

Odhad heritability pomocí analýzy dvojčat

prof. Ing. Tomáš Urban, Ph.D.
urban@mendelu.cz

Dvojčata

Jedinci s identickým genotypem ~ podobně jako inbrední linie

Studie dvojčat (lidé, skot)

Zohlednění společného prostředí – od zygoty po narození, a v dalším období, kdy jsou spolu -> variance mezi páry dvojčat obsahuje varianci společného prostředí (V_{EC}), spojená s V_G

Tento problém lze řešit analýzou dvou typů dvojčat:

monozygotní (MZ) a dizygotní (DZ)

MZ ~ úplní sourozenci

DZ ~ polosourozenci

- ve společném prostředí mají přibližně stejný rozměr jako MZ

Odhad množství genetické variance, ptáme se nakolik méně jsou si podobní DZ než MZ?

Složení komponent variance mezi páry dvojčat

	Mezi páry, σ_b^2	Uvnitř párů, σ_w^2
MZ dvojčata	$V_A + V_D + V_{Ec}$	V_{Ew}
DZ dvojčata	$\frac{1}{2} V_A + \frac{1}{4} V_D + V_{Ec}$	$\frac{1}{2} V_A + \frac{3}{4} V_D + V_{Ew}$
Rozdíl ($MS_{MZ} - MS_{DZ}$)	$\frac{1}{2} V_A + \frac{3}{4} V_D$	$\frac{1}{2} V_A + \frac{3}{4} V_D$

Za předpokladu, že obě složky variance prostředí, V_{Ec} a V_{Ew} , jsou stejné pro MZ a DZ

V_I se neuvažuje; celková genetická variance je stejná u MZ a u DZ

Rozdíly mezi MZ a DZ dvojčaty v obou komponentách odhadují $\frac{1}{2} V_A$ společně se $\frac{3}{4} V_D$

$$(\sigma_{bMZ}^2 - \sigma_{bDZ}^2) + (\sigma_{wMZ}^2 - \sigma_{wDZ}^2)$$

Korelace mezi páry dvojčat je podílem komponenty variance mezi páry fenotypovou variancí ~ dvojnásobek rozdílu mezi MZ korelace a DZ korelace:

$$\text{„Heritabilita“} = \frac{(V_A + \frac{1}{2} V_D)}{V_P}$$

Výpočet se více blíží heritabilitě v širším smyslu, než v úzkém smyslu.

Genetika kvantitativních znaků

Další genetické parametry

- koeficient opakovatelnosti
- genetické korelace

Koeficient opakovatelnosti

Opakované měření stejné vlastnosti na stejném jedinci v průběhu jeho života (nejlépe za stejných podmínek) → podobnost měření je závislá na velikosti genetické determinace.

Koeficient opakovatelnosti udává podíl proměnlivosti zapříčiněný genetickými rozdíly z celkové fenotypové proměnlivosti.

- Na stejném místě (prostorové opakování, topografická)
- Opakování v časovém období (s věkem)

Koeficient opakovatelnosti - r_{op}

$$S_p^2 = S_g^2 + S_{E_p}^2 + S_{GE_p}^2 + S_{E_T}^2 + S_{GE_T}^2$$

$\underbrace{\hspace{10em}}_{S_g^2} \qquad \underbrace{\hspace{10em}}_{S_e^2}$

Měření jsou **rozdílná** → vlivy dočasné jsou různé

Měření jsou **podobná** → vlivy stálé jsou stejné

stálé vlivy: $s = S_g^2 + S_{E_p}^2 + S_{GE_p}^2 = \boxed{S_g^2}$
podobnost

Zdroj proměnlivosti mezi jedinci

dočasné vlivy: $e = S_{E_T}^2 + S_{GE_T}^2 = \boxed{S_e^2}$
rozdílnost

Zdroj proměnlivosti mezi měřeními u jedince

Korelace mezi fenotypovými hodnotami různých měření u všech jedinců populace je OPAKOVATELNOST:

koeficient opakovatelnosti: r_{op}

$$r_{op} = r = \frac{S_g^2}{S_p^2} = \frac{S_g^2 + S_{E_p}^2 + S_{GE_p}^2}{S_p^2} = h_s^2 + \frac{S_{E_p}^2 + S_{GE_p}^2}{S_p^2}$$

$$h^2 = r_{op} - \frac{\sigma_{E_p}^2 + \sigma_{GE_p}^2}{\sigma_p^2}$$

$$h_u^2 < h_s^2 < r_{op}$$

r_{op} je horní hranicí h^2

Př.: 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.

