (BGF90). Proceedings of the 7th world congress on genetics applied to livestock production 20(7).
- Misztal I, Legarra A, Aguilar I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science* 92:4648–4655.
- Misztal I, Aguilar I, Legarra A, Tsuruta S, Johnson DL, Lawlor TJ. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation. *Journal of Dairy Science* 93(2):743-52.
- Misztal I, Tsuruta S, Lourenco DAL. 2013. Bias in singlestep genomic evaluations attributable to unknown parent group estimates. *Journal of Dairy Science* 96:75.
- Misztal I, Vitezica ZG , Legarra A, Aguilar I, Swan AA. 2013. Unknown-parent groups in single-step genomic evaluation. *Journal of Animal Breeding and Genetics* 130:252–258.
- Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. 2014. Manual for BLUPF90 family of programs. University of Georgia, Athens, USA. p. 125.
- Mrode RA. 2005. Linear models for the prediction for animal breeding values. CABI Publishing, Wallingford, UK.
- Legarra A, Aguilar I, Misztal I. 2009. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92:4656–4663.
- Legarra A, Christensen OF, Vitezica ZG, Aguilar I, Misztal I. 2015. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. *Genetics* 200:455–68.

- Phocas F, Laloe D. 2004. Should genetic groups be fitted in BLUP evaluation? Practical answer for the French AI beef sire evaluation. *Genetics Selection Evolution* 36:325–345.
- Powell JE, Visscher PM, Goddard ME. 2010. Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics* 11: 800–805.
- Příbyl J, Haman J, Kott T, Příbylová J, Šimečková M, Vostrý L. 2012. Single-step prediction of genomic breeding value in a small dairy cattle population with strong import of foreign genes. *Czech Journal of Animal Science* 57(4): 151–159.
- Quaas RL, Pollak EJ. 1981. Modified equations for siremodels with groups. *Journal of Dairy Science* 64:1868–1872.
- Quass RL. 1988. Additive Genetic Model with Groups and Relationships. *Journal of Dairy Science* 71: 1338–1345.
- Rosenberg NA, Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* 3:380–390.
- SAS institute Inc. 2013. Base SAS® 9.4 procedures guide: Statistical procedures, second edition. Cary, NC: SAS institute Inc.
- Sebastiani P. 2014. An Efficient Technique for Bayesian Modelling of Family Data Using the BUGS software. *Frontiers in Genetics* 5:390.
- Tekman M, Medlar A, Mozere M, Kleta R, Stanescu H. 2017. HaploForge: a comprehensive pedigree drawing and haplotype visualization web application. *Bioinformatics* 1–7.
- Thompson R (1986) Estimation of realized heritability in a selected population using mixed model methods. *Genetics Selection Evolution* 18:475–483
- Tsuruta S, Misztal I, Lourenco DAL, Lawlor TJ. 2014. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. *Journal of Dairy Science* 97:1–8.
- VanRaden PM. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *Journal of Dairy Science* 75:3136–3144.
- VanRaden PM. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91, 4414–4423.
- VanRaden PM. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *Journal of Dairy Science* 94 :5673–5682.
- Westell RA, Quass RL, Dale Van Vleck L. 1988. Genetic Groups in an Animal Model. *Journal of Dairy Science* 71:1310–1318.
- Wright S. 1922. Coefficients of inbreeding and relationship. *The American Naturalist* 56: 330–338.



Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42:565-569.

