

Heritabilita II

- metody odhadu koeficientu heritability

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

Variance kvantitativní vlastnosti

Projev kvantitativní vlastnosti je ovlivněn genetickými a prostředovými faktory: $P = G + E$

Základní ?

Kolik z celkové variance v populaci je způsobeno genetickými příčinami a kolik prostředovými

$$V_P = V_G + V_E \quad h^2 = \frac{V_A}{V_P} = \frac{V_A}{V_A + V_D + V_I + V_E}$$

Odhalení genetické komponenty variance (určuje stupeň, kterým vlastnosti odpovídají na selekci)

Komponenty (~ složky) variance

Odhad genetických parametrů je synonymem pro odhad komponent pozorovatelné variance (včetně kovariancí) mezi příbuznými jedinci do příčinných komponent (V_A , V_D , V_I , V_{EP} , V_{ET})

Využití znalosti stupně příbuznosti mezi jedinci a z toho vyplývající očekávané kovariance mezi nimi

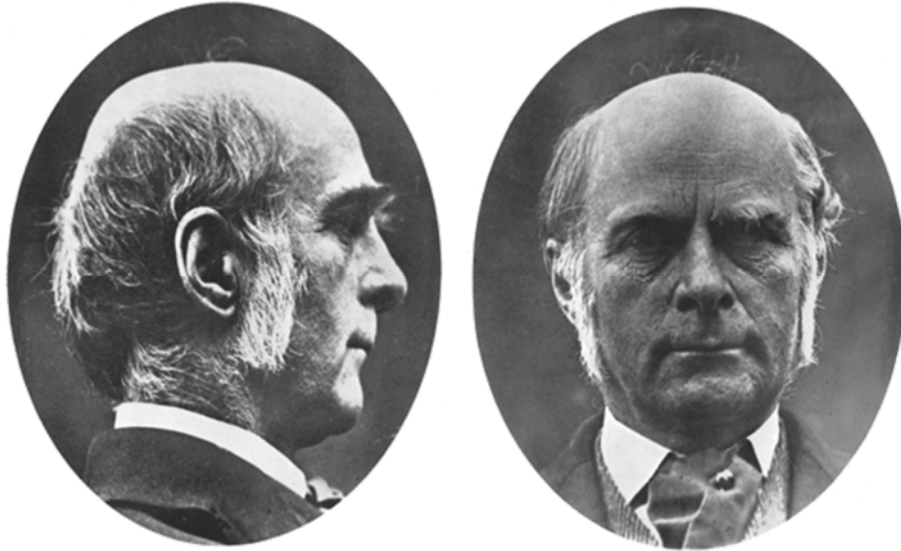
Proč předpokládáme, že komplexní vlastnost může mít genetický základ?

Protože je podobnost mezi příbuznými jedinci pro tu vlastnost!

Francis Galton F.R.S. 1822-1911

Chtěl kvantifikovat tyto rodinné podobnosti.

Založil biometrický přístup ke genetice (regrese a korelace)



Vymyslel korelační koeficient (r)

- Možno měřit stupeň asociace pro proměnnou mezi 2 příbuznými jedinci

Pro vlastnost, která je kompletně geneticky determinovaná, s malým nebo bez vlivu prostředí, se očekává, že r bude stejný jako koeficient příbuznosti

Koeficient příbuznosti vyjadřuje podíl genů sdílených 2 jedinci.

Příklady koeficientů příbuznosti

1. stupeň (rodič, sourozenci, dítě)	$\frac{1}{2}$
2. Stupeň (strýc, neteř, vnuk)	$\frac{1}{4}$
3 . stupeň (bratranec, pra-)	$\frac{1}{8}$
4 . stupeň (bratranec z 2. kolena)	$\frac{1}{16}$

Metody pro rozdělení fenotypové variance do dílčích komponent

Princip: fenotypová podobnost mezi příbuznými jedinci poskytuje informaci o stupni genetických rozdílů mezi jedinci