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	σ_e^2
Celková (c)	4756,24	$df_c = n - 1 = 999$	-	

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

$$MS_e = \sigma_e^2$$

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

$$\text{Odhad variance genetické: } \sigma_g^2 = \frac{MS_a - MS_e}{n_0} = 0,6740$$

$$\text{Odhad variance fenotypové: } \sigma_p^2 = (\sigma_G^2 + \sigma_{Ep}^2 + \sigma_{GEp}^2) + \sigma_{Et}^2 + \sigma_{GEt}^2 = \sigma_g^2 + \sigma_e^2 = 4,7330$$

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

$$r_{op} = \frac{\sigma_g^2}{\sigma_p^2} = 0,1424$$

$$se_{r_{op}} = \sqrt{\frac{2 \cdot (1 - \rho)^2 (1 + (n_0 - 1)\rho)^2}{n_0(n_0 - 1)(p - 1)}} = 0,0317$$

$$Y_{ijkl} = HY_i + M_j + a_k + p_k + b_1 x_{ijkl} + b_2 (x_{ijkl})^2 + e_{ijkl}$$

To estimate variances with the **repeatability model, REML**, was used.

Y_{ijkl} is milk, fat, or protein yield with twice daily milking and complete records of length 240 to 305 d for cow k in herd-year i, calving month j, and lactation l;

HY_i is effect of herd-year i;

M_j is effect of calving month j ;

x_{ijkl} is the cow age at calving;

a_k is additive genetic effect of cow k;

p_k is the permanent environmental effect of cow k;

b_1 and b_2 are the partial regression coefficients for linear and quadratic effects of age at calving;

e_{ijkl} is random residual associated with each record.

TABLE 3. Estimated variance components,¹ heritabilities, and repeatabilities (R) of milk yield for 10 data files with means and empirical standard errors.

Subsets	$\hat{\sigma}_a^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}_e^2$	h^2	R
1	428,434	342,251	705,362	.29	.52
2	392,230	308,267	652,472	.29	.52
3	376,293	261,408	608,725	.30	.51
4	378,277	305,932	703,141	.27	.49
5	452,460	291,879	708,161	.31	.51
6	376,338	292,226	666,288	.28	.50
7	489,450	345,570	597,150	.34	.58
8	435,667	411,703	545,908	.34	.61
9	288,459	288,008	543,053	.26	.51
10	480,144	444,486	522,103	.33	.64
\bar{X}	409,775	329,173	625,236	.30	.54
SE	17,999	17,484	21,573	.01	.01

¹ $\hat{\sigma}_a^2$ = Estimated variance component for additive genetic effects (kilograms squared), $\hat{\sigma}_p^2$ = estimated variance component for permanent environmental effects (kilograms squared), and $\hat{\sigma}_e^2$ = estimated error variance (kilograms squared).

TABLE 5. Estimated variance components,¹ heritabilities, and repeatabilities (R) of protein yield for 10 data files with means and empirical standard errors.

Subsets	$\hat{\sigma}_a^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}_e^2$	h^2	R
1	317.0	321.4	616.3	.25	.51
2	298.9	276.2	569.7	.26	.50
3	280.3	237.2	527.3	.27	.50
4	268.3	296.8	604.4	.23	.48
5	354.7	271.5	604.1	.29	.51
6	273.3	277.3	573.1	.24	.49
7	358.5	403.7	496.6	.28	.61
8	427.7	371.6	456.0	.34	.64
9	227.0	241.3	461.8	.24	.50
10	175.6	247.2	451.2	.20	.48
\bar{X}	298.1	294.4	536.1	.26	.52
SE	21.4	16.8	19.7	.01	.02

¹ $\hat{\sigma}_a^2$ = Estimated variance component for additive genetic effects (kilograms squared), $\hat{\sigma}_p^2$ = estimated variance component for permanent environmental effects (kilograms squared), and $\hat{\sigma}_e^2$ = estimated error variance (kilograms squared).