## 9 Seznam použitých zkratk a symbolů

PH – plemenná hodnota

GEPH – genomická plemenná hodnota

BLUP – nejlepší lineární nevychýlená předpověď

MV – náhodný mendelistický výběr

ssGBLUP – single step nejlepší nevychýlená genomická předpověď

IBD – identity by descendent

IBS – identity by state

LD – linkage disequilibrium – vazebná nerovnováha

GBLUP – genomický BLUP

## 10 Samostatné přílohy

Příloha 1: Kód programu SAS upravující vstupní data.

```
/*.....prirod2.sas.....*/
/*  priprava pro GEPH dlouhovekosti, kontrolni tisky
           iteraci oprava roku                114,149,155
           preseknuti kruhoveho rod          150
...*/
*libname user '/folders/myfolders/temp/';

filename byci '/folders/myfolders/Luboš data/kgm009.txt';/*rod*/
filename kravy '/folders/myfolders/Luboš data/kgm008.txt';/*rod*/

/* zapis */
filename ciseln '/folders/myfolders/Luboš data/ciseln';/*číselník jedinců*/
filename rod '/folders/myfolders/Luboš data/puvd';/*celkový rodokomen*/
filename ped '/folders/myfolders/Luboš data/upgped';/*upg rodokomen*/
filename pedgen '/folders/myfolders/Luboš data/pedgen';/*upg rodokomen*/
filename pedgen2 '/folders/myfolders/Luboš data/pedgen2';/*upg rodokomen*/
filename pedgen3 '/folders/myfolders/Luboš data/pedgen3';/*upg rodokomen*/
filename pedgen4 '/folders/myfolders/Luboš data/pedgen4';/*upg rodokomen*/

dm output 'clear';
dm log 'clear';
/*****soubor užít*****/
/* Generated Code (IMPORT) */
/* Source File: uzit.txt */
/* Source Path: /folders/myfolders/Luboš data */
/* Code generated on: 27.03.19 17:16 */
%web_drop_table(WORK.IMPORT7);
FILENAME REFFILE '/folders/myfolders/Luboš data/uzit.txt';
PROC IMPORT DATAFILE=REFFILE
      DBMS=DLM
      OUT=WORK.IMPORT7;
      GETNAMES=NO;
RUN;
PROC CONTENTS DATA=WORK.IMPORT7; RUN;
%web_open_table(WORK.IMPORT7);
```

```

data uzit;
set work.import7;
sro= var1;
ncj = var2;
drp = var3;
vaha= var4;
keep sro ncj drp vaha;
proc sort; by ncj;
*proc means;
/* .....otci..... */
*libname xx v9 '/folders/myfolders/temp';
*options user=xx;
data males;          title "byci seznam 17";
    attrib cisj o matka format=$14. length=$14;
    attrib pl format=$1. length=$1;
infile byci missover;
input cisj $ 1-6 o $ 8-13 matka $15-28 rok nar 46-49 pl $59 ;
    if rok nar = 0 then rok nar = . ;
    if o*1 = 0 then o = "" ; if matka*1 = 0 then matka = "" ;
    poh1 = 1;
        keep cisj pl poh1;
/*proc means;
proc univariate;*/
proc sort ; by cisj ; run;
/* .....matky..... */
*libname xx v9 '/folders/myfolders/temp';
*options user=xx;
data females;          title "kravy seznam 29";
    attrib cisj o matka format=$14. length=$14;
    attrib pl format=$1. length=$1;
infile kravy missover;
input cisj $1-14 o $16-21 matka $23-36 rok nar 54-57 pl $63 ;
    if rok nar = 0 then rok nar = . ;
    if o*1 = 0 then o = "" ; if matka*1 = 0 then matka = "" ;
    poh1 = 2;
        keep cisj pl poh1;
/*proc means;
proc univariate;*/
proc sort ; by cisj ;