- aditivní komponenta primárně určuje stupeň, kterým se potomci podobají rodičům
 - tak ovládá míru odpovědi vlastnosti na selekci (ΔG)

$$V_A \sim V_{PH}$$

Genetická podobnost

Aditivně genetická podobnost '*additive genetic relationship (A)*' mezi jedinci x a y je a_{xy}

Podíl alel, které jedinci x a y mají společné (IBD)

Rodič – potomek $a = 0,5$

polosourozenci $a = 0,25$

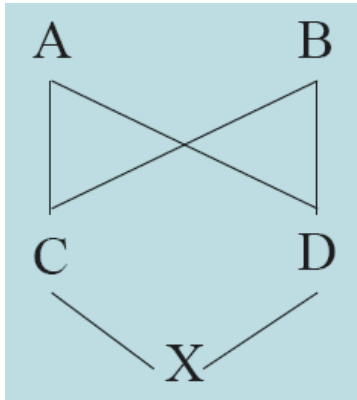
Příbuznost

Zdroj	Stupeň příbuznosti	Regresní koeficient
	R	b
Vlastní užítkovost	1	h^2
Rodič	$\frac{1}{2}$	$\frac{1}{2} h^2$
Prarodič	$\frac{1}{4}$	$\frac{1}{4} h^2$
Praprarodič	$\frac{1}{8}$	$\frac{1}{8} h^2$
VI. sourozenci – dvojče	1	h^2
Vlastní sourozenec	$\frac{1}{2}$	$\frac{1}{2} h^2$
Polosourozenec	$\frac{1}{4}$	$\frac{1}{4} h^2$
Potomek	$\frac{1}{2}$	$\frac{1}{2} h^2$

Aditivně genetická příbuznost

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

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



	A	B	C	D	X
A	1	0	0,5	0,5	0,5
B	0	1	0,5	0,5	0,5
C	0,5	0,5	1	0,5	0,75
D	0,5	0,5	0,5	1	0,75
X	0,5	0,5	0,75	0,75	1,25

$$a_{CD} = (0,5 * 0,5) + (0,5 * 0,5) = 0,5$$

- Metodou úsekových koeficientů

$$= 0,5(a_{CA} + a_{CB}) = 0,5(0,5 + 0,5) = 0,5$$

- Pomocí vzorců nad maticí

$$a_{XX} = 1 + (a_{cd}/2) = 1,25$$

Genetická podobnost

koeficient příbuznosti („*coefficient of relationship*“) (r_{xy})

Podobný k a_{xy} , ale je upraven pro inbriding rodičů (0 – 1)

$$r_{xy} = \frac{a_{xy}}{\sqrt{(1 + F_x)(1 + F_y)}}$$

Podobnost mezi příbuznými jedinci

Měření stejné vlastnosti na dvou příbuzných jedincích X a Y

$$P_X = A_X + E_X$$

$$P_Y = A_Y + E_Y$$

Korelace mezi fenotypy stejné vlastnosti existuje pro pár příbuzných jedinců v důsledku **podobnosti genů** a **podobného prostředí**

Kovariance mezi PH stejné vlastnosti příbuzných jedinců = genetická příbuznost x genetická variance

$$\text{COV}_{A_X A_Y} = r_{XY} \cdot V_A$$

$$r_{xy} = \frac{\text{COV}_{x,y}}{\sigma_x \sigma_y} \quad \text{Obecný vzorec korelace}$$

$$r_{A_X A_Y} = \frac{\text{COV}_{A_X A_Y}}{\sigma_{A_X} \sigma_{A_Y}} = \frac{\text{COV}_{A_X A_Y}}{\sigma_A^2} = \frac{r_{XY} V_A}{V_A}$$

