

Genetika kvantitativních znaků - principy, vlastnosti a aplikace statistiky

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

Genetika kvantitativních vlastností vychází z

Mendelistická genetika

- dědičnost (přenos GI) vlastností s diskrétními hodnotami fenotypu
- individuální efekt genu

Genetika populací kvalitativních znaků

- Model genetiky populací – 1 gen a 2 alely
- změny v alelových frekvencích genů u vlastností s diskrétními hodnotami
- nutné určit, kterou alelu jedinec nese

Mnoho vlastností však nemá diskrétní hodnoty, nýbrž kontinuální variabilitu, ⇒

GENETIKA POPULACÍ KVANTITATIVNÍCH VLASTNOSTÍ
KVANTITATIVNÍ GENETIKA

Základní typy vlastností

Diskontinuální / diskrétní (nespojité) vlastnosti

- výrazně definované fenotypové kategorie
- vlastnosti určované 1 nebo několika geny (**major gen**), bez vlivu prostředí → Mendelistický polymorfismus → většina fenotypové variability

KVALITATIVNÍ VLASTNOST

Kontinuální / spojité / komplexní vlastnosti

- nevýrazně definované kategorie fenotypů
- rozdělení fenotypů vykazuje více nebo méně **kontinuální variabilitu** (lze zjistit jen rozmezí hodnot pro vlastnost)
- vlastnosti jsou určovány geny na mnoha lokusech → **polygenní dědičnost**
- ***geny kvantitativních vlastností (QTL) ≈ polygeny***
- většina genů má malý, ale aditivní účinek (**minor geny**)
- vlivy **prostředí** na projev a distribuci fenotypů
- **každý gen je děděn mendelisticky, ale nelze vidět jeho efekt**

⇒ KVANTITATIVNÍ VLASTNOST

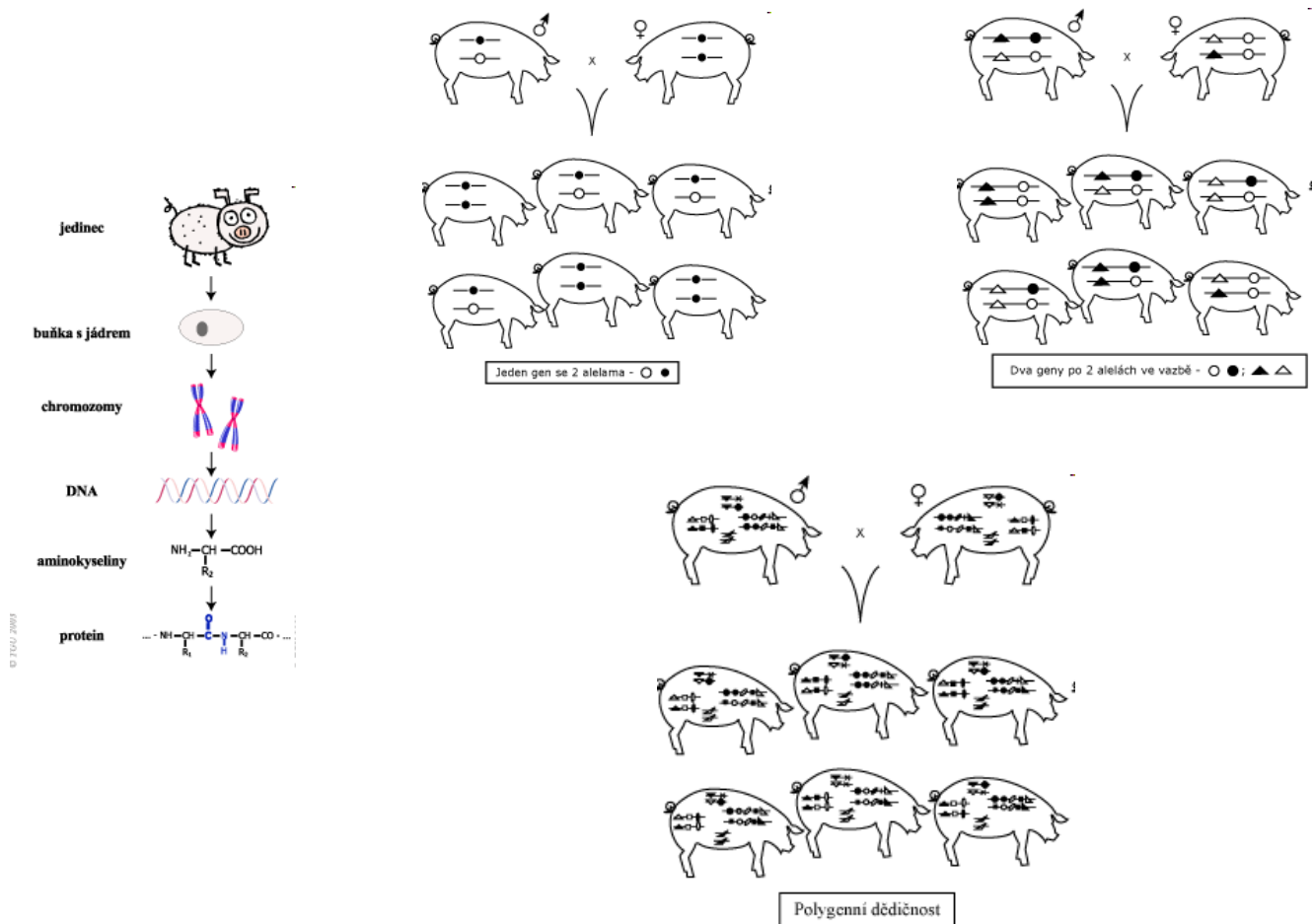
Kvalitativní vlastnosti

- diskontinuální, nespojitá (diskrétní) variabilita
- podmíněna 1 nebo několika málo geny
- monogenní (popř. oligenní) dědičnost
- mendelistické poměry v F_1
- lze určit fenotypovou hodnotu každého genotypu
- vlastnosti jsou hodnoceny podle kvality projevu (rohato - bezrohato, červený - bílý květ, ...)
- geny s interakčními účinky (dominance, epistáze)
- na projev vlastnosti nemá vliv *prostředí*
- lze detekovat efekt jednotlivých genů podílejících se na vlastnosti