```

```
/****** číselník******/
```

```
data ciseln;  
  infile ciseln;  
  input cisj $1-16 rokn $18-21 ncj 39-46 ;  
  proc sort ; by cisj ;  
data vse;  
merge ciseln males; by cisj;  
data vse;  
merge vse females; by cisj;  
data vse;  
set vse;  
if ncj = "." then delete;  
keep ncj pl pohl rokn;  
proc sort; by ncj;  
data pl;  
set upg;  
keep pl;  
proc freq;
```

```
/******spojení s rodokmenem******/
```

```
data rod;  
infile rod ;  
  input ncj 1-7 otec 9-15 matka 17-23 ;  
data celrod;  
merge rod vse; by ncj;  
proc sort; by ncj; /*title "četnosti data narození";  
proc freq; tables rokn;
```

```
/******spojení s užítkovostí******/
```

```
data uz;  
merge uzit celrod; by ncj;  
run;  
proc sort; by ncj;
```

```
/******tvorba UPG******/
```

```
data upg;          /* koeficient = 1  oba rodiče jsou známí • koeficient = 2  jeden rodič je  
neznámý • koeficient = 3  oba rodiče jsou neznámí*/  
set uz;  
if rokn = ".CZO" then rokn = 0;  
do upg = 1;end;  
if otec = 0 & matka = 0 then do; upg = 3;end;
```

```

else if otec = 0 then do;upg = 2 ; end;
else if matka = 0 then do;upg = 2 ;end;
if otec = 0 then otec =otec + 2000000;
if matka = 0 then matka =matka+ 2000000;
  if pl = "." then pl = 0;
  if pl = 0 then pl = 5;
    if pohl = "." then pohl = 2;
    if rokcn <1842 then rokcn = 1842;
data upg;
set upg;
if pl ne 7 and pl ne 1 then pl = 5;
if pohl = 0 then pohl = 2;
data pl2;
set upg;
keep pl;
proc freq;
data pohl;
set upg;
keep pohl;
proc freq;
data datum;
set upg;
keep rokcn;
proc freq;
data upg;
set upg;
/*pohlaví*/
if otec = 2000000 & pohl = 1 then do; otec = otec+100000; end;
  else if otec = 2000000 & pohl = 2 then do; otec = otec + 200000;end;
if matka = 2000000 & pohl = 1 then do; matka = matka + 100000; end;
  else if matka = 2000000 & pohl = 2 then do; matka = matka+200000;end;
data upg;
set upg;
/*plemeno*/
if otec = >2000000 & pl = 1 then do; otec = otec+10000; end;
  else if otec = >2000000 & pl = 7 then do; otec = otec+70000;end;
    else if otec = >2000000 & pl = 5 then do; otec = otec+50000;end;
if matka = >2000000 & pl = 1 then do; matka = matka+10000; end;
  else if matka = >2000000 & pl = 7 then do; matka = matka+70000;end;
    else if matka = >2000000 & pl = 5 then do; matka = matka+50000;end;

