

3.1.2. Výpočet odhadu koeficientu dědivosti pomocí **analýzy variance** užitkových hodnot u příbuzných jedinců.

a) Analýza variance skupin polosourozenců

V jednom chovu byl sledován přírůstek živé hmotnosti u býčků – polosourozenců ze strany otce, ve věku tří měsíců. Náhodně bylo z této neinbrední populace vybráno 40 polosourozenců, po pěti otcích (vybalancovaný design pokusu). Zjistěte pomocí analýzy variance skupin polosourozenců odhad koeficientu dědivosti a jeho střední chybu.

n	O ₁	O ₂	O ₃	O ₄	O ₅
1	717	732	603	648	690
2	704	694	731	669	650
3	753	691	737	693	788
4	700	631	678	718	678
5	675	683	747	606	611
6	793	592	763	669	674
7	691	680	687	657	658
8	687	618	618	600	717
Σ	5720	5321	5564	5260	5466

Statistický model jednofaktorové analýzy variance:

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

y_{ij} – užitkovost j-tého potomka po i-tém otci
 μ – obecný průměr populace
 a_i – vliv i-tého otce
 e_{ij} – ostatní nahodilé vlivy

Předpočítané součty za skupiny podle otců $Y_{i\cdot}$, jejich druhé mocniny $Y_{i\cdot}^2$ a součty čtverců $\sum y_{ij}^2$.

skupina	$Y_{i\cdot}$	$Y_{i\cdot}^2$	$Y_{i\cdot}^2/n_i$	$\sum y_{ij}^2$
1	5720	32718400	4089800,00	4100638
2	5321	28313041	3539130,13	3554379
3	5564	30958096	3869762,00	3894894
4	5260	27667600	3458450,00	3469684
5	5466	29877156	3734644,50	3753878
$Y_{\cdot\cdot} =$	27331		18691786,63	18773473
$Y_{\cdot\cdot}^2 =$	746983561		$\sum Y_{i\cdot}^2/n_i$	$\sum \sum y_{ij}^2$

Počet otců $p = 5$

Celkový počet potomků $n = 40$

Vážený/průměrný počet potomků na otce $n_i = n_0 = 8$

Výpočet součtu čtverců odchylek od průměru:

- mezi otci

$$SS_a = \sum_{i=1}^p \frac{Y_{i\cdot}^2}{n_i} - \frac{Y_{\cdot\cdot}^2}{n} = \quad =$$

- uvnitř skupin podle otců (reziduální)

$$SS_e = \sum_{i=1}^p \sum_{j=1}^{m_j} y_{ij}^2 - \sum_{i=1}^p \frac{Y_{i\cdot}^2}{n_i} = \quad =$$

- pro celý pokus

$$SS_c = \sum_{i=1}^p \sum_{j=1}^{m_j} y_{ij}^2 - \frac{Y_{..}^2}{n} = 18773473 - (746983561/40) =$$

Tabulka analýzy variance:

Proměnlivost	SS	df	MS	složení MS
Mezi skupinami (a)	SS _a	p - 1	SS _a /df	= $\sigma_e^2 + n_0\sigma_g^2$
Uvnitř skupin (e)	SS _e	n - p	SS _e /df	= σ_e^2
Celková (c)				

Tady končí statistika a začíná genetika!

Výpočet odhadu variance genetické podle otců:

$$MS_a = \sigma_e^2 + n_0\sigma_g^2 = MS_e + n_0\sigma_g^2$$

$$\sigma_g^2 = \frac{MS_a - MS_e}{n_0} =$$

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2 =$$

$$\rho = r_i = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} =$$

$$h^2 = 4\rho = 4 \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 4 \frac{\sigma_g^2}{\sigma_p^2} =$$

