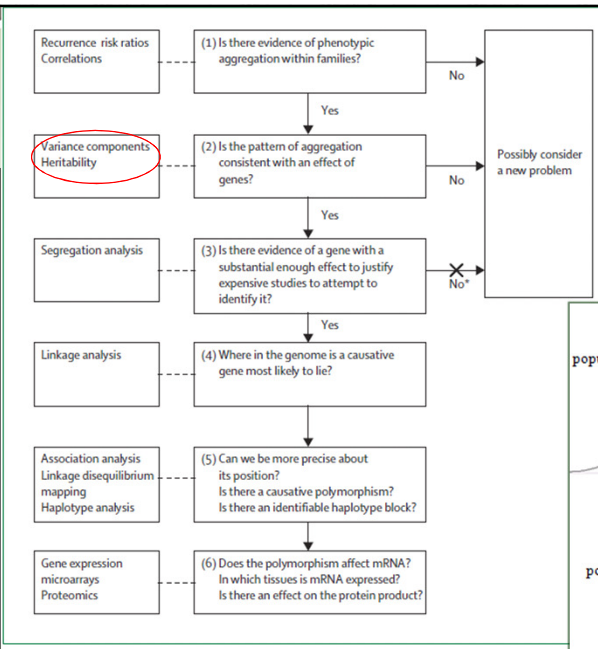
Päritavuskoefitsiendi hindamisest

Tanel Kaart
Statistilise geneetika seminar
04.02.2011

Epidemioloogias, aretuses, …

Burton, P.R., Tobin, M.D. and Hopper, J.L. (2005). **Key concepts in genetic epidemiology**. *Lancet* 366, 941–951.


Recurrence risk ratios
Correlations

(1) Is there evidence of phenotypic aggregation within families?

Variance components
Heritability

(2) Is the pattern of aggregation consistent with an effect of genes?

Segregation analysis

(3) Is there evidence of a gene with a substantial enough effect to justify expensive studies to attempt to identify it?

Linkage analysis

(4) Where in the genome is a causative gene most likely to lie?

Association analysis
Linkage disequilibrium mapping
Haplotype analysis

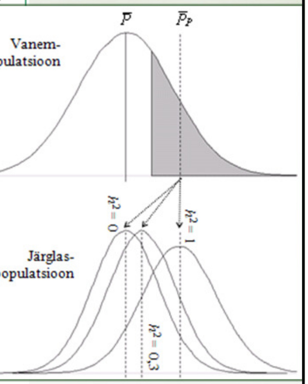
(5) Can we be more precise about its position?
Is there a causative polymorphism?
Is there an identifiable haplotype block?

Gene expression
microarrays
Proteomics

(6) Does the polymorphism affect mRNA?
In which tissues is mRNA expressed?
Is there an effect on the protein product?

Possibly consider a new problem

Figure 1: **Framework outlining systematic approach to identification and characterisation of genetic determinants of complex disease**
*It is probably illogical to stop trying to identify genetic determinants of disease simply because segregation analysis fails to provide significant evidence of major gene.



Geneetiline mudel

Fenotüüp = Genotüüp + Keskkond

$$P = \bar{P} + G + E$$

$$\mathrm{var}(P) = \mathrm{var}(G) + \mathrm{var}(E) + 2\,\mathrm{cov}(G,E)$$

$$G = A + D + \ldots$$

$$\mathrm{var}(G) = \mathrm{var}(A) + \mathrm{var}(D) + \ldots$$

G – genotüübiefekt, E – keskkonnaefekt, A – aretusväärtus, D – dominantsiefekt

---

$$A = ½A_S + ½A_D + MS$$

$$\mathrm{var}(A) = ¼\mathrm{var}(A_S) + ¼\mathrm{var}(A_D) + \mathrm{var}(MS)$$

Definitsioon

Päritavus (päritavuskoefitsient; *heritability*) on indiviidide-vahelise geneetilise erinevusega seletatav osa populatsiooni üldisest fenotüübilisest varieeruvusest.

