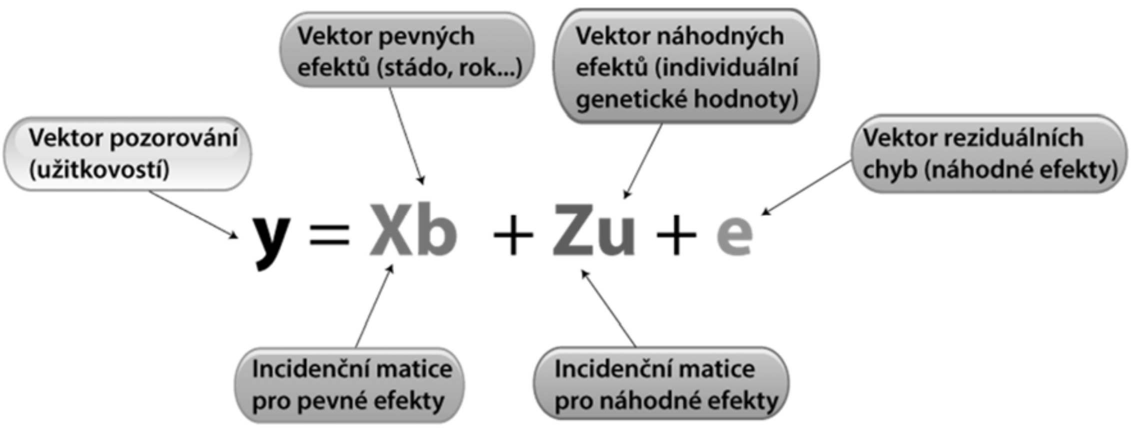
Odhad Var(A) a plemenných hodnot

- ANOVA odhaduje komponenty variance při jednom typu příbuznosti a vybalancovaných datech
- Reálné soubory jsou s nevybalancovanými daty a širokou rodokmenovou strukturou
- Téměř většina šlechtění zvířat je založena na modelech:
  - **REML** (*restricted maximum likelihood*) pro odhad variancí
  - **BLUP** (*best linear unbiased predictors*) pro předpověď plemenných hodnot

Smíšený model



y, **X** , **Z** – zaznamenané hodnoty

Odhad pevných efektů **b**
Odhad náhodných efektů **u**, **e**

Příklad **Sire modelu**

- Chceme odhadnout plemenné hodnoty tří otců (sire), každý byl pářen s náhodnými matkami (dam) a každý měl dva potomky, vyvíjející se ve dvou různých prostředích (stájích).

Pozorování	y	Otec	Stáj
y_{111}	9	1	1
y_{121}	12	1	2
y_{211}	11	2	1
y_{212}	6	2	1
y_{311}	7	3	1
y_{321}	14	3	2

Základní model

$$\mathbf{y}_{ijk} = \mathbf{b}_j + \mathbf{u}_i + \mathbf{e}_{ijk}$$

Pevný efekt prostředí (stáje)

Plemenná hodnota i-tého otce

- Vektory a matice:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_{111} \\ y_{121} \\ y_{211} \\ y_{212} \\ y_{311} \\ y_{321} \end{bmatrix} = \begin{bmatrix} 9 \\ 12 \\ 11 \\ 6 \\ 7 \\ 14 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

v programu R

```
y <- matrix(c(9,12,11,6,7,14),6,1)
```

Vytvoří sloupcový vektor **y**, s užitkovostmi dcer, s **6**ti řádky a **1** sloupcem

```
X <- matrix(c(1,0,1,1,1,0,
              0,1,0,0,0,1),6,2)
```

Vytvoří designovou matici $\mathbf{X}_{6,2}$ (pro pevný efekt stáje)

```
Z <- matrix(c(1,1,0,0,0,0,
              0,0,1,1,0,0,
              0,0,0,0,1,1),6,3)
```

Vytvoří designovou matici $\mathbf{Z}_{6,3}$ (pro náhodný efekt otce)