Korelace mezi PH příbuzných jedinců je rovna koeficientu příbuznosti

$$r_{P_X P_Y} = \frac{\text{COV}_{A_X A_Y} + \text{COV}_{E_X E_Y}}{\sigma_{P_X} \sigma_{P_Y}} = \frac{\text{COV}_{A_X A_Y}}{V_P} = \frac{r_{XY} V_A}{V_P} = r_{XY} h^2$$

$$\text{COV}_{E_X E_Y} = 0$$

Předpoklad: zvířata jsou chována v různém prostředí bez prostře. korelace

Korelace mezi fenotypy příbuzných jedinců je rovna koeficientu příbuznosti x heritabilita

Závěr

Fenotypová podobnost mezi příbuznými jedinci závisí na stupni příbuznosti (podobné geny) a heritabilitě

$$r_{P_X P_Y} = r_{XY} h^2$$

Genetická podobnost mezi příbuznými jedinci závisí na stupni příbuznosti: kovariance mezi PH:

$$\text{COV}_{A_X A_Y} = r_{XY} V_A$$

Způsoby odhadování heritability

Odpověď na selekci:

$$\hat{h}^2 = \frac{\Delta G}{d}$$

Regrese potomek – rodič:

$$\hat{h}^2 = 2 b_{P-R}$$

Regrese potomek – střed rodičů:

$$\hat{h}^2 = b_{P-OM}$$

Korelace polosourozenců:

$$\hat{h}^2 = 4 \hat{\rho}$$

Korelace vlastních sourozenců:

$$\hat{h}^2 = 2 \hat{\rho}$$

VCE s AM:

$$\hat{h}^2 = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_P^2}$$

Metody odhadu heritability na základě:

- podobnost rodičů a potomků
- rozklad proměnlivosti
- neparametrické metody
- selekční experimenty

1. podobnost rodičů a potomků

- regrese potomka na rodiče: $h^2 = 2 \cdot b_{xy}$
- regrese potomka na průměr rodičů: $h^2 = b_{xy}$
- regrese průměru potomků na průměr rodičů: $h^2 = b_{xy}$
- korelace mezi sourozenci: $h^2 = 2 \cdot r_{xy}$
- korelace mezi polosourozenci: $h^2 = 4 \cdot r_{xy}$

Regrese potomek – rodič

$$\hat{h}^2 = 2 \cdot b_{p,r}$$

Model statistický: $Z_i = \beta X_i + e_i$

Z_i – střední hodnota potomka i-tého otce

β - regrese Z na X

X_i – pozorovaná hodnota u i-tého otce

e_i – náhodná chyba spojená s Z_i

Př.: Ve velké neinbrední populaci slepic byla zjišťována ve věku osmi týdnů živá hmotnost kohoutů s přesností na jeden gram. V pohlavní dospělosti bylo 17 kohoutků (otcové) náhodně pářeno se slepicemi a u jejich potomků samčího pohlaví byla opět zjišťována živá hmotnost ve stejném věku. Zjišťovaný genetický parametr, koeficient dědivosti, se bude vztahovat na populaci otců.

Výpočet vstupních hodnot regresního stavu

$$SS_{xz} = \sum_i x_i \cdot z_i - \frac{X \cdot Z}{n}$$

$$SS_x = \sum_i x_i^2 - \frac{X^2}{n}$$

$$SS_z = \sum_i z_i^2 - \frac{Z^2}{n}$$

$$b_{zx} = \frac{\sum_i x_i \cdot z_i - \frac{X \cdot Z}{n}}{\sum_i x_i^2 - \frac{X^2}{n}}$$

n	x_i	z_i	x_i^2	z_i^2	$x_i \cdot z_i$
1	601	910	361201	828100	546910
2	733	983	537289	966289	720539
3	793	976	628849	952576	773968
4	795	1050	632025	1102500	834750
5	818	1080	669124	1166400	883440
6	838	1040	702244	1081600	871520
7	854	1040	729316	1081600	888160
8	880	1025	774400	1050625	902000
9	882	994	777924	988036	876708
10	895	1030	801025	1060900	921850
11	952	1021	906304	1042441	971992
12	953	1078	908209	1162084	1027334
13	961	964	923521	929296	926404
14	979	976	958441	952576	955504
15	995	1110	990025	1232100	1104450
16	997	1041	994009	1083681	1037877
17	1040	1035	1081600	1071225	1076400
Σ	14966	17353	13375506	17752029	15319806

$$\Sigma x_i = X, \quad \Sigma z_i = Z.$$

$$h^2 = 2 \cdot b_{zx}$$

Následuje výsledek této regresní analýzy (variance a kovariance), který byl získán v programu SAS.