$$h^2 = \mathrm{var}(G) \;/\; \mathrm{var}(P)$$

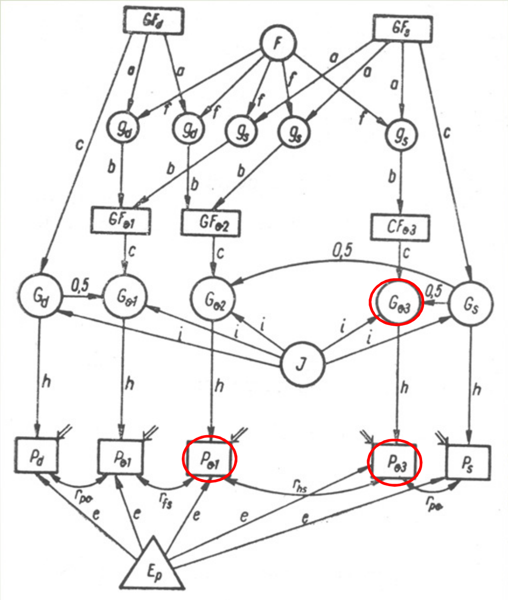
(**päritavus laiemas mõttes** peegeldab kogu võimalikku geneetilist mõju)

$$h^2 = \mathrm{var(A)} \;/\; \mathrm{var}(P)$$

(**päritavus kitsamas mõttes** mõõdab üksnes
alleelide aditiivsest toimest tingitud varieeruvuse osakaalu)

$$A = h^2 \times P$$

$$(b_{P\rightarrow A} = \mathrm{cov}(A,P) \;/\; \mathrm{var}(P) = \mathrm{var}(A) \;/\; \mathrm{var}(P) = h^2)$$



Tähistus

Tähistused h ja h^2 tulenevad Sewall Wright'i töödest.

$r(P,G) = h$

(korrelatsioon fenotüübi- ja genotüübiväärtuste vahel)

$r(P_{o1},P_{o3}) = r_{hs} = 0{,}25\times h^2$

(korrelatsioon poolõvede fenotüübiväärtuste vahel)

Kasutusel on ka H^2 ja h^2_w (päritavus laiemas mõttes)

Omadused

- Et $0 \le \mathrm{var}(A) \le \mathrm{var}(G) \le \mathrm{var}(P)$, siis $0 \le h^2 \le h^2_w \le 1$.
- Päritavuskoefitsient näitab tunnuse geneetilise muutlikkuse osa antud geneetilise struktuuriga populatsioonis konkreetsetes keskkonnatingimustes.
- Päritavuskoefitsient ei näita tunnuse päriliku tingituse määra ega mehhanismi üksik-indiviidide arengus.
- …

Eeldused

¤ Matemaatilised/epidemioloogilised eeldused,
(kasutatava statistilise mudeli vastavus andmetele):

- kõik kasutatavad andmed on usaldusväärsed ja nende struktuur vastavab sobitatavale mudelile;
- vaatlused on sõltumatud (tinglikustatult mudeli liikmete suhtes);
- uuritavate tunnuste ja hinnatavate efektide jaotused on korrektsed.

¤ Bioloogilised (geneetilised) eeldused:

- kehtib Hardy-Weinbergi seadus;
- rakendatav geneetiline mudel on korrektne (keskkond ja genotüüp on sõltumatud jmt);
- registreeritud ja kasutatavad sugulussidemed on täielikud ja korrektsed;
- …

Keskkonna ja genotüübi sõltumatus

Oletame lihtsuse mõttes, et $\mathrm{var}(G) = \mathrm{var}(E) = \sigma^2$.

Siis keskkonna ja genotüübi sõltumatuse korral

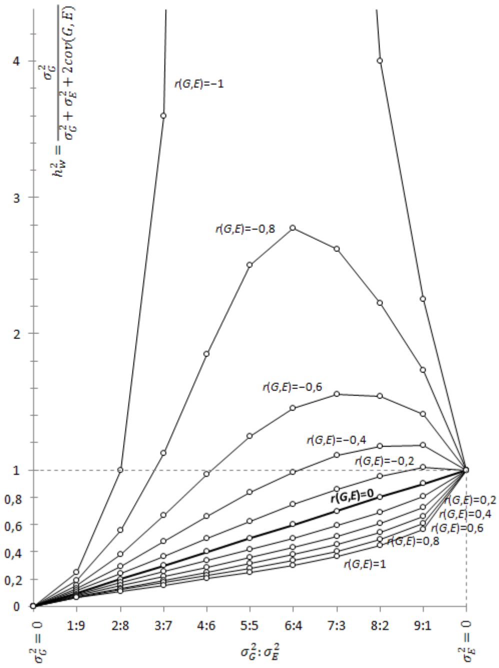
$$\mathrm{cov}(G,E) = 0,$$
$$\mathrm{var}(P) = \mathrm{var}(G) + \mathrm{var}(E) + 2\,\mathrm{cov}(G,E) = 2\sigma^2$$

ja $h^2_w = ½$.