Kvantitativní vlastnosti

- **kontinuální**, spojitá variabilita
- podmíněna mnoha geny na více lokusech
- **polygenní dědičnost**
- poměry v F_1 nejsou mendelistické
- rozdělení fenotypů vykazuje více nebo méně kontinuální variabilitu (lze určit rozmezí hodnot)
- vlastnosti jsou kvantifikovány měřením, vážením, počítáním, ...
- vlastnosti jsou determinovány geny velkého účinku (nepřispívají kvantitativně) a větším počtem genů malého účinku (polygeny), většina genů má **aditivní** účinek
- projev vlastnosti modifikuje **vliv prostředí**
- nelze rozpoznat účinek jednotlivých genů podílejících se na vlastnosti

Přesto se všechny geny se dědí jako informační jednotky, t.j. u diploidních organismů je každý gen v buňce přítomen 2x (jeden od otce a druhý od matky), bez ohledu determinují-li vlastnost kvalitativní nebo kvantitativní (rozdíl lze pozorovat v jejich fenotypovém projevu + specifické odchylky jako např. imprinting genů!)



Hodnocení genetické variability u kvantitativních vlastností

Klasický přístup – jsou změřeny fenotypy jedinců se známými příbuzenskými vztahy a genetické a prostředové zdroje fenotypové variance jsou určeny statistickými metodami (ANOVA, ML,...)

Příbuzní jedinci sdílejí určitý podíl společných genů (to je kvantifikováno koeficientem příbuznosti r_{ij})

Molekulárně statistický přístup – nejnovější technika mapování lokusů kvantitativních vlastností (**QTL** – quantitative trait loci) pomocí markerů a asociování s fenotypovými hodnotami

Hledají se sekvence DNA pomocí genetických markerů (SNP) po celém genomu a jejich statistické asociace k fenotypové variabilitě, popř. expresi (opět má slovo statistika!)

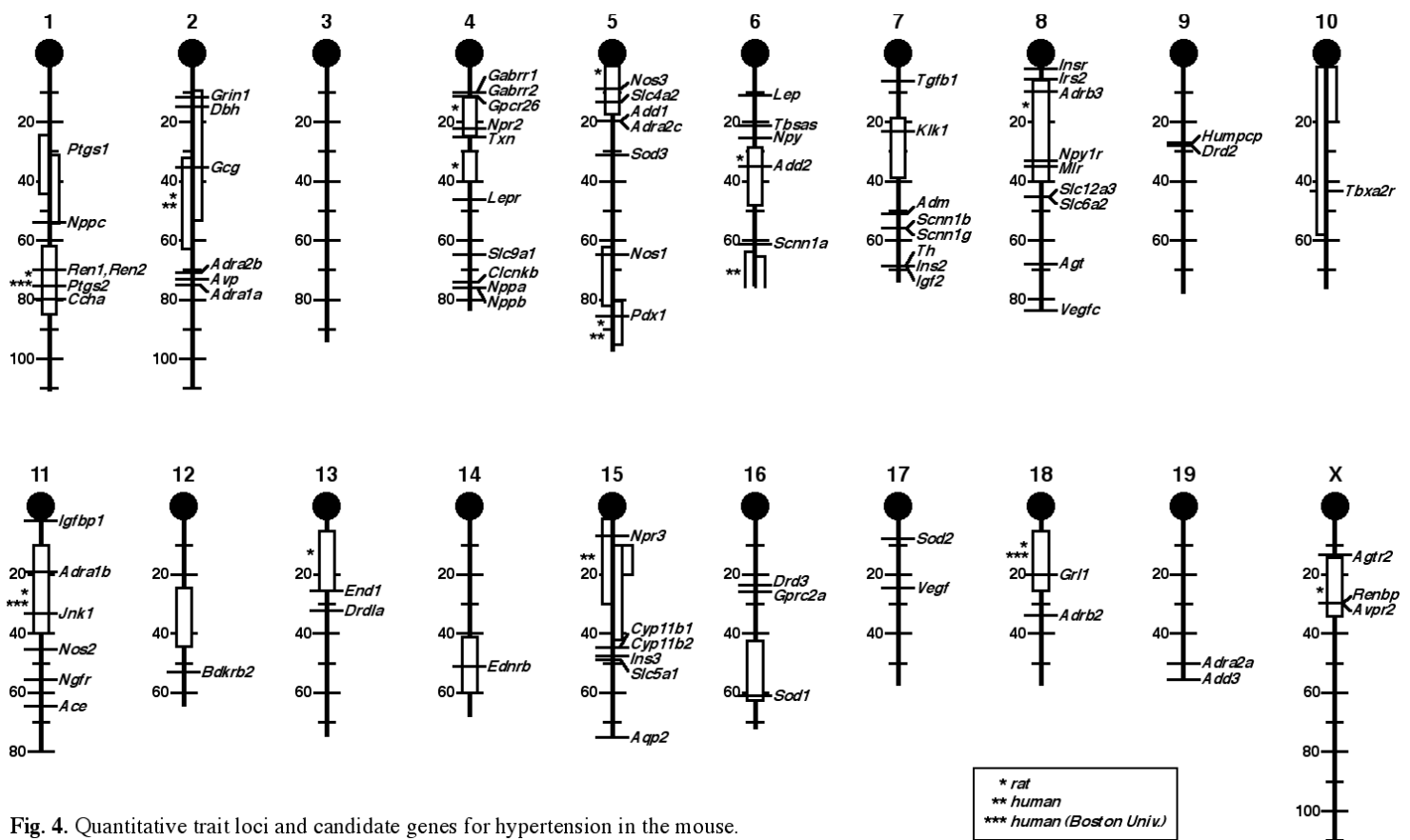
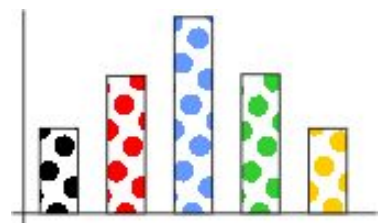
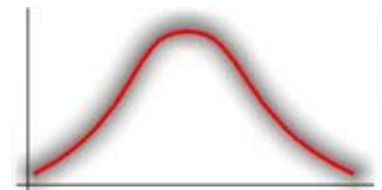


Fig. 4. Quantitative trait loci and candidate genes for hypertension in the mouse.