The REG Procedure

Dependent Variable: y
Analysis of Variance

Source	DF	Squares	Sum of Square	Mean F Value	Pr > F
Model	1	9256.15665	9256.15665	4.72	0.0463
Error	15	29443	1962.86015		
Corrected Total	16	38699			

Root MSE	44.30418	R-Square	0.2392
Dependent Mean	1020.76471	Adj R-Sq	0.1885
Coeff Var	4.34029		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	831.44274	87.84239	9.47	<.0001
x	1	0.21505	0.09903	2.17	0.0463

Střední chyba odhadu

$$se_{b_{zx}} = \sqrt{\frac{s_{b_{zx}}^2}{SS_x}} = \sqrt{\frac{SS_z - \frac{(SS_{xz})^2}{SS_x}}{SS_x \cdot (n-2)}}$$

$$se_{h^2} = 2 \cdot se_{b_{zx}}$$

• U některých druhů hospodářských zvířat může být otec připarován na několik matek a každá matka může mít několik potomků. Vliv otců může být odstraněn na základě výpočtu **regrese potomků na matku uvnitř otců** podle modelu:

- Model: $Z_{ij} = \mu + \alpha_i + \beta(x_{ij} - x_{..}) + e_i$
 - Z_{ij} – střední hodnota potomků z páření i-tého otce a j-té matky
 - α_i – vliv i-tého otce
 - μ – obecný střed populace
 - β – regrese Z na X
 - X_{ij} – pozorovaná hodnota j-té matky pářené s i-tým otcem
 - $x_{..}$ – fenotypový průměr
 - e_i – náhodná chyba spojená s Z_{ij} (odchylka od Z_{ij})

Tabulka analýzy variance a kovariance:

Zdroj proměnlivosti	Součet čtverců odchylek		Stupeň volnosti	Průměrný čtverec		Součet produktů xz	Střední produkt xz
	SS_x	SS_z	df	MS_x	MS_z	SP	MP
Mezi otci	SS_{sx}	SS_{sz}	$df_s = s-1$	<i>tato data nejsou třeba k výpočtům</i>			
Mezi matkami uvnitř otců	SS_{dx}	SS_{dz}	$df_d = d - s$	MS_{dx}	MS_{dz}	SP_{dxz}	MP_{dxz}
Celková	SS_{cx}	SS_{cz}	$df_c = d - 1$	-	-		

s – počet otců (sire), d – počet matek (dam), S – součet čtverců, MS – střední čtverec, Sp – součet produktů, MP – střední produkt (variance)

$$cov_{d_{ZX}} = \frac{SP_{d_{ZX}}}{df_d} = MP_{d_{ZX}}$$

Odhad regrese a dědivosti mezi matkami a dcerami uvnitř otců:

$$b_{zx} = \frac{cov_{d_{ZX}}}{var_{d_x}} = \frac{MP_{d_{ZX}}}{\sigma_{d_x}^2} = \frac{\frac{SP_{d_{ZX}}}{d-s}}{\frac{S_{d_{ZX}}}{d-s}} = \frac{SP_{d_{ZX}}}{S_{d_x}} \quad h^2 = 2 \cdot b_{zx}$$

Odhad střední chyby koeficientu regrese a dědivosti:

$$s_{b_{zx}}^2 = \frac{S_{d_z} - \frac{(SP_{d_{ZX}})^2}{S_{d_x}}}{d-s-1} \quad se_{b_{zx}} = \sqrt{\frac{s_{b_{zx}}^2}{S_{d_x}}} \quad se_{h^2} = 2 \cdot se_{b_{zx}}$$