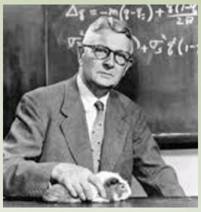
Keskkonna ja genotüübi sõltuvuse korral aga

$$h^2_w = \sigma^2 / 2[\sigma^2 + \mathrm{cov}(G,E)] = 1 / [2(1 + r_{GE})],$$

sest $\mathrm{cov}(G,E) = r_{GE}\,\sigma^2$.



Hindamismeetodid

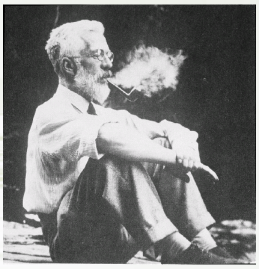


Kaks peamist koolkonda

¤ Regressioon- ja korrelatsioonanalüüs (Sewall Wright, Jay Laurence Lush; rajakoefitsientide meetod, selektsiooniindeksid)

Idee: sugulased on omavahel sarnasemad!

$$h^2 = b_{X\to Y} / 2f_{XY} \text{ või } h^2 = r_{XY} / 2f_{XY}$$



¤ Dispersioonanalüüs ja segamudelid (Sir Ronald Aylmer Fisher; dispersioonikomponendid, intraklasskorrelatsioonikordaja)

Idee: sugulased erinevad üksteisest vähem!

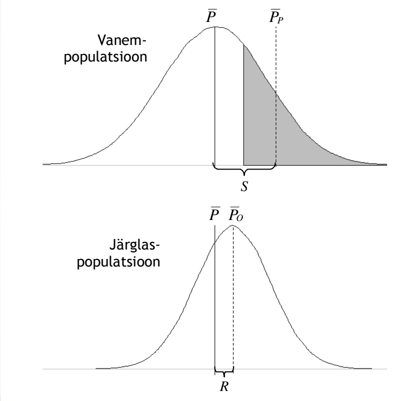
$$h^2 = t / 2f_{XY},\ t = \sigma_B^2 / (\sigma_B^2 + \sigma_W^2)$$

Selektsioonieksperimendid

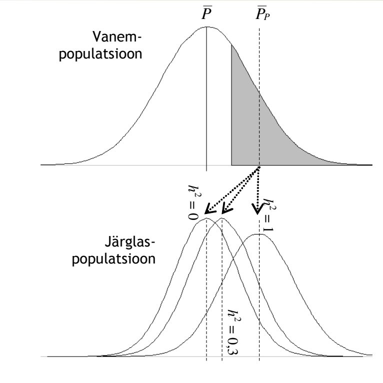
Selektsiooniefekti (R) ja selektsioonidiferentsi (S) suhe annab **realiseerunud** e **efektiivse päritavuse**:

$$h^2 = \frac{\bar{P}_O - \bar{P}}{\bar{P}_P - \bar{P}} = \frac{R}{S}$$

$$R = h^2 \times S$$



Valikuedu ja selektsioonidiferents



Statistiline mudel *vs* geneetiline mudel

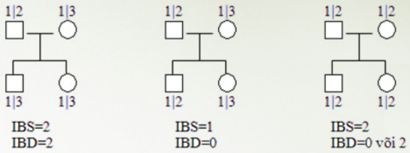
Statistiline mudel (*observational components*)	Geneetiline mudel (*causal components*)
$b_{X\to Y}$; r_{XY}; σ^2; …	var(A); var(G); var(E_C); …
$\Downarrow$	$\Downarrow$
$\widehat{h^2}$; …	h^2; …

Vaja, et mõista hinnatud päritavuskoefitsiendi olemust, ning et hinnata/kahtlustada/…

- dominantsiefekti,
- püsivat keskkonnamõju,
- emapoolset geneetilist efekti,
- imprintingut,
- epigeneetilist päritavust,
- …

Indiviidide vaheline sugulus

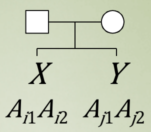
Indiviidide vaheline **suguluskoefitsient** (*coancestry*, *kinship coefficient*, …) on tõenäosus, et ühe indiviidi juhuslikult valitud lookusest juhuslikult valitud alleel on **päritolult identne** (*IBD*) teise indiviidi vastavast lookusest juhuslikult valitud alleeliga.