Výpočet střední chyby odhadu koeficientu dědivosti:

a) jako čtyřnásobek střední chyby intraklasního korelačního koeficientu (při stejném počtu pozorování ve skupinách):

$$se_{h^2} = 4.s_\rho = 4 \cdot \sqrt{\frac{2 \cdot (1-\rho)^2 (1 + (n_0 - 1)\rho)^2}{n_0(n_0 - 1)(p - 1)}} =$$

b) na základě jeho velikosti váženém počtu jedinců ve skupině polosourozenců a počtu skupin polosourozenců:

$$se_{h^2} = \left(h^2 + \frac{4}{n_0} \right) \cdot \sqrt{\frac{2}{p}} =$$

$$h^2 \pm se_{h^2} = \pm$$

Odhad koeficientu dědivosti na základě výpočtu analýzy variance polosourozenců byl svou hodnotou a jeho střední chyba byla Tento příklad je nereprezentativní z důvodu nízkého počtu sledování polosourozenců. Vypočítaný odhad koeficientu dědivosti je nepoužitelný pro svou vysokou střední chybu odhadu.

Výsledek ANOVY vypočítaný pomocí zobecněného lineárního modelu GLM v programu SAS

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	17197.60000	4299.40000	1.84	0.1428
Error	35	81686.37500	2333.89643		
Corrected Total	39	98883.97500			

Výsledek odhadu otcovské variance pomocí smíšeného modelu metodou REML v programu SAS**The Mixed Procedure**

```

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

```

```

Class Level Information
Class  Levels  Values
otec   5      1 2 3 4 5

```

```

Iteration History

Iteration  Evaluations  -2 Res Log Like  Criterion
          0           1      420.05357726
          1           1      419.26631336  0.00000000

```

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm Estimate

```

otec      245.69
Residual  2333.90

```

<- variance genetická dle otců

ANOVA v programu R:

```

data6 <- read.table("K:/R/data6.csv", header=T, sep=";") #načte data do tabulky, bez hlavičky a odstraní středník
anova.data6 <- lm(potomek~otec, data=data6) # zápis pomocí lineárního modelu (lm)
anova(anova.data6) # ukáže souhrnnou tabulku ANOVA

```

```

Analysis of Variance Table
Response: potomek
          Df    Sum Sq  Mean Sq  F value  Pr(>F)
otec       4    17198    4299.4    1.8422  0.1428
Residuals 35    81686    2333.9

```

Smíšený lineární model (REML) v programu R:

```

lmer.data6 <- lmer(potomek ~ 1+(1|otec),data=data6)
summary(lmer.data6)

```

```

Linear mixed model fit by REML ['lmerMod']
Formula: potomek ~ 1 + (1 | otec)
Data: data6

REML criterion at convergence: 419.3

Scaled residuals:
    Min       1Q   Median       3Q      Max
-1.77733 -0.49544 -0.04508  0.50849  2.16799

Random effects:
 Groups   Name      Variance Std.Dev.
otec     (Intercept)  245.7    15.67
Residual                2333.9   48.31
Number of obs: 40, groups: otec, 5

Fixed effects:
              Estimate Std. Error t value
(Intercept)   683.27     10.37    65.91

```

b) Analýza variance skupin vlastních sourozenců a polosourozenců

Metoda analýzy variance skupin vlastních sourozenců a polosourozenců umožňuje simultánní výpočet odhadu koeficientu dědivosti, protože lze nezávisle na sobě odhadnout ze stejného sledování koeficienty dědivosti mezi polosourozenci ze strany otce nebo matky, a úplnými sourozenci na základě obou rodičů:

$$h_O^2 \quad h_M^2 \quad h_{O+M}^2$$

Byl sledován snáškový test nosnic v určitém chovu

Otec	Matka	Užitkovost potomků	$Y_{ij\bullet}$	n_{ij}
1	1	78, 55, 50, 79, 46, 91, 54, 79, 63, 47	642	10
	2	55, 66, 53, 75, 73, 63, 71, 61	517	8
	3	62, 81, 45, 60, 72, 63, 53	436	7
	4	69, 89, 58, 90, 65, 73, 95	539	7
	5	78, 76, 81, 80, 82, 49, 85, 91, 64	686	9
			$Y_{i\bullet\bullet} = 2820$	$n_i = 41$
2				
3				
.				
.				
5	21	72, 68, 88, 51, 47, 52, 71, 87	536	8
	22	95, 56, 73, 76, 91, 53, 91	535	7
	23	75, 92, 58, 50, 52, 87, 88, 69, 51	622	9
	24	87, 77, 66, 93, 56, 84, 94, 97	654	8
	25	75, 47, 77, 63, 62, 78, 85	487	7
	26	66, 52, 64, 66, 64, 92, 74, 63, 58, 84	683	10
			$Y_{i\bullet\bullet} = 3517$	$n_i = 49$
$Y_{\bullet\bullet\bullet} = 15\,904$		$\sum\sum\sum y_{ijk}^2 = 1\,216\,770$	$n = 218$	

počet otců: $p = 5$
 počet matek: $m = 26$
 počet potomků: $n = 218$

Model dvoufaktorové hierarchické analýzy variance:

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

y_{ij} – užitkovost j-tého potomka po i-tém otce
 μ – obecný průměr populace
 a_i – vliv i-tého otce
 b_{ij} – vliv j-té matky pod i-tým otcem
 e_{ij} – ostatní nahodilé vlivy

Výpočet mezihodnot:

$$\sum \sum \frac{Y_{ij\bullet}^2}{n_{ij}} = 1\,175\,398,79$$

$$\sum \frac{Y_{i\bullet\bullet}^2}{n_i} = 1\,165\,967,12$$

$$\frac{Y_{\bullet\bullet\bullet}^2}{n} = 1\,160\,262,46$$

Výpočet součtu čtverců odchylek od průměru:

- mezi skupinami otců

$$SS_a = \sum_{i=1}^p \frac{Y_{i\bullet}^2}{n_i} - \frac{Y_{\bullet\bullet}^2}{n} =$$

- mezi skupinami matek uvnitř otců

$$SS_b = \sum_{i=1}^p \sum_{j=1}^{m_i} \frac{Y_{ij\bullet}^2}{n_{ij}} - \sum_{i=1}^p \frac{Y_{i\bullet}^2}{n_i} =$$

- reziduální (mezi sourozenci uvnitř rodin)

$$SS_c = \sum_{i=1}^p \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} y_{ijk}^2 - \sum_{i=1}^p \sum_{j=1}^{m_i} \frac{Y_{ij\bullet}^2}{n_{ij}} =$$

- pro celý pokus

$$SS_c = \sum_{i=1}^p \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} y_{ijk}^2 - \frac{Y_{\bullet\bullet}^2}{n} =$$

Tabulka analýzy variance:

Zdroj proměnlivosti	Součet čtverců odchylek (SS)	Stupeň volnosti (df)	Průměrný čtverec (MS)	Složení MS ¹
- mezi otci	SS _a =	df _a = k - 1 =	MS _a =	$\sigma_e^2 + (m_2 \sigma_{g_M}^2) + (n_3 + \sigma_{g_o}^2)$
- mezi matkami uvnitř otců	SS _b =	df _b = p - k =	MS _b =	$\sigma_e^2 + (n_1 \sigma_{g_M}^2)$
- mezi potomky uvnitř skupin podle matek	SS _c =	df _c = n - p =	MS _c =	σ_e^2
- celková	SS _c =	df _c = n - 1 =		