Mouse Models of Blood Pressure Regulation and Hypertension: Fumihiko Sugiyama, Ken-ichi Yagami, and Beverly Paigen ; Current Hypertension Reports (2001). 3:41-48

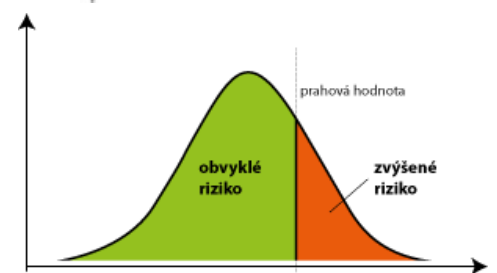
Příklady kvantitativních vlastností

Vlastnosti s plynulou kontinuální proměnlivostí - výnos obilí, rezistence k nemocem u rostlin i živočichů, přírůstek hmotnosti, výška v kohoutku, obsah tuku v mase, IQ, schopnost naučit se, krevní tlak, ...



Vlastnosti meristické - počet selat ve vrhu, počet zrn v klase

Vlastnosti prahové (threshold) – projev nemoci (schizofrenie, cukrovka), výskyt dvojčat



Základní otázky

Základní zásada QG: viditelná variabilita je způsobena kombinací působení mnoha genů a faktorů prostředí

Nebo: fenotyp = „genotyp“ + prostředí

Základní otázky (a odpovědi?):

1. Jaká je genetická podstata kvantitativních vlastností? (jsou podmíněny normálními geny podléhající Mendelovým pravidlům).
2. Jak můžeme odlišit efekty genů od efektů prostředí? (např. inbreedingem lze eliminovat genetickou varianci).
3. Jak můžeme předpovědět a kontrolovat výsledky z křížení? (např. umělou selekcí).

Kvantitativní geny jsou mendelovské geny

V r. **1909** Herman Nilsson-Ehle (Švédsko) provedl řadu experimentů s pšenicí - barva zrna.

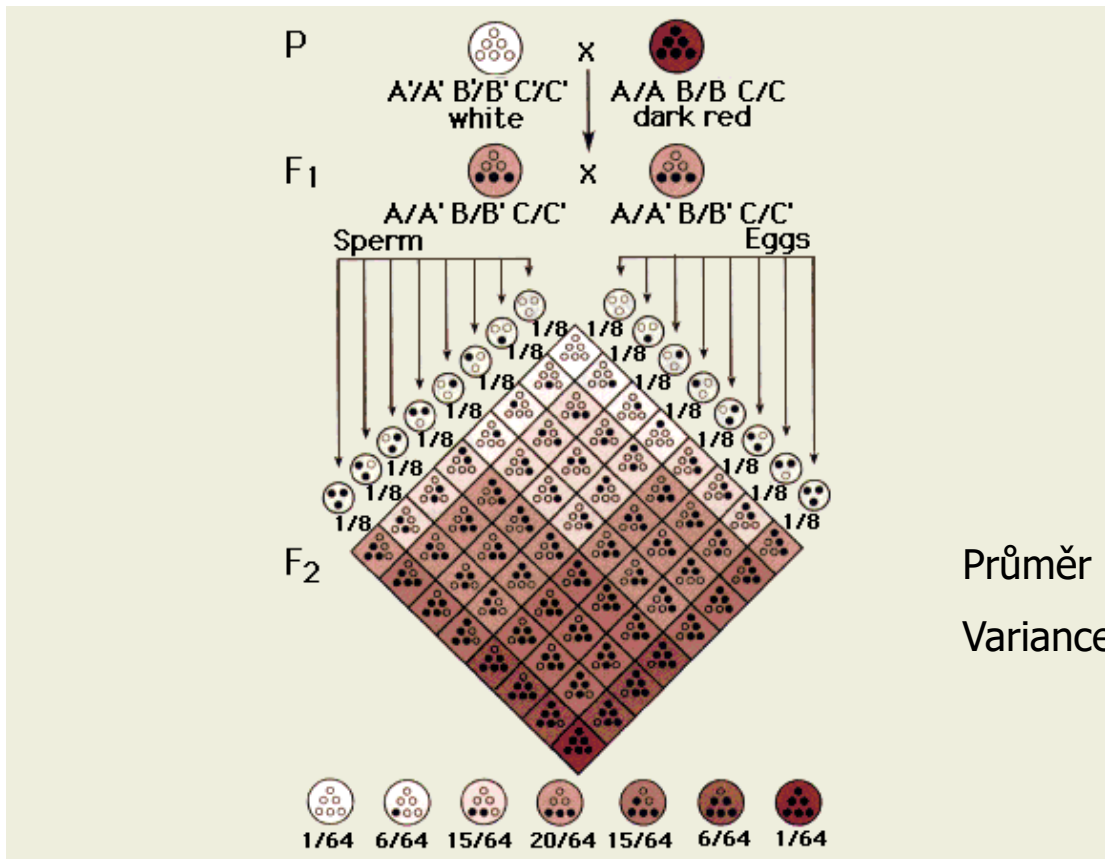
Současná pšenice je hexaploidní, výsledek 3 různých druhů tvořící stabilního hybridu, allopolyploid. > 3 podobné, ale trochu odlišné genomy pšenice, nazvané A, B a D.

Každý genom má jeden gen, který ovlivňuje barvu zrna, a každý tento lokus má alelu pro červenou barvu a alelu pro bílou barvu.

- *červené* alely: A, B a D,
- *bílé* alely a, b a d.

Forma dědičnosti těchto alel je „aditivita“. Množství červeného pigmentu v zrně je úměrné přítomnému počtu červených alel, od 0 do 6.