```

```
/*rok narození*/
```

```
if otec = >2000000 & rok = < 1950 then do; otec = otec+1940; end;
```

```
  else if otec = >2000000 & rok >1950 & rok = <1954 then do; otec = otec+1950;end;
```

```
  else if otec = >2000000 & rok >1954 & rok = <1958 then do; otec = otec+1954;end;
```

```
  else if otec = >2000000 & rok >1958 & rok = <1962 then do; otec = otec+1958;end;
```

```
  else if otec = >2000000 & rok >1962 & rok = <1966 then do; otec = otec+1962;end;
```

```
  else if otec = >2000000 & rok >1966 & rok = <1970 then do; otec = otec+1966;end;
```

```
  else if otec = >2000000 & rok >1970 & rok = <1974 then do; otec = otec+1970;end;
```

```
  else if otec = >2000000 & rok >1974 & rok = <1978 then do; otec = otec+1974;end;
```

```
  else if otec = >2000000 & rok >1978 & rok = <1982 then do; otec = otec+1978;end;
```

```
  else if otec = >2000000 & rok >1982 & rok = <1986 then do; otec = otec+1982;end;
```

```
  else if otec = >2000000 & rok >1986 & rok = <1990 then do; otec = otec+1986;end;
```

```
  else if otec = >2000000 & rok >1990 & rok = <1994 then do; otec = otec+1990;end;
```

```
  else if otec = >2000000 & rok >1994 & rok = <1998 then do; otec = otec+1994;end;
```

```
  else if otec = >2000000 & rok >1998 & rok = <2002 then do; otec = otec+1998;end;
```

```
  else if otec = >2000000 & rok >2002 & rok = <2006 then do; otec = otec+2002;end;
```

```
  else if otec = >2000000 & rok >2006 & rok = <2010 then do; otec = otec+2006;end;
```

```
  else if otec = >2000000 & rok >2010 & rok = <2014 then do; otec = otec+2010;end;
```

```
  else if otec = >2000000 & rok = >2014 then do; otec = otec+2014;end;
```

```
if matka = >2000000 & rok = < 1950 then do; matka = matka+1940; end;
```

```
  else if matka = >2000000 & rok >1950 & rok = <1952 then do; matka =  
matka+1950;end;
```

```
  else if matka = >2000000 & rok >1952 & rok = <1954 then do; matka =  
matka+1952;end;
```

```
  else if matka = >2000000 & rok >1954 & rok = <1956 then do; matka =  
matka+1954;end;
```

```
  else if matka = >2000000 & rok >1956 & rok = <1958 then do; matka =  
matka+1956;end;
```

```
  else if matka = >2000000 & rok >1958 & rok = <1960 then do; matka =  
matka+1958;end;
```

```
  else if matka = >2000000 & rok >1960 & rok = <1962 then do; matka =  
matka+1960;end;
```

```
  else if matka = >2000000 & rok >1962 & rok = <1964 then do; matka =  
matka+1962;end;
```

```
  else if matka = >2000000 & rok >1964 & rok = <1966 then do; matka =  
matka+1964;end;
```

```
  else if matka = >2000000 & rok >1966 & rok = <1968 then do; matka =  
matka+1966;end;
```

```
  else if matka = >2000000 & rok >1968 & rok = <1970 then do; matka =  
matka+1968;end;
```

```
else if matka = >2000000 & rokn >1970 & rok = <1972 then do; matka =
matka+1970;end;
else if matka = >2000000 & rokn >1972 & rok = <1974 then do; matka =
matka+1972;end;
else if matka = >2000000 & rokn >1974 & rok = <1976 then do; matka =
matka+1974;end;
else if matka = >2000000 & rokn >1976 & rok = <1978 then do; matka =
matka+1976;end;
else if matka = >2000000 & rokn >1978 & rok = <1980 then do; matka =
matka+1978;end;
else if matka = >2000000 & rokn >1980 & rok = <1982 then do; matka =
matka+1980;end;
else if matka = >2000000 & rokn >1982 & rok = <1984 then do; matka =
matka+1982;end;
else if matka = >2000000 & rokn >1984 & rok = <1986 then do; matka =
matka+1984;end;
else if matka = >2000000 & rokn >1986 & rok = <1988 then do; matka =
matka+1986;end;
else if matka = >2000000 & rokn >1988 & rok = <1990 then do; matka =
matka+1988;end;
else if matka = >2000000 & rokn >1990 & rok = <1992 then do; matka =
matka+1990;end;
else if matka = >2000000 & rokn >1992 & rok = <1994 then do; matka =
matka+1992;end;
else if matka = >2000000 & rokn >1994 & rok = <1996 then do; matka =
matka+1994;end;
else if matka = >2000000 & rokn >1996 & rok = <1998 then do; matka =
matka+1996;end;
else if matka = >2000000 & rokn >1998 & rok = <2000 then do; matka =
matka+1998;end;
else if matka = >2000000 & rokn >2000 & rok = <2002 then do; matka =
matka+2000;end;
else if matka = >2000000 & rokn >2002 & rok = <2004 then do; matka =
matka+2002;end;
else if matka = >2000000 & rokn >2004 & rok = <2006 then do; matka =
matka+2004;end;
else if matka = >2000000 & rokn >2006 & rok = <2008 then do; matka =
matka+2006;end;
else if matka = >2000000 & rokn >2008 & rok = <2010 then do; matka =
matka+2008;end;
```



```

else if matka = >2000000 & rok >2010 & rok = <2012 then do; matka =
matka+2010;end;
else if matka = >2000000 & rok >2012 & rok = <2014 then do; matka =
matka+2012;end;
else if matka = >2000000 & rok >2014 & rok = <2016 then do; matka =
matka+2014;end;
else if matka = >2000000 & rok >2016 then do; matka = matka+2016;end;
run;
/*úprava skupin podle počtu jedinců*/
data upg;
set upg;
if matka = 2272010 then matka = 2272008;
if matka = 2252004 then matka = 2251998;
if matka = 2252002 then matka = 2251998;
if matka = 2252000 then matka = 2251998;
if matka = 2251994 then matka = 2251998;
if matka = 2251996 then matka = 2251998;
if matka = 2271950 then matka = 2271952;
if matka = 2251992 then matka = 2251990;
if matka = 2251988 then matka = 2251990;
if matka = 2251986 then matka = 2251990;
if matka = 2212002 then matka = 2211994;
if matka = 2211998 then matka = 2211994;
if matka = 2211996 then matka = 2211994;
if matka = 2211968 then matka = 2211964;
if matka = 2211966 then matka = 2211964;
if matka = 2211962 then matka = 2211964;
if matka = 2211960 then matka = 2211958;
if matka = 2211958 then matka = 2211958;
if matka = 2211956 then matka = 2211958;
if matka = 2252010 then matka = 2252008;
if matka = 2252006 then matka = 2252008;
if matka = 2211954 then matka = 2211958;
if matka = 2211950 then matka = 2211940;
if matka = 2211952 then matka = 2211940;
if matka = 2111950 then matka = 2111952;
if matka = 2111954 then matka = 2111952;
if matka = 2111956 then matka = 2111958;
if matka = 2111960 then matka = 2111958;
if matka = 2172000 then matka = 2172002;
if matka = 2171998 then matka = 2172002;