TABLE 4. Estimated variance components,¹ heritabilities, and repeatabilities (R) of fat yield for 10 data files with means and empirical standard errors.

Subsets	$\hat{\sigma}_a^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}_e^2$	h^2	R
1	662.2	451.3	1077.0	.29	.51
2	576.8	393.5	943.3	.30	.51
3	552.0	339.7	890.3	.31	.50
4	511.2	450.4	973.2	.26	.50
5	678.3	372.1	982.2	.33	.52
6	463.7	446.9	926.9	.25	.50
7	811.9	476.4	833.4	.38	.61
8	633.4	485.7	768.9	.34	.59
9	423.1	394.0	787.6	.26	.51
10	345.3	390.8	786.8	.23	.48
\bar{X}	565.8	420.1	897.0	.30	.52
SE	41.2	14.5	30.5	.01	.01

¹ $\hat{\sigma}_a^2$ = Estimated variance component for additive genetic effects (kilograms squared), $\hat{\sigma}_p^2$ = estimated variance component for permanent environmental effects (kilograms squared), and $\hat{\sigma}_e^2$ = estimated error variance (kilograms squared).

Význam r_{op}

Určuje horní hranici koeficientu heritability (obsahuje i vlivy dominance a interakce genů, navíc i složku stálých vlivů prostředí a interakci genů a prostředí).

Pro odhad není nutné mít skupiny příbuzných jedinců.

Upřesňuje stanovení skutečně geneticky podmíněné užitkové hodnoty (odhad genotypové odchylky).

Pro výpočet odhadu koeficientu heritability.

Zpřesnění účinnosti selekce.

Fenotypové, genetické, genotypové a paratypové korelace

Vztahy mezi vlastnostmi

- **pozitivní** – zvýšením jednoho znaku se zvyšuje i druhý
- **negativní** – zvýšením jednoho znaku se snižuje druhý

Fenotypová korelace

- závislost mezi pozorovanými hodnotami P_x a P_y je daná kombinací závislostí genotypových hodnot a účinků prostředí;
- korelační páry = dvojice měření znaku x a znaku y u téhož jedince;

r_P

$r_{P_x P_y}$

Genotypová a genetická korelace

genotypová = závislost mezi genotypovými hodnotami znaků x a y ;

r_G ;

r_{GxGy} ;

genetická = závislost aditivních hodnot obou znaků – častější;

- vyjadřuje rozsah, ve kterém dvě měření odrážejí, co je geneticky stejná vlastnost

(délka křídel : délka těla $r = 0,75$)

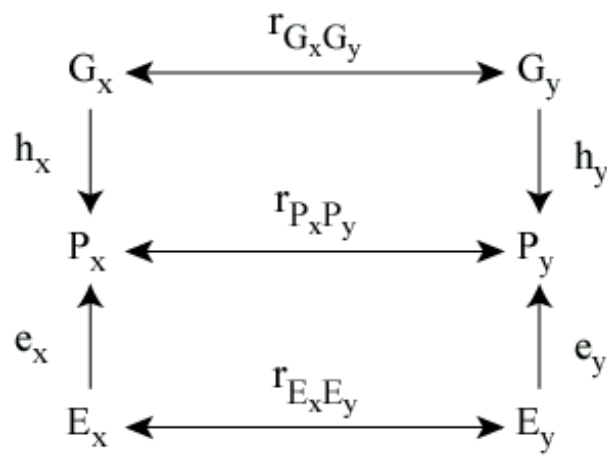
Paratypová korelace

Prostřed'ová korelace

Závislost mezi efekty prostředí

r_E

r_{ExEy}



Ze schématu vyplývá rozklad fenotypové korelace:

$$r_{P_x P_y} = h_x \cdot h_y \cdot r_{G_x G_y} + e_x \cdot e_y \cdot r_{E_x E_y}$$

$$r_{P_x P_y} = h_x \cdot h_y \cdot r_{G_x G_y} + \sqrt{(1 - h_x^2)(1 - h_y^2)} \cdot r_{E_x E_y}$$