- průměrný počet potomků pro matku: $n_1 = 218/26 = 8,38$ - průměrný počet matek pro otce: $m_2 = 26/5 = 5,20$ - průměrný počet potomků pro otce: $n_3 = 218/5 = 43,60$ **Tady končí statistika a začíná genetika!****Výpočet odhadu komponent variance:****a) genetická variance podle matek**

$$\sigma_{g_M}^2 = \frac{MS_b - MS_c}{n_1} =$$

b) genetická variance podle otců

$$\sigma_{g_o}^2 = \frac{MS_a - MS_c - m_2 \sigma_{g_M}^2}{n_3} =$$

Když $n_1 = m_2$, pak $\sigma_{g_o}^2 = \frac{MS_a - MS_b}{n_3}$

c) variance prostředí $\sigma_e^2 = MS_c =$ ¹ n_1 – vážený počet potomků na jednu matku m_2 – vážený počet matek na jednoho otce n_3 – vážený počet potomků na jednoho otce

Výpočet odhadu koeficientů dědivosti:

a) podle otců

$$h_O^2 = 4\rho_O = 4 \frac{\sigma_{g_O}^2}{\sigma_{g_O}^2 + \sigma_{g_M}^2 + \sigma_e^2} = 4 \frac{\sigma_{g_O}^2}{\sigma_P^2} =$$

Střední chyba koeficientu dědivosti:

$$se_{h_O^2} = \left(h_O^2 + \frac{4}{n_3} \right) \sqrt{\frac{2}{p}} =$$

b) podle matek

$$h_M^2 = 4\rho_M = 4 \frac{\sigma_{g_M}^2}{\sigma_{g_O}^2 + \sigma_{g_M}^2 + \sigma_e^2} = 4 \frac{\sigma_{g_M}^2}{\sigma_P^2} =$$

Střední chyba koeficientu dědivosti:

$$se_{h_M^2} = \left(h_M^2 + \frac{4}{n_1} \right) \sqrt{\frac{2}{m}} =$$

c) podle matek i otců

$$h_{O+M}^2 = 2\rho_{O+M} = 2 \frac{\sigma_{g_O}^2 + \sigma_{g_M}^2}{\sigma_{g_O}^2 + \sigma_{g_M}^2 + \sigma_e^2} = 2 \frac{\sigma_{g_O}^2 + \sigma_{g_M}^2}{\sigma_P^2} =$$

Střední chyba koeficientu dědivosti:

$$se_{h_{O+M}^2} = 4 \sqrt{h_{O+M}^2} \sqrt{\frac{1}{n}} =$$

$h_O^2 \pm se_{h_O^2} =$
$h_M^2 \pm se_{h_M^2} =$
$h_{O+M}^2 \pm se_{h_{O+M}^2} =$

Zjištěné odhady koeficientů dědivosti na základě analýzy variance skupin úplných sourozenců a polosourozenců byly svými hodnotami a jejich střední chyby byly v důsledku sledování

Výpočet v SASu

Výsledek při použití procedury GLM v SAS

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	15136.32859	605.45314	2.81	<.0001
Error	192	41371.21270	215.47507		
Corrected Total	217	56507.54128			

R-Square	0.267864	Coeff Var	20.12096	Root MSE	14.67907	y Mean	72.95413
----------	----------	-----------	----------	----------	----------	--------	----------

Source	DF	Type I SS	Mean Square	F Value	Pr > F
otec	4	5704.659442	1426.164861	6.62	<.0001
matka(otec)	21	9431.669144	449.127102	2.08	0.0051

Výsledek při použití procedury Mixed v SAS, metodou REML

The Mixed Procedure
Model Information

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

The Mixed Procedure

Covariance Parameter
Estimates

Cov Parm	Estimate
otec	21.4241
otec(matka)	28.8662
Residual	215.02

V programu R:

```
data7 <- read.table("K:/R/data7.csv", header=T, sep=";") #načte data, bez hlavičky a odstraní středník
anova.data7 <- lm(y~ OTEC/MATKA, data=data7) # zápis pomocí lineárního modelu (lm)
anova(anova.data7) # ukáže souhrnnou tabulku ANOVA
```

```
Analysis of Variance Table
Response: y
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
OTEC	4	5705	1426.16	6.6187	5.199e-05 ***
OTEC:MATKA	21	9432	449.13	2.0844	0.005103 **
Residuals	192	41371	215.48		