```

if matka = 2171996 then matka = 2172002;  
if matka = 2171994 then matka = 2172002;  
if matka = 2171990 then matka = 2172002;  
if matka = 2171988 then matka = 2172002;  
if matka = 2171986 then matka = 2172002;  
if matka = 2171984 then matka = 2172002;  
if matka = 2171982 then matka = 2172002;  
if matka = 2171978 then matka = 2171980;  
if matka = 2171974 then matka = 2171978;  
if matka = 2171964 then matka = 2171966;  
if matka = 2171968 then matka = 2171966;  
if matka = 2171950 then matka = 2171966;  
if matka = 2171952 then matka = 2171966;  
if matka = 2171954 then matka = 2171966;  
if matka = 2171956 then matka = 2171966;  
if matka = 2171958 then matka = 2171966;  
if matka = 2171960 then matka = 2171966;  
if matka = 2171962 then matka = 2171966;  
if matka = 2151978 then matka = 2151986;  
if matka = 2151980 then matka = 2151986;  
if matka = 2151982 then matka = 2151986;  
if matka = 2151984 then matka = 2151986;  
if matka = 2151976 then matka = 2151976;  
if matka = 2151974 then matka = 2151976;  
if matka = 2272006 then matka = 2272008;  
if matka = 2272004 then matka = 2272002;  
if matka = 2111982 then matka = 2111996;  
if matka = 2111984 then matka = 2111996;  
if matka = 2111986 then matka = 2111996;  
if matka = 2111988 then matka = 2111996;  
if matka = 2111992 then matka = 2111996;  
if matka = 2111994 then matka = 2111996;  
if matka = 2111972 then matka = 2111970;  
if matka = 2111964 then matka = 2111962;  
if matka = 2111968 then matka = 2111966;  
if matka = 2171992 then matka = 2171980;  
if matka = 2111990 then matka = 2111996;  
if matka = 2251986 then matka = 2251984;  
if matka = 2211992 then matka = 2211990;  
if matka = 2211976 then matka = 2211974;  
if matka = 2211966 then matka = 2211964;

if matka = 2211962 then matka = 2211964;  
if matka = 2171976 then matka = 2171974;  
if matka = 2151968 then matka = 2151966;  
if matka = 2111980 then matka = 2111978;  
if matka = 2111954 then matka = 2111952;  
if matka = 2111950 then matka = 2111952;