Výpočet korelace – obecně

Zohledňuje variance (rozptyly) obou vlastností a jejich vzájemnou kovarianci:

$$r_{P_x P_y} = \frac{\sum [(x - \bar{x}) \cdot (y - \bar{y})]}{\sqrt{\sum (x - \bar{x})^2 \cdot \sum (y - \bar{y})^2}}$$

Zjednodušeně: $r_{P_x P_y} = \frac{\text{COV}_{xy}}{\sqrt{\text{var}(x) \cdot \text{var}(y)}}$

Vznik genetických korelací

Pleiotropní působení genů

= gen ovlivňuje více znaků ve stejném čase

Vazba genů

= geny lokalizovány v jedné vazbové skupině

- čím blíže, tím je vazba silnější
- není stálá – crossing over
- vazba mezi skupinami genů

• zavedení genů do populace

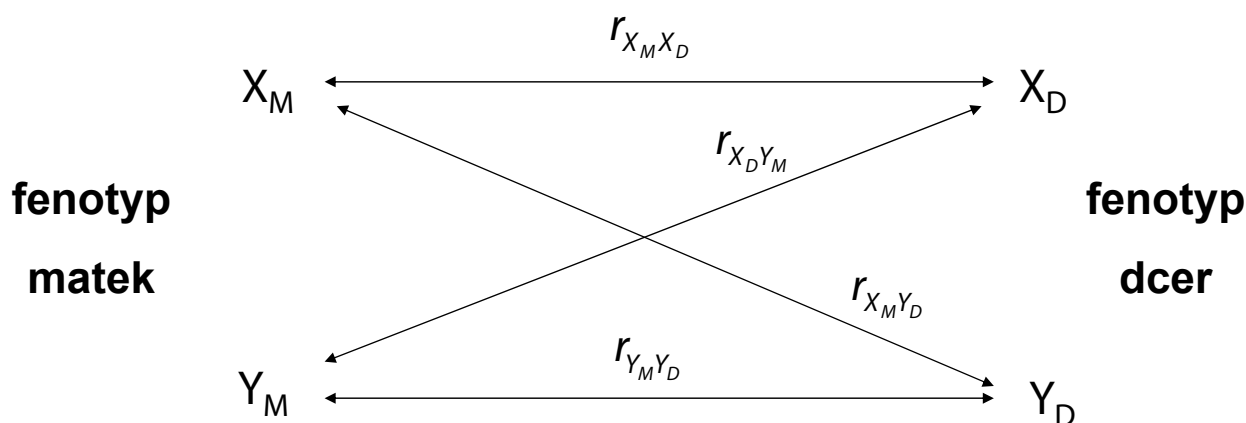
- ~ dočasné korelace
- dané intenzivním využíváním vybraných jedinců
- Při změně intenzity rozmnožování určitých jedinců genetické korelace zanikají
- narušení náhodným pářením

Metody odhadu genetické korelace

- metoda korelace křížem
- analýza variance a kovariance
- realizovaná genetická korelace na základě výsledku selekce

1. Korelace křížem

- známe fenotypovou hodnotu znaku x, y u rodiče a potomka



$$r_{G_X G_Y} = \frac{r_{X_D Y_M} + r_{X_M Y_D}}{2 \cdot \sqrt{r_{X_M X_D} \cdot r_{Y_M Y_D}}}$$

Křížová korelace různých vlastností u matek a dcer

Křížová korelace stejných vlastností u matek a dcer

2. Analýza variance a kovariance

- Jednofaktorová
- Dvoufaktorová
- u skupin příbuzných jedinců
- Cíl: určit **variance a kovariance** genetické a prostředové
- Počítáme 2 ANOVY pro 2 vlastnosti a navíc analýzu kovariance současně pro obě vlastnosti

$$r_{G_x G_y} = \frac{\text{cov}(G_{xy})}{\sqrt{\text{var}(G_x) \cdot \text{var}(G_y)}}$$

$$r_{E_x E_y} = \frac{\text{cov}(E_{xy})}{\sqrt{\text{var}(E_x) \cdot \text{var}(E_y)}}$$

$$r_{P_x P_y} = \frac{\text{cov}(P_{xy})}{\sqrt{\text{var}(P_x) \cdot \text{var}(P_y)}}$$