Průměry a variance pro $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$

- **Průměry**:
  - E(u) = E(e) = 0
  - E(y) = Xb
- **Variance**:
  - **R** je VCV matice pro rezidua (prostředí); předpoklad že $\mathbf{R} = \sigma^2_e \mathbf{I}$
  - **G** je VCV matice pro plemenné hodnoty
  - VCV matice pro **y** je: $\mathbf{V} = \mathbf{ZGZ}` + \mathbf{R}$

Odhady pevných efektů a předpovědi náhodných efektů

- Ve smíšeném modelu jsou pozorovány **y**, **X** a **Z**
- **b**, **u**, **R** a **G** jsou obecně neznámé
- Provádí se dva současné odhady:

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

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

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

(Henderson, 1963)

- Předpokládáme, že rezidua nejsou korelována a $\mathbf{R} = \sigma^2_e \mathbf{I}$ (6×6)
  - $\sigma^2_e = 6$ -> $\mathbf{R} = 6\mathbf{I}$

- VCV matice **G**: předpoklad, že otcové jsou nepříbuzní a **G** je diagonální matice (3×3) s prvky σ^2_G = variance otců, kde
  $\sigma^2_G = \sigma^2_A/4$
  - $\sigma^2_A = 8$ -> $\mathbf{G} = 8/4*\mathbf{I}$

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

$$\mathbf{V} = \frac{8}{4}\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} + 6\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 8 & 2 & 0 & 0 & 0 & 0 \\ 2 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 2 & 0 & 0 \\ 0 & 0 & 2 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 2 \\ 0 & 0 & 0 & 0 & 2 & 8 \end{bmatrix} \qquad \mathbf{V}^{-1} = \frac{1}{30}\begin{bmatrix} 4 & -1 & 0 & 0 & 0 & 0 \\ -1 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & -1 & 0 & 0 \\ 0 & 0 & -1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & -1 \\ 0 & 0 & 0 & 0 & -1 & 4 \end{bmatrix}$$

v programu R

Vytvoří jednotkovou matici 3×3 (~ 3 otci)	`I3 <- diag(c(1,1,1))`
Vytvoří genetickou variančně kovarianční matici $\mathbf{G}_{3,3}$	`G <- 8/4*I3`
Vytvoří prostřeďovou variančně kovarianční matici $\mathbf{R}_{6,6}$	`R <- 6* diag(c(1,1,1,1,1,1))`
Vytvoří fenotypovou variančně kovarianční matici $\mathbf{V}_{6,6}$	`V <- Z%*%G%*%t(Z) + R`
Vytvoří inverzi matice **V**	`invV <- solve(V)`

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \begin{pmatrix} 8{,}22222 \\ 13{,}05556 \end{pmatrix}$$

v programu R

```
b <- solve(t(X)%*%invV%*%X) %*% (t(X)%*%invV%*%y)
```

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\mathbf{b}\right) = \begin{pmatrix} -0{,}055556 \\ 0{,}111111 \\ -0{,}055556 \end{pmatrix}$$

v programu R

```
u <- G%*%t(Z)%*%invV %*% (y-(X%*%b))
```

Odvozená **soustava normálních rovnic smíšeného modelu** (Henderson):

- $\mathbf{u} \sim (0, \mathbf{G})$, $\mathbf{e}$ $(0, \mathbf{R})$, $\mathrm{cov}(\mathbf{u}, \mathbf{e}) = 0$

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

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \frac{1}{6}\begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \qquad \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} = \left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\right)' = \frac{1}{6}\begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} = \frac{5}{6}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} = \frac{1}{6}\begin{bmatrix} 33 \\ 26 \end{bmatrix} \qquad \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} = \frac{1}{6}\begin{bmatrix} 21 \\ 17 \\ 21 \end{bmatrix}$$

v programu R

```
XRX <- t(X)%*%solve(R)%*%X
        XRZ <- t(X)%*%solve(R)%*%Z
ZRX <- t(Z)%*%solve(R)%*%X
        ZRZG <- t(Z)%*%solve(R)%*%Z + solve(G)

XRy <- t(X)%*%solve(R)%*%y

ZRy <- t(Z)%*%solve(R)%*%y
```