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

Otcové	Matky x_{ij}	Potomci z_{ij}	$X_{i\cdot}$	$Z_{i\cdot}$	n_i	X_{ij}^2	z_{ij}^2	$x_{ij} \cdot z_{ij}$
1	754	808				568516	652864	609232
1	648	700	2283	2228	3	419904	490000	453600
1	881	720				776161	518400	634320
2	740	725				547600	525625	536500
2	712	840	2264	2365	3	506944	705600	598080
2	812	800				659344	640000	649600
3	765	780				585225	608400	596700
3	807	840	1572	1620	2	651249	705600	677880
4	969	850				938961	722500	823650
4	849	802	2550	2482	3	720801	643204	680898
4	732	830				535824	688900	607560
5	740	806	1481	1641	2	547600	649636	596440
5	741	835				549081	697225	618735
6	831	830				690561	688900	689730
6	639	800	2203	2134	3	408321	640000	511200
6	733	504				537289	254016	369432
Σ	12353	12470			16	9643381	9830870	9653557

mezi matkami
uvnitř otců

SS pro X

$$SS_{d_x} = \sum_i \sum_j x_{ij}^2 - \sum_i \frac{X_{i\cdot}^2}{n_i}$$

SS pro Y

$$SS_{d_z} = \sum_i \sum_j z_{ij}^2 - \sum_i \frac{Z_{i\cdot}^2}{n_i}$$

Výpočet součtu
produktů
(součinů)
vlastností X a Z

$$SP_{d_{zx}} = \sum_i \sum_j x_{ij} \cdot z_{ij} - \sum_i \frac{X_{i\cdot} \cdot Z_{i\cdot}}{n_i}$$

$$cov_{d_{zx}} = MP_{d_{zx}} = \frac{SP_{d_{zx}}}{d - s}$$

$$b_{zx} = \frac{SP_{d_{zx}}}{SS_{d_x}}$$

$$h^2 = 2 \cdot b_{zx}$$

Následuje výsledek této analýzy variance a kovariance, který byl získán v programu SAS

The GLM Procedure

Dependent Variable: P

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	31134.0501	5189.0083	0.58	0.7409
Error	9	80929.6999	8992.1889		
Corrected Total	15	112063.7500			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
O	5	30330.58333	6066.11667	0.67	0.6534
M	1	803.46673	803.46673	0.09	0.7718

Parameter	Estimate	Std. Error	t Value	Pr > t
Intercept	637.7152160	252.2943565	2.53	0.0324
O 1	28.6599564	77.9408706	0.37	0.7216
O 2	74.9615501	77.7257832	0.96	0.3600
O 3	93.4869989	88.2822372	1.06	0.3172
O 4	104.4042276	86.6005504	1.21	0.2587
O 5	108.5484483	86.5896538	1.25	0.2416
O 6	0.0000000	.	.	.
M	0.1002516	0.3353824	0.30	0.7718

P - užítkovost potomků; M - užítkovost matek; O - otcové

2. Rozklad proměnlivosti

- z analýzy variance
- teoreticky vychází z předpokladu podobnosti příbuzných
- Čím si jsou příbuzní podobnější náhodní jedinci v populaci, tím pro danou vlastnost odhadujeme vyšší hodnotu koeficientu heritability

Využití užitečnosti rodin k odhadu heritability

• Vysoká heritabilita

– **nízká** variance v rodinách

– **vysoká** variance mezi rodinami

• Nízká heritabilita

– **vysoká** variance v rodinách

– **nízká** variance mezi rodinami

	Typ podobnosti rodiny								
	Žádná			Střední			Vysoká		
Otec	1	2	3	1	2	3	1	2	3
Hodnota jedinců	1	1	1	3	2	1	1	2	3
	2	2	2	2	1	3	1	2	3
	3	3	3	1	2	3	1	2	3
Var mezi rodinami	Nízká			Střední			Velká		
Var v rodinách	Velká			Střední			Nízká		