Vztah mezi obsahem bílkovin (x) a tuku (y) v mléce dojnic byl sledován v osmi chovech (k), u skupin polosester (b). Výpočet odhadu byl proveden u 487 dojnic (n), dcer po 69 otců (b) bez ohledu na pořadí laktace.

$$\begin{aligned} \text{\% bílkovin} & \quad x_{ijk} = \mu + a_i + b_{ij} + e_{ijk} \\ \text{\% tuku} & \quad y_{ijk} = \mu + a_i + b_{ij} + e_{ijk} \end{aligned}$$

Zdroj proměnlivosti	součet čtverců		stupně volnosti	součet součinnů	střední čtverec (variance)		střední produkt (kovariance)
	SS _x	SS _y	F	SP _{xy}	MS _x	MS _y	MP _{xy}
- mezi chovy	SS _{a_x} = 4,7959	SS _{a_y} = 3,6014	f _a = k - 1 = 7	SP _{a_{xy}} = 2,2283	MS _{a_x} = 0,685128	MS _{a_y} = 0,514485	MP _{a_{xy}} = 0,318328
- mezi otci uvnitř chovů	SS _{o_x} = 6,1125	SS _{o_y} = 12,6068	f _o = b - k = 61	SP _{o_{xy}} = 4,4461	MS_{o_x} = 0,100204	MS_{o_y} = 0,206668	MP_{o_{xy}} = 0,072887
- mezi dojnicemi uvnitř skupin otců	SS _{e_x} = 18,1950	SS _{e_y} = 32,5841	f _e = n - b = 418	SP _{e_{xy}} = 10,1466	MS_{e_x} = 0,043528	MS_{e_y} = 0,077952	MP _{e_{xy}} = 0,024274
- celková	SS _{c_x} = 29,1034	SS _{c_y} = 48,7923	f _c = n - 1 = 486	SP _{c_{xy}} = 16,8210	-	-	-

Protože ve skupinách byly různé počty pozorování, tak byl vypočítán vážený počet pozorování $n_0 = 6,6494$.

$$\sigma_{g_x}^2 = \frac{MS_{O_x} - MS_{e_x}}{n_0} = 0,00852$$

$$\sigma_{g_y}^2 = \frac{MS_{O_y} - MS_{e_y}}{n_0} = 0,01936$$

$$\text{cov}_{g_{xy}} = \frac{\text{cov}_{O_{xy}} - \text{cov}_{e_{xy}}}{n_0} = \frac{MP_{O_{xy}} - MP_{e_{xy}}}{n_0} = 0,00731$$

$$\hat{r}_{G_x G_y} = \frac{\text{COV}_{g_{xy}}}{\sqrt{\sigma_{g_x}^2 \cdot \sigma_{g_y}^2}} = 0,569$$

$$\hat{r}_{E_x E_y} = \frac{\text{COV}_{e_{xy}} - 3\text{COV}_{g_{xy}}}{\sqrt{\sigma_{e_x}^2 - 3\sigma_{g_x}^2} \sqrt{\sigma_{e_y}^2 - 3\sigma_{g_y}^2}} = \frac{MP_{e_{xy}} - 3\text{COV}_{g_{xy}}}{\sqrt{MS_{e_x} - 3\sigma_{g_x}^2} \sqrt{MS_{e_y} - 3\sigma_{g_y}^2}} = 0,1239$$

$$\hat{r}_{P_x P_y} = \frac{\text{COV}_{c_{xy}}}{\sqrt{\sigma_{c_x}^2 \cdot \sigma_{c_y}^2}} = \frac{SP_{c_{xy}}}{\sqrt{SS_{c_x} \cdot SS_{c_y}}} = 0,4464$$