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

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

$$\begin{bmatrix} b_1 \\ b_2 \\ u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 1 & 2 & 1 \\ 0 & 2 & 1 & 0 & 1 \\ 1 & 1 & 5 & 0 & 0 \\ 2 & 0 & 0 & 5 & 0 \\ 1 & 1 & 0 & 0 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 33 \\ 26 \\ 21 \\ 17 \\ 21 \end{bmatrix} = \begin{bmatrix} 8{,}2222 \\ 13{,}0556 \\ -0{,}0556 \\ 0{,}1111 \\ -0{,}0556 \end{bmatrix}$$

v programu R

```
LS1 <- cbind(XRX, XRZ)

LS2 <- cbind(ZRX, ZRZG)
```

Vytvoří velkou matici levé strany (**LS**)

```
LS <- rbind(LS1, LS2)
```

Vytvoří velkou matici pravé strany (**PS**)

```
PS <- rbind(XRy, ZRy)
```

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

$$[\mathbf{LS}] \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = [\mathbf{PS}]$$

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = [\mathbf{LS}]^{-1}[\mathbf{PS}]$$

Vektor řešení **bu**

```
bu <- solve(LS)%*%PS
```

Animal Model – viz přednáška č. 11

Předpokládáme, že naměřená užitkovost krávy je ovlivněna jen stádem, ve kterém je chována, věkem (tj. pořadím laktace) a genotypem (tj. jedincem se svou jedinečnou genetickou výbavou).

jedince	stádo	laktace	užitkovost
1	1	1	4500
2*	1	1	5000
3	1	2	6500
4	2	2	8000
5*	2	1	7000

V 1. stádě jsou tři dojnice, z toho dvě jsou na první laktaci a jedna na druhé laktaci. Ve 2. stádě jsou dvě krávy, jedna na první a dvě na druhé laktaci. Podle původu víme, že dojnice č. 2 a 5 mají společného otce* – jsou tedy polosestry. Jiné příbuzenské vztahy nejsou známy.

V populaci byl odhadnuta hodnota $h^2 = 0{,}25$.

modelová rovnice: $y_{ijkl} = S_i + L_j + u_k + e_{ijkl}$

maticový zápis: $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$

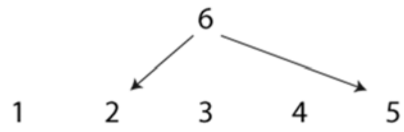
Odvozená soustava normálních rovnic smíšeného modelu:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}K \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} \qquad K = \frac{1-h^2}{h^2}$$

Řešení

- Matice A

Jedince	Otec
1	**0**
2*	**6**
3	**0**
4	**0**
5*	**6**



$a_{ii} = 1 + 0{,}5(a_{om})$

$a_{ij} = 0{,}5(a_{jo} + a_{jm})$

	1	**2**	**3**	**4**	**5**	**otec**
1	1	0	0	0	0	0
2	0	1	0	0	0,25	0,5
3	0	0	1	0	0	0
4	0	0	0	1	0	0
5	0	0,25	0	0	1	0,5
otec	0	0,5	0	0	0,5	1

Designová matice **X**

```
X1 <- matrix(c(1,1,1,0,0,0,0,0,1,1),5)
X1
    [,1] [,2]
[1,]   1    0
[2,]   1    0
[3,]   1    0
[4,]   0    1
[5,]   0    1
```

Pro pevný efekt stádo

```
X2 <- matrix(c(1,1,0,0,1,0,0,1,1,0),5)
X2
    [,1] [,2]
[1,]   1    0
[2,]   1    0
[3,]   0    1
[4,]   0    1
[5,]   1    0
```