if otec = 2272014 then otec = 2272006;  
if otec = 2272010 then otec = 2272006;  
if otec = 2252002 then otec = 2251994;  
if otec = 2251998 then otec = 2251994;  
if otec = 2212002 then otec = 2211994;  
if otec = 2211998 then otec = 2211994;  
if otec = 2211950 then otec = 2211954;  
if otec = 2172002 then otec = 2211994;  
if otec = 2211998 then otec = 2211994;  
if otec = 2171998 then otec = 2172002;  
if otec = 2171994 then otec = 2172002;  
if otec = 2171986 then otec = 2171982;  
if otec = 2171990 then otec = 2172002;  
if otec = 2171982 then otec = 2172002;  
if otec = 2171966 then otec = 2171962;  
if otec = 2171954 then otec = 2171958;  
if otec = 2171950 then otec = 2171958;  
if otec = 2151986 then otec = 2151978;  
if otec = 2151982 then otec = 2151978;  
if otec = 2111990 then otec = 2111994;  
if otec = 2111986 then otec = 2111994;  
if otec = 2111982 then otec = 2111994;  
if otec = 2251986 then otec = 2251984;  
if otec = 2211992 then otec = 2211990;  
if otec = 2211976 then otec = 2211974;  
if otec = 2211966 then otec = 2211964;  
if otec = 2211962 then otec = 2211964;  
if otec = 2171976 then otec = 2171974;  
if otec = 2151968 then otec = 2151966;  
if otec = 2111980 then otec = 2111978;  
if otec = 2111954 then otec = 2111952;  
if otec = 2111950 then otec = 2111952;  
\*/

```

data f;
set upg;
i = matka;
keep i;
data p;
set upg;
i = otec;
keep i;
data freq;
set f p;
if i < 2000000 then delete;
proc freq;
/*následuje zařazení gen. skupin do rodokmenu*/
data otec;
set upg;
keep otec;
proc sort; by otec;
data otec;
set otec; by otec; if first.otec;
data otec;
set otec;
if otec < 2000000 then delete;
ncj = otec;
keep ncj;
data matka;
set upg;
keep matka;
proc sort; by matka;
data matka;
set matka; by matka; if first.matka;
data matka;
set matka;
if matka < 2000000 then delete;
ncj = matka;
keep ncj;
/*vložení upg do rodokmenu*/
data ped;
set otec matka;
proc sort;by ncj;
data ped;
set ped; by ncj; if first.ncj;

```

```

data ped;
set upg ped;
proc sort;by ncj;
DATA CIS;
SET ped; BY ncj; if first.ncj;
KEEP ncj; run;
DATA CIS;
SET CIS;
KEEP ncj NC;
NC = _n_;
data pedn;
set ped;
if matka= "." then delete;
proc sort; by matka;
data cmat;
set cis;
matka = ncj;
keep matka nc;
data pedn;
set pedn;
if ncj = "." then delete;
proc sort; by matka;
data uzitm;
merge pedn cmat; by matka;
run;
data uzitm;
set uzitm;
matka = nc;
drop nc;
proc sort; by otec;
data cot;
set cis;
otec = ncj;
keep otec nc;
data uzito2;
merge uzitm cot; by otec;
otec = nc;
drop nc;
proc sort; by ncj;

data uzito2;

```

```

set uzito2; by ncj; if first.ncj;
data uzito2;
set uzito2;
if ncj = "." then delete;
run;
/**soubor užítkovovství***/
data uzit;
set uzito2;
if drp = "." then delete;
/**rodokmen***/
data puvod;
set uzito2;
if otec = "." then delete;
keep ncj otec matka upg;
run;

```

#### Příloha 2: Četnosti roku narození

rok narození	četnost
1842	1964
1843	1
1845	1
1846	1
1847	2
1849	2
1850	1
1851	1
1852	2
1853	2
1854	4
1855	4
1856	6
1857	2
1858	3
1859	6
1860	10
1861	12
1862	7
1863	11
1864	16

1865	16
1866	13
1867	6
1868	9
1869	14
1870	19
1871	14
1872	5
1873	11
1874	21
1875	26
1876	12
1877	5
1878	11
1879	20
1880	15
1881	5
1882	5
1883	11
1884	15
1885	6
1886	4
1887	10
1888	15
1889	11
1890	13
1891	16
1892	17
1893	12
1894	12
1895	21
1896	35
1897	36
1898	21
1899	11
1900	25
1901	40
1902	37
1903	23
1904	22