Správnost výpočtu odhadů korelací je možno **ověřit rozkladem fenotypové korelace**

$$h_x^2 = 4 \cdot \rho = 4 \frac{\sigma_{g_x}^2}{\sigma_{g_x}^2 + \sigma_{e_x}^2} = 4 \frac{\sigma_{g_x}^2}{\sigma_{P_x}^2} = 0,655 \quad h_x = \sqrt{h_x^2} = 0,8093 \quad e_x = \sqrt{e_x^2} = \sqrt{1 - h_x^2} = 0,5874$$

$$h_y^2 = 4 \cdot \rho = 4 \frac{\sigma_{g_y}^2}{\sigma_{g_y}^2 + \sigma_{e_y}^2} = 4 \frac{\sigma_{g_y}^2}{\sigma_{P_y}^2} = 0,7957 \quad h_y = \sqrt{h_y^2} = 0,8921 \quad e_y = \sqrt{e_y^2} = \sqrt{1 - h_y^2} = 0,4520$$

$$\hat{r}_{P_x P_y} = h_x \cdot h_y \cdot r_{G_x G_y} + e_x \cdot e_y \cdot r_{E_x E_y} = 0,4438$$

3. Z výsledku selekce

- Realizovaná genetická korelace 2 vlastností na základě výsledků selekce, podle výše genetického zisku
 - selekční pokus
 - selekce v běžných podmínkách

a) Selekční pokus

- $\Delta G_{X/YS}$ genetický zisk v selektované části populace
- $\Delta G_{X/Y}$ genetický zisk v neselektované části populace

$$r_{G_X G_Y} = \frac{\Delta G_{XS}}{\Delta G_Y} \cdot \frac{\Delta G_{YS}}{\Delta G_X}$$

b) Selekcce v běžných podmínkách

- provádíme selekci dle jednoho znaku
- stanovíme dosažený genetický zisk pro oba znaky
- známe-li heritabilitu obou znaků, pak při selekci podle X

$$r_{G_X G_Y} = \frac{\Delta G_Y}{i_X h_X \sigma_{G_Y}} = \frac{\Delta G_Y}{i_X h_X h_Y \sigma_{P_Y}}$$

Korelovaný selekční efekt

- o kolik se v generaci po selekci na znak X změní znak Y, který nebyl předmětem selekce
- vlastnosti ve vazbě (skot: mléčné složky)

a) *přímá selekce* – selekční zisk za jednu generaci lze získat za pomoci vztahu:

$$\Delta G_x = i \cdot h_x^2 \cdot \sigma_{Px}$$

$$\Delta G_y = i \cdot h_y^2 \cdot \sigma_{Py}$$

b) *nepřímá selekce* – selekční zisk za jednu generaci lze získat pomocí vztahu **korelovaného selekčního zisku (efektu)**. Při selekci na jeden znak, lze ze znalosti genetické korelace s druhým znakem predikovat o kolik se změní druhý znak, na který neprobíhala selekce.

$$\begin{aligned} \text{kor. } \Delta G_y &= \Delta G_x \cdot b_{G_{xy}} \\ b_{G_{yx}} &= r_{GxGy} \cdot \frac{\sigma_{Gy}}{\sigma_{Gx}} \end{aligned}$$

$$\begin{aligned} \text{kor. } \Delta G_x &= \Delta G_y \cdot b_{G_{xy}} \\ b_{G_{xy}} &= r_{GxGy} \cdot \frac{\sigma_{Gx}}{\sigma_{Gy}} \end{aligned}$$

$$\text{kor. } \Delta G_x = i \cdot h_x \cdot h_y \cdot r_{GxGy} \cdot \sigma_{Px}$$

$$\text{kor. } \Delta G_y = i \cdot h_x \cdot h_y \cdot r_{GxGy} \cdot \sigma_{Py}$$

Příklady genetických korelací

skot	r_P	r_G	r_E
produkce mléka : % tuku	-0,26	-0,38	-0,18
prod. mléka v 1 : 2 laktaci	0,40	0,75	0,26
prasata			
přírůstek : hřbetní tuk	0,00	0,13	-0,18
přírůstek : výkrmnost	0,66	0,69	0,64
drůbež			
hm. těla : hm. vejce	0,33	0,42	0,23
hm. těla : snáška	0,01	-0,17	0,08