Pro pevný efekt pořadí laktace

```
X <- cbind(X1,X2)
X
    [,1] [,2] [,3] [,4]
[1,]   1    0    1    0
[2,]   1    0    1    0
[3,]   1    0    0    1
[4,]   0    1    0    1
[5,]   0    1    1    0
```

Spojení do jedné designové matice X

Matice aditivně genetické příbuznosti **A**

```
A <- matrix(c(1,0,0,0,0,0,   0,1,0,0,0.25,0.5,  0,0,1,0,0,0,
0,0,0,1,0,0,  0,0.25,0,0,1,0.5,  0,0.5,0,0,0.5,1),6)
A
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1 0.00    0    0 0.00  0.0
[2,]    0 1.00    0    0 0.25  0.5
[3,]    0 0.00    1    0 0.00  0.0
[4,]    0 0.00    0    1 0.00  0.0
[5,]    0 0.25    0    0 1.00  0.5
[6,]    0 0.50    0    0 0.50  1.0
```

Vektor užitkovostí **y**, designová matice **Z**

```
y <- matrix(c(4500,5000,6500,8000,7000),5,1)
y
     [,1]
[1,] 4500
[2,] 5000
[3,] 6500
[4,] 8000
[5,] 7000

h2 <- 0.25
K <- (1-h2)/h2
K
[1] 3
```

```
Z <- diag(1,5)
Z
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    0    0    0    0
[2,]    0    1    0    0    0
[3,]    0    0    1    0    0
[4,]    0    0    0    1    0
[5,]    0    0    0    0    1
```

Vytvoření matice soustavy normálních rovnic

```
XX <- t(X)%*%X
XZ <- t(X)%*%Z
ZX <- t(Z)%*%X
ZZ <- t(Z)%*%Z
AK <- round(K*solve(A))          současně i zaokrouhlí
/* Abychom mohli spojit matice ZZ(5x5) a AK (6x6) musíme přidat řádek a sloupec nul do matice ZZ, aby vznikla matice o rozměrech 6x6
Stejně musíme upravit i matice XZ (+ 1 sloupec nul) a ZX (+ 1 řádek nul)
Přidáním nul se nic nemění – jen je pak možné spojit tyto submatice do matice levé strany LS */
Z0 <- matrix(c(0,0,0,0,0))
ZZ1 <- cbind(ZZ,Z0)
Z01 <- matrix(c(0,0,0,0,0,0),1,6)
ZZ2 <- rbind(ZZ1,Z01)

XZ0 <- cbind(XZ,matrix(c(0,0,0,0)))
ZX0 <- rbind(ZX,matrix(c(0,0,0,0),1,4))
ZZAK <- ZZ2+AK
```

Spojení dílčích matic do jedné matice levé strany (**LS**)

```
LS1 <- cbind(XX,XZ0)
LS2 <- cbind(ZX0,ZZAK)
LS <- rbind(LS1,LS2)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	3	0	2	1	1	1	1	0	0	0
[2,]	0	2	1	1	0	0	0	1	1	0
[3,]	2	1	3	0	1	1	0	0	1	0
[4,]	1	1	0	2	0	0	1	1	0	0
[5,]	1	0	1	0	4	0	0	0	0	0
[6,]	1	0	1	0	0	5	0	0	0	-2
[7,]	1	0	0	1	0	0	4	0	0	0
[8,]	0	1	0	1	0	0	0	4	0	0
[9,]	0	1	1	0	0	0	0	0	5	-2
[10,]	0	0	0	0	0	-2	0	0	-2	5

Konstrukce matice pravé strany (**PS**)

```
Xy <- t(X)%*%y
Zy <- t(Z)%*%y
PS <- rbind(Xy,Zy,matrix(c(0)))
```

/* rovněž u matice pravé strany musíme > PS přidat nulu, aby vznikl vektor o 10 řádcích */