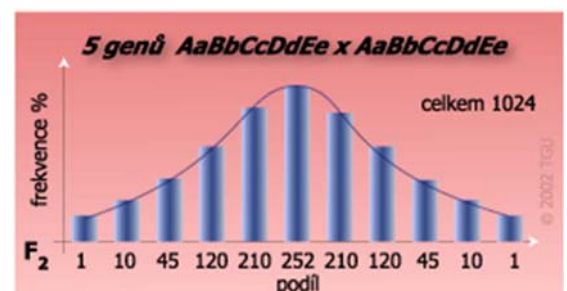
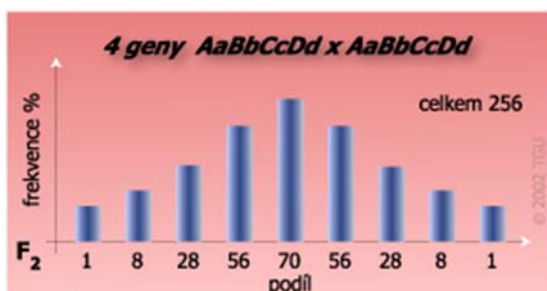
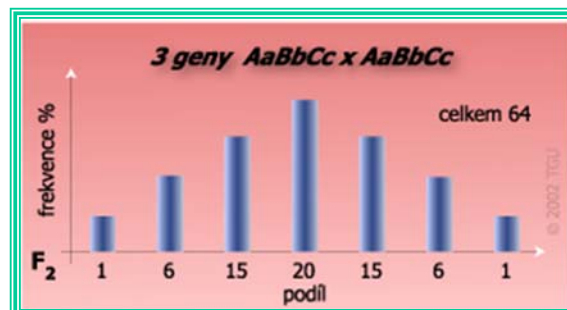
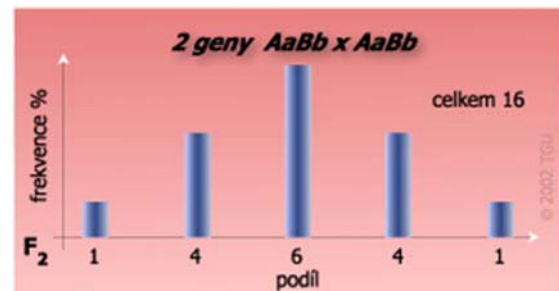
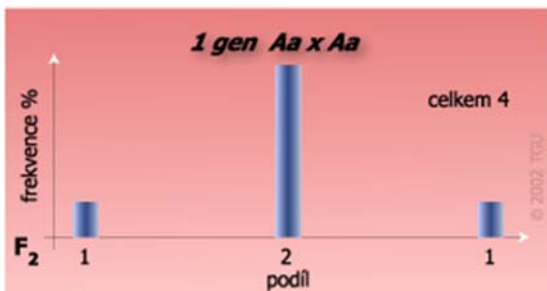
Segregace a volná kombinovatelnost tří genů a aditivním působením