Olgu indiviidid X ja Y genotüüpidega $A_{i1}A_{i2}$ ning $A_{j1}A_{j2}$.

$$\begin{aligned} f_{XY} &= \mathrm{P}(2 \text{ juhuslikult valitud alleeli on } IBD) \\ &= \Sigma_{i,j}\, \mathrm{P}(A_i \equiv A_j \mid \text{juhuslikult valitud alleelid on } A_i \text{ ja } A_j) \\ &\quad \times \mathrm{P}(\text{juhuslikult valiti alleelid } A_i \text{ ja } A_j) \\ &= \tfrac{1}{4}\, [\mathrm{P}(A_{i1}\equiv A_{j1}) + \mathrm{P}(A_{i1}\equiv A_{j2}) + \mathrm{P}(A_{i2}\equiv A_{j1}) + \mathrm{P}(A_{i2}\equiv A_{j2})] \end{aligned}$$

Indiviidide vaheline sugulus



Olgu X ja Y **täisõved** (FS). Siis

$$f_{XY} = \text{P(2 juhuslikult valitud alleeli on IBD)}$$
$$= \tfrac{1}{4}\,[\underbrace{\text{P}(A_{i1}\equiv A_{j1})}_{\frac{1}{2}\times\frac{1}{2}} + \underbrace{\text{P}(A_{i1}\equiv A_{j2})}_{\frac{1}{2}\times\frac{1}{2}} + \underbrace{\text{P}(A_{i2}\equiv A_{j1})}_{\frac{1}{2}\times\frac{1}{2}} + \underbrace{\text{P}(A_{i2}\equiv A_{j2})}_{\frac{1}{2}\times\frac{1}{2}}] = \tfrac{1}{4}$$

Sugulusaste	Suguluskoefitsient f (või ϕ, ½r, …)
Indiviid ise	½
Ühemunarakukaksikud (MZ)	½
Täisõved (FS)	¼
Poolõved (HS)	⅛
Vanem ja järglane (PO)	¼
Vanavanem ja lapselaps	⅛

Indiviidide vaheline geneetiline kovariatsioon

Olgu indiviidid X ja Y genotüüpidega $A_{i1}A_{i2}$ ning $A_{j1}A_{j2}$.

Tähistagu a_{i1}, a_{i2}, a_{j1} ja a_{j2} alleeli- ning d_{i1i2} ja d_{j1j2} dominantsiefekte.

$$G_X = a_{i1} + a_{i2} + d_{i1i2} \text{ ja } G_Y = a_{j1} + a_{j2} + d_{j1j2}$$

$$\begin{aligned}\text{cov}(G_X, G_Y) = {} & \text{cov}(a_{i1}, a_{j1}) + \text{cov}(a_{i1}, a_{j2}) + \text{cov}(a_{i1}, d_{j1j2}) \\ & + \text{cov}(a_{i2}, a_{j1}) + \text{cov}(a_{i2}, a_{j2}) + \text{cov}(a_{i2}, d_{j1j2}) \\ & + \text{cov}(d_{i1i2}, a_{j1}) + \text{cov}(d_{i1i2}, a_{j2}) + \text{cov}(d_{i1i2}, d_{j1j2}).\end{aligned}$$

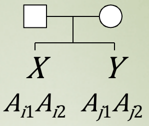
$\text{cov}(a, d) = 0$;

$$\text{cov}(a_x, a_y) = \begin{cases} 0, & \text{kui } x \not\equiv y\ (\neg IBD) \\ \text{var}(A)/2\,, & \text{kui } x \equiv y\ (IBD) \end{cases},$$

sest $\text{var}(A) = \text{var}(a_1 + a_2) = 2\text{var}(a_1)$, mistõttu $\text{var}(a_1) = \text{cov}(a_1, a_1) = \text{var}(A)/2$

$$\text{cov}(d_{xy}, d_{wz}) = \begin{cases} 0, & \text{kui } xy \not\equiv wz \text{ (mõlemad alleelid } \neg IBD) \\ \text{var}(D), & \text{kui } xy \equiv wz \text{ (mõlemad alleelid } IBD) \end{cases}.$$

Indiviidide vaheline geneetiline kovariatsioon



Olgu X ja Y genotüübiväärtustega

$$G_X = a_{i1} + a_{i2} + d_{i1i2} \text{ ja } G_Y = a_{j1} + a_{j2} + d_{j1j2}$$

täisõved (FS).