```
lmer.hierarchy <- lmer(y ~ 1+(1|OTEC/MATKA),data=data7)
summary(lmer.hierarchy)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ 1 + (1 | OTEC/MATKA)
Data: data7

REML criterion at convergence: 1809.8

Scaled residuals:
    Min       1Q   Median       3Q      Max
-1.9314 -0.8292 -0.0625  0.8188  2.0051

Random effects:
 Groups      Name          Variance Std.Dev.
MATKA:OTEC (Intercept)  28.87    5.373
OTEC       (Intercept)  21.42    4.629
Residual                   215.02   14.663
Number of obs: 218, groups: MATKA:OTEC, 26; OTEC, 5

Fixed effects:
              Estimate Std. Error t value
(Intercept)  73.701     2.544    28.97
```

3.2.2. U 250 prasnic ve velkochovu byly sledovány počty všech narozených selat za jejich první čtyři vrhy. Vypočítejte odhad **koeficientu opakovatelnosti** této užitkové vlastnosti včetně jeho střední chyby.

Tabulka výsledků jednofaktorové analýzy variance:

Proměnlivost	SS	df	MS	složení MS
Mezi skupinami jedinců mezi prasnicemi (a)	1681,99	$df_a = p - 1 = 249$	6,777	$= \sigma_e^2 + n_0 \sigma_g^2$
Uvnitř skupin (e)	3044,25	$df_e = n - p = 750$	4,059	$= \sigma_e^2$
Celková (c)	4756,24	$df_c = n - 1 = 999$	-	

$p = 250$ počet prasnic
 $n = 1000$ počet sledovaných vrhů
 $k = 4$ počet opakování u jedné prasnice, zde platí: $k = n_0$ vážený počet potomků

$$\text{Při nestejném počtu sledování je nutno } n_0 \text{ zjistit: } n_0 = \frac{1}{p-1} \left(n - \frac{\sum_i n_i^2}{n} \right)$$

Odhad variance genetické: $MS_a = \sigma_e^2 + n_0 \sigma_g^2$

$$\sigma_g^2 = \frac{MS_a - MS_e}{n_0} =$$

Odhad variance prostřed'ové: $\sigma_e^2 = MS_e$

Odhad variance fenotypové: $\sigma_p^2 = (\sigma_G^2 + \sigma_{Ep}^2 + \sigma_{GEp}^2) + \sigma_{Et}^2 + \sigma_{GEt}^2 = \sigma_g^2 + \sigma_e^2$

$$\sigma_p^2 =$$

Výpočet odhadu intraklasního koeficientu korelace $\rho = r_{op}$ - **koeficientu opakovatelnosti**:

$$r_{op} = \frac{\sigma_g^2}{\sigma_p^2} =$$

Stanovení významnosti intraklasního koeficientu korelace pomocí F-testu:

$$F = \frac{MS_a}{MS_e} =$$

Tabulkové hodnoty: $F_{(249;750;0,05)} = 1,22$ a $F_{(249;750;0,01)} = 1,32$

Hodnota intraklasního koeficientu korelace je vysoce průkazná.

Výpočet **střední chyby koeficientu opakovatelnosti** (při stejném počtu opakování $n_0 = k$):

$$se_{r_{op}} = \sqrt{\frac{2 \cdot (1 - \rho)^2 (1 + (n_0 - 1)\rho)^2}{n_0 (n_0 - 1)(p - 1)}} =$$

$$\text{Při nestejném počtu sledování je: } se_{r_{op}} = \sqrt{\frac{2 \cdot (n - 1)(1 - \rho)^2 (1 + (n_0 - 1)\rho)^2}{n_0^2 (n - p)(p - 1)}}$$

Výsledná hodnota: $r_{op} \pm se_{r_{op}} = \pm$

² někde se uvádí i určení stupňů volnosti pro $df_c = p(k-1)$ (výsledek je však stejný jako $n-p$)