Průměr = 3
 Variance = 1,5

Počet aditivních alel: 0 1 2 3 4 5 6

Přechod od kvality ke kvantitě



Př.: 2 geny, A & B kontrolují vlastnost

Alela A = 4 jednotky

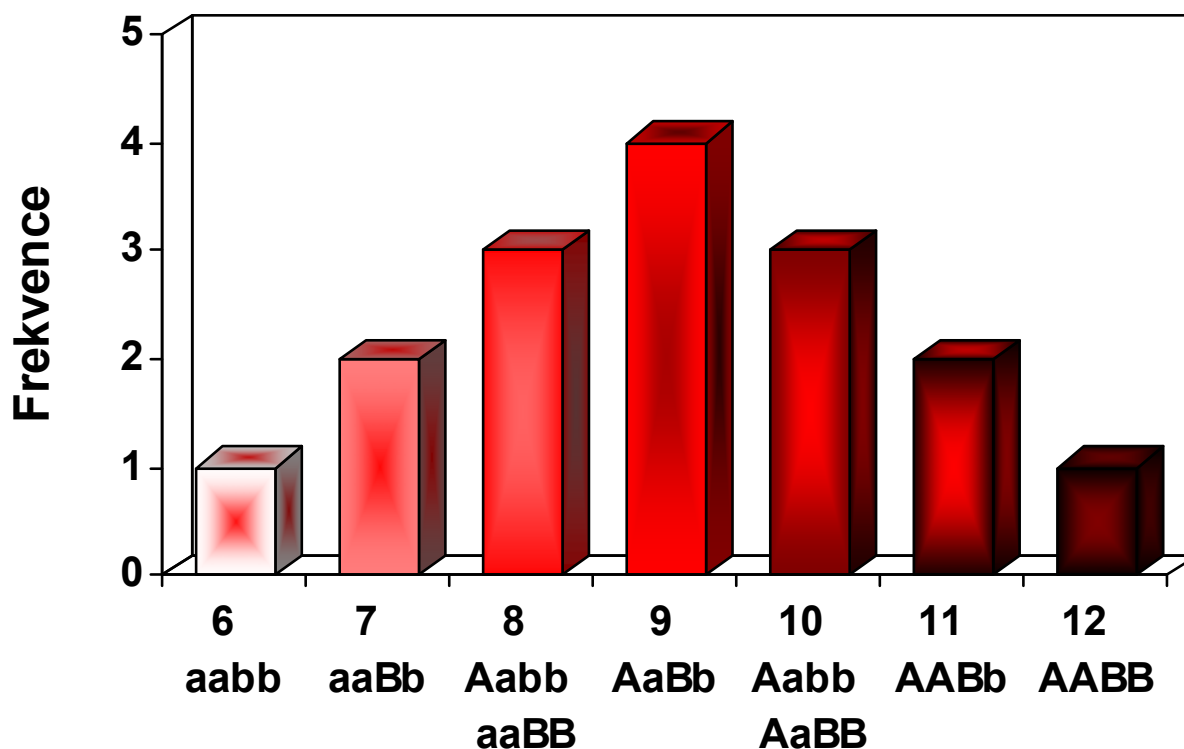
B = 2 jednotky

a = 2 jednotky

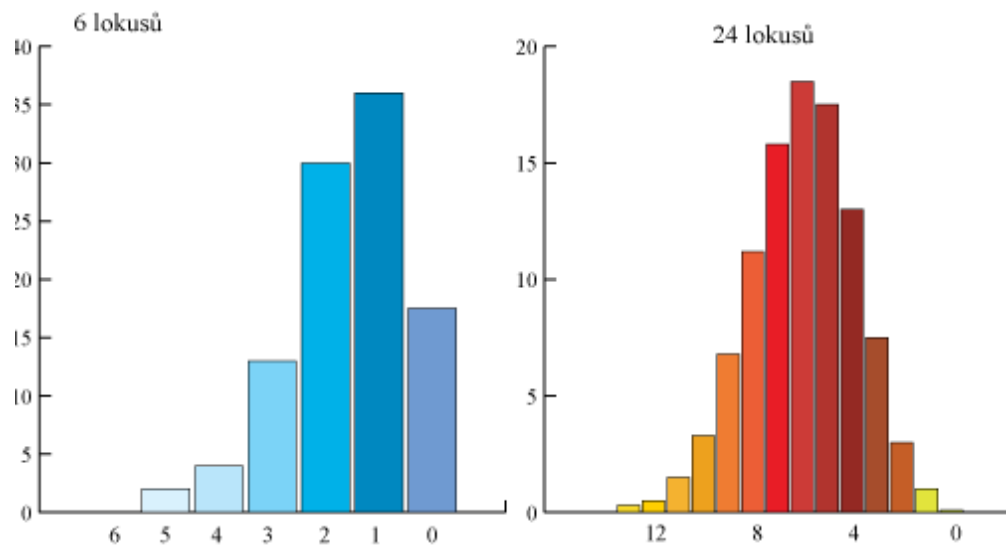
b = 1 jednotka

Genotyp	G. poměr v F ₂	Metrická hodnota
<i>AABB</i>	1	12
<i>AABb</i>	2	11
<i>AAbb</i>	1	10
<i>AaBB</i>	2	10
<i>AaBb</i>	4	9
<i>Aabb</i>	2	8
<i>aaBB</i>	1	8
<i>aaBb</i>	2	7
<i>aabb</i>	1	6

Grafické znázornění příkladu



Očekávaná distribuce za současné segregace dvou alel na každém lokusu:



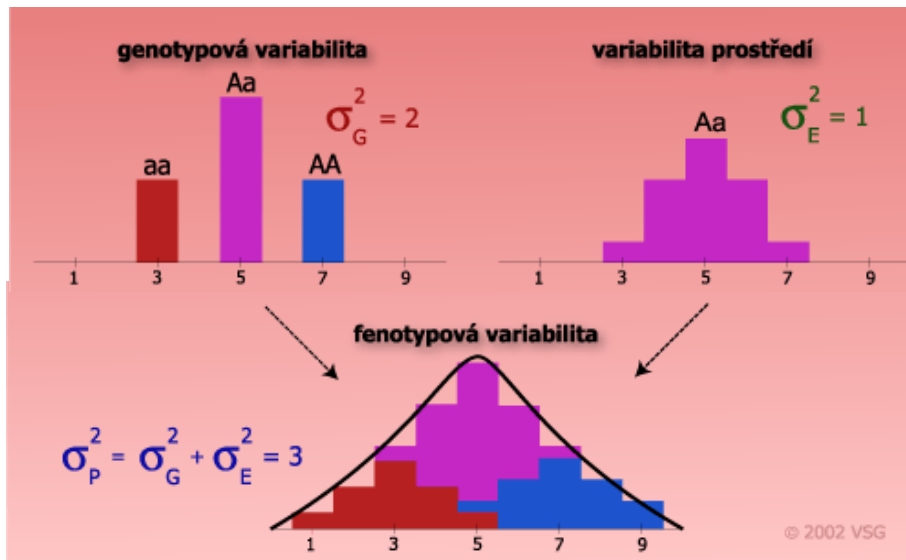
Vlivy prostředí na kvantitativní vlastnosti

Rok	Výnos genotypů (bu/akr)		
	Roughrider	Seward	Agassiz
1986	47,9	55,9	47,5
1987	63,8	72,5	59,5
1988	23,1	25,7	28,4
1989	61,6	66,5	60,5
1990	0	0	0
1991	60,3	71,0	55,4
1992	46,6	49,0	41,5
1993	58,2	62,9	48,8
1994	41,7	53,2	39,8
1995	53,1	65,1	53,5

Poznámka ...

Všechny rostliny v r. 1990 nepřežimovaly.

Modifikující vliv prostředí



Důležitý koncept v kvantitativní genetice

Fenotyp = genetické faktory + faktory prostředí

$$P = G + E$$

Variabilita → variance (V_P)

Pro pochopení heritability je třeba znát více kvantitativní variabilitu.

Co to je, přesně, co pozorujeme, když studujeme kvantitativní vlastnost?

Variabilitu mezi fenotypy!

-> **fenotypová variance** (V_P) je základní informace používaná v kvantitativní genetice

Fenotyp vyplývá z genotypu interagujícího s prostředím

Lze říci, že V_P je způsobena variancí genů (genotypů) a variancí prostředí:

$$V_P = V_G + V_E$$

Studované otázky QG

Jaký je podíl genetických faktorů a prostředí na fenotypu?

Kolika geny je ovlivněna vlastnost?

Je příspěvek genů stejný?

Jak alely na různých lokusech interagují: aditivně? epistaticky?

Jak rychle se změní vlastnost pod selekcí?

Kontinuální variabilita - normální distribuce

- Problémy kontinuálních vlastností lze rozdělit do dvou typů:
 - většina je ovlivněna alelami na dvou a více lokusech a segregace jakéhokoliv genu v rodokmenu je zastřena segregací ostatních genů ovlivňujících vlastnost,
 - většina kontinuálních vlastností jsou ovlivněny faktory prostředí stejně jako geny a genetická segregace je zastřena prostředovými efekty.
- Pokud se u mnoha kontinuálních vlastností fenotypy seskupí do vhodných intervalů a zobrazen pomocí pruhového grafu, pak bude distribuce fenotypů vyhovovat **normálnímu rozložení**, které je symetrické, s křivkou zvonovitého tvaru. Hladká křivka je normálním rozložením, které nejlépe odpovídá datům.

– Rovnice normální křivky

(funkci hustoty pravděpodobnosti) je:

$$f_{(x)} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Popis kvantitativních vlastností

Tyto vlastnosti nemohou být analyzovány tradičními technikami mendelistické genetiky (~ genetikou populací kvalitativních znaků)