```
      [,1]
 [1,] 16000
 [2,] 15000
 [3,] 16500
 [4,] 14500
 [5,]  4500
 [6,]  5000
 [7,]  6500
 [8,]  8000
 [9,]  7000
[10,]     0
```

Určení determinantu LS a zobecněná inverze

```
det <- round(det(LS))
> det
[1] 0
> bu <- solve(LS)%*%PS
Error in solve.default(LS) :
  system is computationally singular: reciprocal condition number = 1.33628e-17
```

/* Protože determinant matice LS je roven nule, je tato matice singulární a nelze ji invertovat -> jedním z řešení je použít zobecněnou inverzi…

Je nutné si nahrát balíček MASS z nabídky: Packages -> Load Packages */

```
bu <- ginv(LS)%*%PS
> bu
          [,1]
 [1,] 2302.23138  ~ odhadnutá odchylka stáda 1
 [2,] 4229.74702  ~ odhadnutá odchylka stáda 2 (odchylka 1927)
 [3,] 2547.96760  ~ odhadnutá odchylka 1. laktace
 [4,] 3984.01080  ~ odhadnutá odchylka 2. laktace (odchylka 1436)
 [5,]  -87.54974        ~ OPH krávy č. 1
 [6,]   47.47015        ~ OPH krávy č. 2
 [7,]   53.43945        ~ OPH krávy č. 3
 [8,]  -53.43945        ~ OPH krávy č. 4
 [9,]   61.96703        ~ OPH krávy č. 5
[10,]   43.77487        ~ OPH otce krav č. 2 a 5
```

Závěry:

• Stáda se liší v chovatelské péči o 1972 kg mléka, druhá laktace převyšuje první o 1436 kg mléka. Nejlepší kráva je č. 5 (OPH = +62 kg) a nejhorší je kráva č. 1 (OPH = -88 kg). Genetický rozdíl mezi nimi je 150kg mléka.

• Kráva č. 4 je druhá nejhorší s plemennou hodnotou -53 kg mléka, přestože v rámci ledovaného souboru dosahuje nejvyšší užitkovost (8000 kg mléka). Při pozornějším ledování však zjistíme, že je na druhé laktaci, tzn., že její vysoká užitkovost je dána vyšším stupněm tělesné dospělosti (+ 1436 kg mléka) a je ve stádě s lepší chovatelskou péčí (+ 1927 kg mléka). Jestliže o tyto položky, které jsou dány technikou chovu, se praví její užitkovost, dostane se na podprůměrnou úroveň.

• Naopak její vrstevnice – kráva č. 5 – je na první laktaci a na druhé laktaci lze tedy u ní očekávat užitkovost 7000 + 1436 = 8436 kg mléka. Kráva č. 5 je proto po korekci +436 kg mléka lepší než kráva č. 4, což činí v plemenné hodnotě rozdíl (v odhadu rozdílu genetického založení) 115 kg mléka (62 + 53).

• U krav č. 2 a č. 5 jsou při odhadu plemenné hodnoty využity vlastní užitkovosti zároveň vzájemný příbuzenský vztah zásluhou společného otce. Jejich plemenné hodnoty jsou proto stanoveny přesněji než u ostatních krav. Plemenná hodnota otce e stanovena na základě užitkovosti těchto dcer a činí +44 kg mléka.

• Jak ukazuje příklad, nelze se při výběru do plemenitby řídit naměřenou užitkovostí, neboť ta je ovlivněna několika činiteli.

• Na základě odhadu plemenných hodnot dáme přednost zařazení do plemenitby krávám podle tohoto pořadí:

Rozdíly v užitkovostech působené chovatelským prostředím jsou mnohem větší, než genetické rozdíly mezi zvířaty. Naměřená užitkovost je ovlivněna větším počtem významných faktorů, a proto jsou soustavy rovnic složitější a zahrnují více efektů.

pořadí	kráva	OPH	užitkovost
1	5	+62	7000
2	3	+53	6500
3	2	+47	5000
4	4	-53	8000
5	1	-88	4500

Porovnejte s příkladem na konci přednášky č. 9