h^2

nízká

střední

vysoká

Celková var = var mezi rodinami + var v rodinách

Stená jako *cov* v rodinách

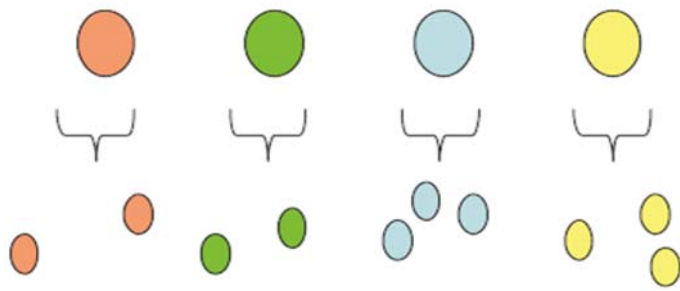
Odhad dědivosti

Nutná variance mezi a v rodinách

Více rodin > větší přesnost odhadu

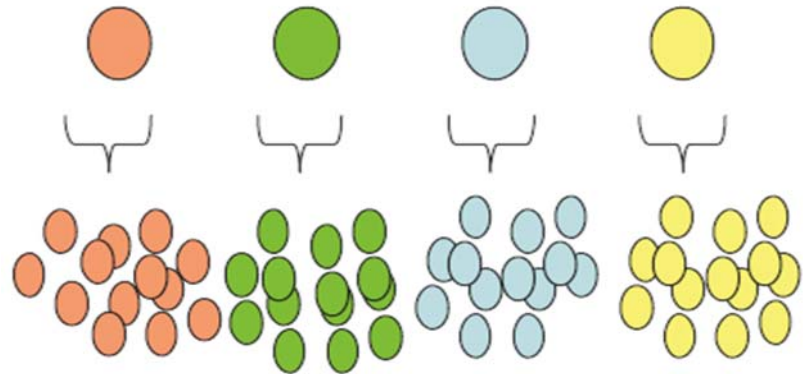
Variance mezi rodinami = kovariance v rodinách

Vysoká variance mezi rodinami (jsou odlišné) = vysoká kovariance v rodinách (členové v rodinách jsou podobní)



Špatný odhad

Dobrý odhad



Odhad h^2 u rodin polosourozenců

• **Variance mezi rodinami polosourozenců**

= kovariance mezi polosourozenci = $\frac{1}{4} V_A$

• **Variance v rodinách polosourozenců**

= reziduální varianci = $V_P - \frac{1}{4} V_A = \frac{3}{4} V_A + V_E + V_D$

$$h^2 = \frac{V_A}{V_P} = \frac{4V_S}{V_P}$$

Odhad h^2 u rodin vlastních sourozenců

- Komplikovanější odhad > společné prostředí > V_{EC} (efekt maternálního prostředí) a V_D (efekt dominance)
- **Variance mezi rodinami úplných sourozenců**
= kovariance mezi úplnými sourozenci
= $\frac{1}{2} V_A + V_{EC} + \frac{1}{4} V_D$
- **Variance v rodinách úplných sourozenců**
= reziduální varianci = $V_P - (\frac{1}{2} V_A + V_{EC} + \frac{1}{4} V_D)$
= $\frac{1}{2} V_A + V_E + \frac{3}{4} V_D$

Často bývá h^2 přeodhadnuto, protože je obtížné oddělit genetické a ostatní efekty

Metody odhadu genetických parametrů

Analýza variance (ANOVA) – **nejmenší čtverce**

Balancovaná data

Nebalancovaná data

Hendersonovy metody (SAS, Harvey, SPSS, ...)

Věrohodnostní metody – **maximální věrohodnost**

Maximální věrohodnost (ML)

Restringovaná maximální věrohodnost (REML)

Bayesian Methods

Monte Carlo, Gibbs Sampling

Komponenty variance	Odhadované parametry
V_A	Heritabilita
V_E	
V_{EP}	Opakovatelnost
kovariance	Korelace