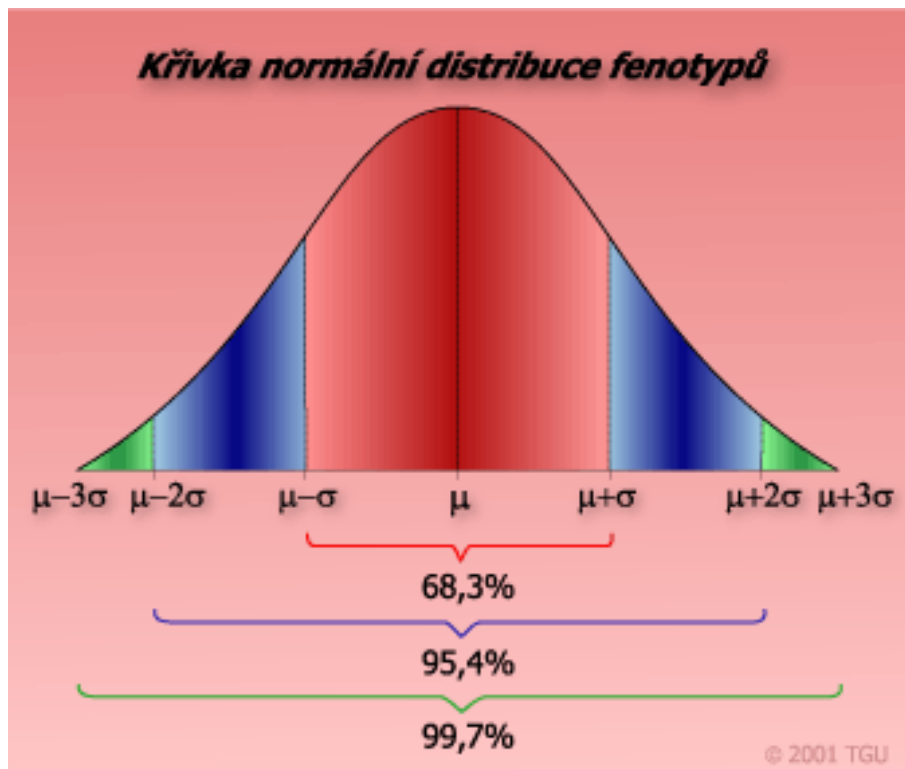
Statistické analýzy

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x_i}{n}$$

$$s_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{\sum (x_i - \bar{x})^2}{n} = V$$

$$s_x = \sqrt{s_x^2}$$

Graf distribuce fenotypové proměnlivosti



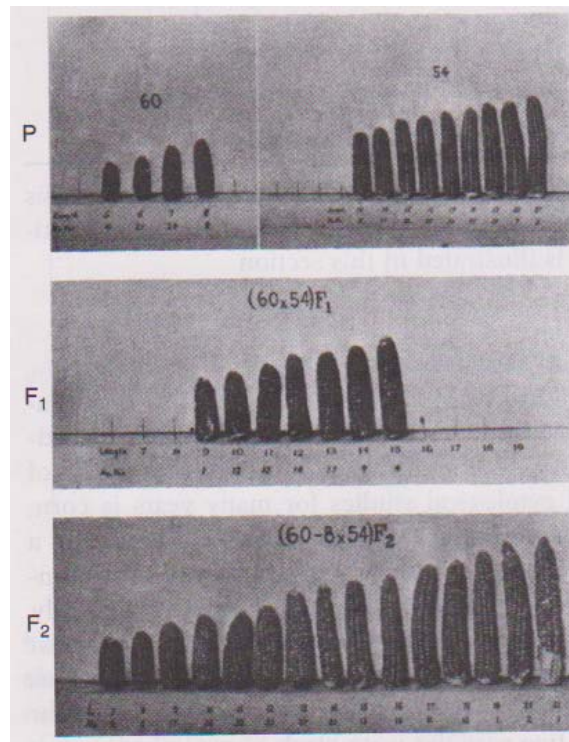
Variance je změřená
variabilita distribuce

$\pm 1 \sigma$ zahrnuje 68,3 %

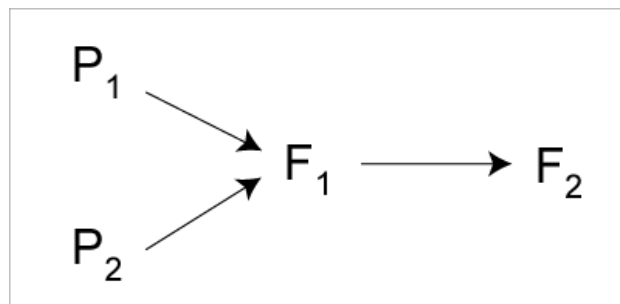
$\pm 2 \sigma$ zahrnuje 95,5 %

$\pm 3 \sigma$ zahrnuje 99,7 %

Příklad: Délka ouška u kukuřice a Mendelův design pokusu



Generace	Průměr (cm)	s_x (cm)
P_1	16,80	0,817
P_2	6,63	1,887
F_1	12,12	1,519
F_2	12,89	2,252

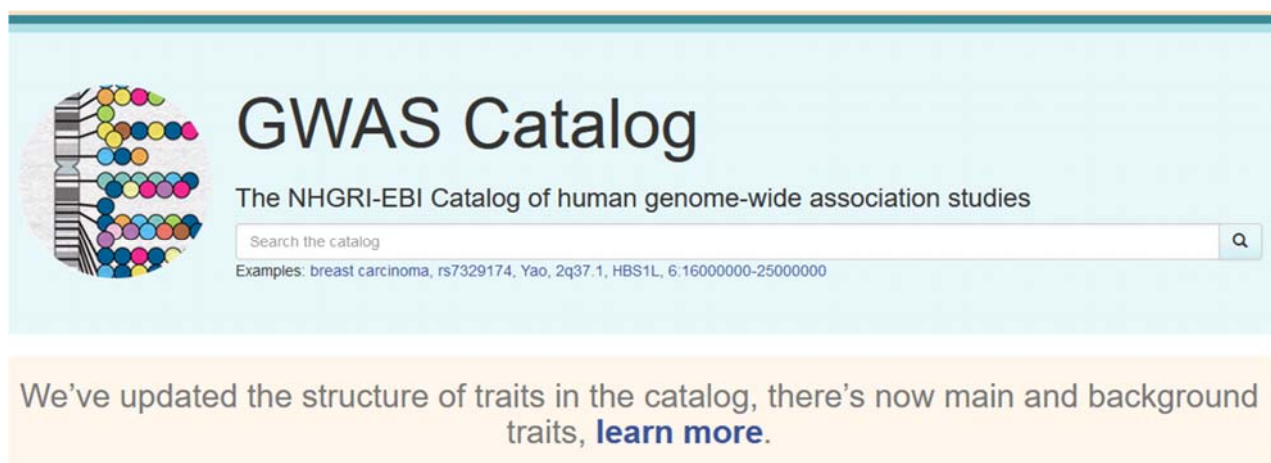


Generace	Průměr (cm)	s_x (cm)
P_1	16,80	0,817
P_2	6,63	1,887
F_1	12,12	1,519
F_2	12,89	2,252

•závěry:

- průměrná délka ouška linie P_2 je menší, s_x je větší \Rightarrow je zde větší variabilita než v linii s delšími oušky P_1
- variabilita v populaci F_1 je v důsledku variability prostředí
- průměr vlastnosti v F_1 populaci je střední mezi oběma rodiči a průměr F_2 populace je přibližně stejný jako v F_1
- F_2 populace má větší variabilitu než F_1
- extrémní hodnoty distribuce by měly být rovny hodnotám rodičů, protože část populace bude mít jejich genotyp

<http://www.ebi.ac.uk/gwas/home>



GWAS Catalog
The NHGRI-EBI Catalog of human genome-wide association studies

Search the catalog

Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6.16000000-25000000

We've updated the structure of traits in the catalog, there's now main and background traits, [learn more](#).