1905	35
1906	42
1907	33
1908	22
1909	27
1910	40
1911	43
1912	23
1913	18
1914	35
1915	41
1916	35
1917	22
1918	31
1919	38
1920	40
1921	17
1922	22
1923	29
1924	41
1925	19
1926	12
1927	20
1928	37
1929	28
1930	23
1931	22
1932	36
1933	36
1934	23
1935	28
1936	36
1937	38
1938	39
1939	35
1940	49
1941	55
1942	42
1943	43
1944	56



1945	64
1946	65
1947	53
1948	62
1949	67
1950	82
1951	78
1952	88
1953	100
1954	94
1955	107
1956	105
1957	134
1958	131
1959	135
1960	153
1961	179
1962	165
1963	201
1964	202
1965	214
1966	237
1967	232
1968	242
1969	261
1970	284
1971	286
1972	273
1973	358
1974	361
1975	366
1976	339
1977	417
1978	470
1979	475
1980	513
1981	489
1982	551
1983	584
1984	551

1985	522
1986	787
1987	714
1988	801
1989	722
1990	848
1991	832
1992	869
1993	802
1994	816
1995	1290
1996	965
1997	966
1998	956
1999	944
2000	1007
2001	1056
2002	989
2003	990
2004	1056
2005	1009
2006	1269
2007	892
2008	975
2009	841
2010	1038
2011	943
2012	1011
2013	1009
2014	979
2015	951
2016	1035
2017	1325
2018	1210

Příloha 3: Parametrový soubor programu blupf90 pro model bez genetických skupin

```
#      paramST
# single trait BLUP animal model
# rph = sro + animal + e
DATAFILE
uzit.txt
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
2
OBSERVATION(S)
3
WEIGHT(S)
4
EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT [EFFECT NESTED]
1 1 cross # sro, cross classified fixed effect, rank of variable in data file,
2 46196 cross # animal genetic effect, cross classified random effect,
RANDOM_RESIDUAL VALUES
0.75
RANDOM_GROUP
2
RANDOM_TYPE
add_animal
FILE
pedgen6.txt(CO)VARIANCES
0.25
OPTION saveAscii
OPTION SNP_file genot2
OPTION AlphaBeta 0.8 0.2
OPTION readGIinvert ggmat.txt
#OPTION readA22 A2mat
OPTION conv_crit 1e-17
OPTION maxrounds 2000
```

Příloha 4: Parametrový soubor programu blupf90 pro model s genetickými skupinami

```
#      paramST
# single trait BLUP animal model
# rph = sro + animal + e
DATAFILE
uzit
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
2
OBSERVATION(S)
3
WEIGHT(S)
4
EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT [EFFECT NESTED]
1 1 cross # sro, cross classified fixed effect, rank of variable in data file,
2 46207 cross # animal genetic effect, cross classified random effect,
RANDOM_RESIDUAL VALUES
0.75
RANDOM_GROUP
2
RANDOM_TYPE
add_animal
FILE
puvodupg11.txt
(CO)VARIANCES
0.25
OPTION saveAscii
OPTION SNP_file genot2
OPTION AlphaBeta 0.8 0.2
OPTION readGIinvert ggmata_convert.txt
#OPTION readA22 A2mat
OPTION conv_crit 1e-17
OPTION maxrounds 2000
```

Příloha 5: Parametrový soubor programu blupf90 pro model s genetickými skupinami

```
#      paramST
# single trait BLUP animal model
# rph = sro + animal + e
```

DATAFILE  
uzit.txt  
NUMBER\_OF\_TRAITS  
1  
NUMBER\_OF\_EFFECTS  
2  
OBSERVATION(S)  
3  
WEIGHT(S)  
4  
EFFECTS: POSITIONS\_IN\_DATAFILE NUMBER\_OF\_LEVELS TYPE\_OF\_EFFECT [EFFECT NESTED]  
1 1 cross # sro, cross classified fixed effect, rank of variable in data file,  
2 46300 cross # animal genetic effect, cross classified random effect,  
RANDOM\_RESIDUAL VALUES  
0.75  
RANDOM\_GROUP  
2  
RANDOM\_TYPE  
add\_an\_upg  
FILE  
puvodupg2.txt  
(CO)VARIANCES  
0.25  
OPTION conv\_crit 1e-17  
OPTION maxrounds 2000