Download

Download a full copy of the GWAS Catalog in spreadsheet format as well as current and older versions of the GWAS diagram in SVG format.

Summary statistics

Documentation and access to full summary statistics for GWAS Catalog studies where available.

Submit

Submit summary statistics to GWAS Catalog.

Documentation

Including FAQs, our curation process, training materials, related resources, a list of abbreviations and API documentation.

Diagram

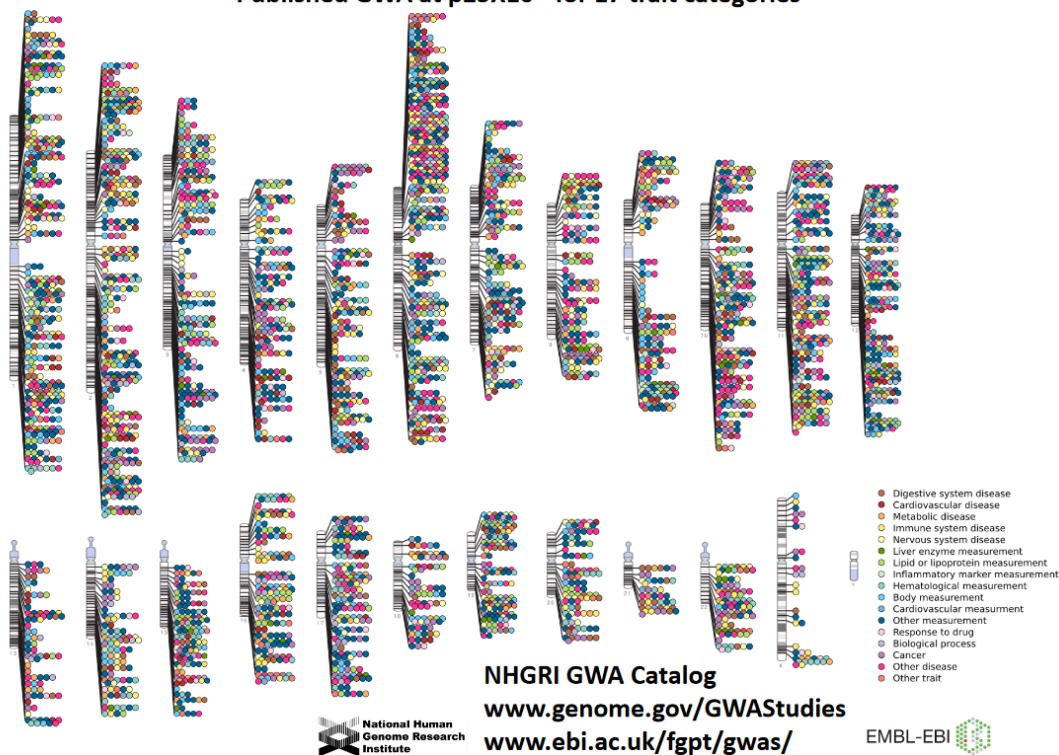
Explore an interactive visualisation of all SNP-trait associations with genome-wide significance ($p \leq 5 \times 10^{-8}$).

Ancestry

An introduction to our ancestry curation process.

Published Genome-Wide Associations through 12/2013

Published GWA at $p \leq 5 \times 10^{-8}$ for 17 trait categories



NHGRI GWA Catalog

www.genome.gov/GWASudies

www.ebi.ac.uk/fgpt/gwas/



2017

SNPs associated with trait 'smoking behavior' in region 6q14.1

SNP	p-Value	EFO mapping	GWAS trait	Study
rs12207675	2×10^{-10}	smoking behavior	Waist circumference adjusted for BMI (joint analysis main effects and smoking interaction)	Justice AE et al., 2017
rs12207675	7×10^{-10}	smoking behavior	Waist circumference adjusted for BMI (adjusted for smoking behaviour)	Justice AE et al., 2017



Show SNPs for 2018

Show SNPs for

- Digestive system disease 714
- Cardiovascular disease 412
- Metabolic disease 173
- Immune system disease 756
- Nervous system disease 928
- Liver enzyme measurement 67
- Lipid or lipoprotein measurement 416
- Inflammatory marker measurement 218
- Hematological measurement 2109
- Body weights and measures 976
- Cardiovascular measurement 546
- Other measurement 3214
- Response to drug 223
- Biological process 333
- Cancer 726
- Other disease 551
- Other trait 2538

- Digestive system disease 968
- Cardiovascular disease 674
- Metabolic disease 226
- Immune system disease 1201
- Nervous system disease 1065
- Liver enzyme measurement 154
- Lipid or lipoprotein measurement 464
- Inflammatory marker measurement 326
- Hematological measurement 2249
- Body weights and measures 1158
- Cardiovascular measurement 679
- Other measurement 5044
- Response to drug 275
- Biological process 631
- Cancer 979
- Other disease 1160
- Other trait 3871

Year	Number of QTL
1994	10
1995	9
1996	1
1997	2
1998	99
1999	72
2000	129
2001	98
2002	232
2003	379
2004	343
2005	234
2006	164
2007	350
2008	887
2009	875
2010	5,680
2011	19,073
2012	4,675
2013	25,434
2014	8,142
2015	5,337
2016	25,894
2017	15,409
2018	6,824
2019	10,680
2020	43,461
2021	2,422

Chromosome	Number of QTL
1	4,174
2	4,650
3	4,430
4	5,456
5	8,098
6	23,469
7	4,094
8	2,155
9	2,364
10	3,410
11	4,678
12	2,125
13	3,636
14	18,958
15	2,199
16	2,405
17	5,992
18	3,185
19	4,246
20	5,485
21	4,179
22	1,712
23	2,015
24	1,254
25	3,597
26	14,545
27	1,775
28	1,167
29	5,657
30	8
X	25,797

Analýza variance - ANOVA

Příklad jednofaktorové analýzy variance hodnocení fenotypové proměnlivosti snášky u slepic jednoho chovu. Na základě jednofaktorové analýzy variance chceme zjistit, zda existují průkazné rozdíly ve snášce u potomstva po čtyřech kohoutech, chceme znát, zda „genetický“ vliv otec byl průkazný nebo neprůkazný.

Ve sledování jsou čtyři skupiny slepic a v každé je potomstvo po jednom ze čtyř kohoutů, jedná se tedy o skupiny polosourozenců.

	i = 1 (k ₁)		i = 2 (k ₂)		i = 3 (k ₃)		i = 4 (k ₄)	
	y_{ij}	y_{ij}^2	y_{ij}	y_{ij}^2	y_{ij}	y_{ij}^2	y_{ij}	y_{ij}^2
1	172	29584	137	18769	209	43681	172	29584
2	253	64009	197	38809	187	34969	166	27556
3	239	57121	200	40000	183	33489	121	14641
4	203	41209	198	39204	215	46225	115	13225
5	247	61009	131	17161	215	46225	134	17956
6	256	65536	145	21025	134	17956	134	17956
7					185	34225	127	16129
8					218	47524	115	13225
9					155	24025	138	19044
10					214	45796	116	13456
11					184	33856	124	15376
12					176	30976	197	38809
13							118	13924
14							129	16641
15							113	12769
16							138	19044
17							133	17689
18							116	13456
19							112	12544
20							114	12996
21							110	12100
22							124	15376
23							118	13924
24							138	19044
$\sum_{j=1}^{m_j}$	1370	318468	1008	174968	2275	438947	3122	416464
	$\sum_{j=1}^{m_j} y_{i\cdot} = Y_{i\cdot}$		$\sum y_{ij}^2$					

Model jednofaktorové analýzy variance: $y_{ij} = \mu + a_i + e_{ij}$

y_{ij} – užítkovost j-tého potomka po i-tém otci

μ – obecný průměr populace

a_i – vliv i-tého otce

e_{ij} – ostatní nahodilé vlivy

kohouti	$Y_{i\cdot}$	$y_{i\cdot}$	$Y_{i\cdot}^2$	$Y_{i\cdot}^2/n_i$	$\sum y_{ij}^2$	n_i
k1	1370	228,33	1876900	312816,67	318468	6
k2	1008	168,00	1016064	169344,00	174968	6
k3	2275	189,58	5175625	431302,08	438947	12
k4	3122	130,08	9746884	406120,17	416464	24
$Y_{..} = 7775$			1319582,92	1348847		$n = 48$
$Y^2_{..} = 60450625$			$\sum Y_{i\cdot}^2/n_i$	$\sum \sum y_{ij}^2$		$i = 4$

Výpočet součtu čtverců odchylek od průměru:

- pro celý pokus $S_c = \sum_{i=1}^p \sum_{j=1}^{m_j} y_{ij}^2 - \frac{Y_{..}^2}{n} =$

- pro skupiny $S_a = \sum_{i=1}^p \frac{Y_{i\cdot}^2}{n_i} - \frac{Y_{..}^2}{n} =$

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

Tabulka analýzy variance:

Zdroj proměnlivosti	Součet čtverců odchylek (S)	Stupeň volnosti (f)	Průměrný čtverec (MS)
mezi skupinami	S_a	$f_a = i - 1$	$MS_a = \frac{S_a}{f_a}$
uvnitř skupin	S_e	$f_e = n - i$	$MS_e = \frac{S_e}{f_e}$
celková	S_c	$f_c = n - 1$	$s_y^2 = \frac{S_c}{f_c}$

Proměnlivost	S	f	MS	F
Mezi skupinami	60194,90	3	20064,97	30,17
Uvnitř skupin	29264,08	44	665,09	
Celková	89458,98	47		

Hodnota F testu:

$$F = \frac{MS_a}{MS_e} = \frac{20064,97}{665,09} = 30,17$$

Vyhodnocení F testu: $F_{\text{vyp.}} > < F_{\text{tab.}}$

$F_{(3; 44; 0,05)} = 2,28$

$F_{(3; 44; 0,01)} = 4,26$

Následuje výsledek této analýzy variance, která byla provedena v programu SAS.

The ANOVA Procedure

Dependent Variable: snaska

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	60194.89583	20064.96528	30.17	<.0001
Error	44	29264.08333	665.09280		
Corrected Total	47	89458.97917			

R-Square	Coeff Var	Root MSE	snaska Mean
0.672877	15.92143	25.78939	161.9792

Source	DF	Anova SS	Mean Square	F Value	Pr > F
kohouti	3	60194.89583	20064.96528	30.17	<.0001

Různé způsoby výpočtů variance otcovské v SASu

```
proc glm ;
class otec;
MODEL y = otec /solution;

proc mixed;
class otec;
MODEL y = ;
random otec;

proc nested;
CLASS otec;
VAR y;

proc varcomp maxiter=50 method=reml;
class otec;
model y = otec ;
```

Syntaxe v programu R

Načtení dat z EXCELU (soubor ve formátu csv) a provedení 1f ANOVA

```
data2 <- read.table("C:/R/data2.csv") #načte data, ale i s hlavičkou a oddělené
#středníkem

data2 <- read.table("C:/R/data2.csv", header=T, sep=";") #načte data do tabulky, bez hlavičky
#a odstraní středník

anova.data2=aov(snaska~otec, data=data2) # provede 1 faktorovou ANOVA (efekt otce)

summary(anova.data2) # ukáže souhrnnou tabulku ANOVA

print(model.tables(anova.data2,"means"),digits=3) # zpráva průměrů a počtu subjektů

boxplot(y~kohouti,data=data2) # grafický souhrn

TukeyHSD(anova.data2) # následné testování pomocí Tukey vícenásobného porovnání průměrů

plot(TukeyHSD(anova.data2), ordered=T)

-----

lm.data2 <- lm(y~ kohouti, data=data2)# alternativní zápis pomocí lineárního modelu (lm)

anova(lm.data2) # ukáže souhrnnou tabulku ANOVA

summary(lm.data2) # odhad úrovní efektů
```