Siis on võimalikud kolm varianti:

Variant	Tõenäosus	$\mathrm{cov}(G_X, G_Y)$
#IBD = 0	¼	0
#IBD = 1	½	$\mathrm{var}(A) / 2$
#IBD = 2	¼	$\mathrm{var}(A) + \mathrm{var}(D)$

$$\begin{aligned}\mathrm{cov}(G_X, G_Y) &= \tfrac{1}{2} \times \mathrm{var}(A)/2 + \tfrac{1}{4} \times [\mathrm{var}(A) + \mathrm{var}(D)] \\ &= \tfrac{1}{2} \times \mathrm{var}(A) + \tfrac{1}{4} \times \mathrm{var}(D)\end{aligned}$$

Indiviidide vaheline geneetiline kovariatsioon

Üldine valem:

$$\mathrm{cov}(G_X, G_Y) = 2 f_{XY} \mathrm{var}(A) + g_{XY} \mathrm{var}(D)$$

Sugulusaste	f_{XY} (või ϕ_{XY}, ...)	g_{XY} (või Δ_{XY}, ...)	$\mathrm{cov}(G_X, G_Y)$
Indiviid ise	½	1	$\mathrm{var}(A) + \mathrm{var}(D)$
Ühem. kaksikud	½	1	$\mathrm{var}(A) + \mathrm{var}(D)$
Täisõved	¼	¼	½ $\mathrm{var}(A)$ + ¼ $\mathrm{var}(D)$
Poolõved	⅛	0	¼ $\mathrm{var}(A)$
Vanem ja järglane	¼	0	½ $\mathrm{var}(A)$
V.vanem ja l.laps	⅛	0	¼ $\mathrm{var}(A)$
Abikaasad	0	0	0

Vanemate-järglaste vaheline regressioon

Järglane ja üks vanem:

$$b_{\mathrm{P}\to\mathrm{O}} = \frac{\mathrm{cov}(P_\mathrm{P}, P_\mathrm{O})}{\mathrm{var}(P_\mathrm{P})} = \frac{½\mathrm{var}(A)}{\mathrm{var}(P)} = ½h^2 \;\Rightarrow\; \boxed{h^2 = 2 \times b_{\mathrm{P}\to\mathrm{O}}}$$

$$r_\mathrm{PO} = \frac{\mathrm{cov}(P_\mathrm{P}, P_\mathrm{O})}{\sqrt{\mathrm{var}(P_\mathrm{P}) \times \mathrm{var}(P_\mathrm{O})}} = \frac{½\mathrm{var}(A)}{\mathrm{var}(P)} = ½h^2, \text{ kui } \mathrm{var}(P_\mathrm{P}) = \mathrm{var}(P_0)$$

Järglane ja vanemate keskmine:

$$b_{\bar{\mathrm{P}}\to\mathrm{O}} = \frac{\mathrm{cov}(P_{\bar{\mathrm{P}}}, P_\mathrm{O})}{\mathrm{var}(P_{\bar{\mathrm{P}}})} = \frac{½\mathrm{var}(A)}{½\mathrm{var}(P)} = h^2 \;\Rightarrow\; \boxed{h^2 = b_{\bar{\mathrm{P}}\to\mathrm{O}}}, \text{ kui } \mathrm{var}(P_\mathrm{P1}) = \mathrm{var}(P_\mathrm{P2})$$

$$r_{\bar{\mathrm{P}}\mathrm{O}} = \frac{\mathrm{cov}(P_{\bar{\mathrm{P}}}, P_\mathrm{O})}{\sqrt{\mathrm{var}(P_{\bar{\mathrm{P}}}) \times \mathrm{var}(P_\mathrm{O})}} = \frac{½\mathrm{var}(A)}{\sqrt{½}\mathrm{var}(P)} = \sqrt{½}h^2, \text{ kui } \ldots$$

Järglaste keskmine ja üks vanem:

$$b_{\mathrm{P}\to\bar{\mathrm{O}}} = \frac{\mathrm{cov}(P_\mathrm{P}, P_{\bar{\mathrm{O}}})}{\mathrm{var}(P_\mathrm{P})} = \frac{½\mathrm{var}(A)}{\mathrm{var}(P)} = ½h^2 \;\Rightarrow\; \boxed{h^2 = 2 \times b_{\mathrm{P}\to\bar{\mathrm{O}}}}$$

$$r_{\mathrm{P}\bar{\mathrm{O}}} = \frac{\mathrm{cov}(P_\mathrm{P}, P_{\bar{\mathrm{O}}})}{\sqrt{\mathrm{var}(P_\mathrm{P}) \times \mathrm{var}(P_{\bar{\mathrm{O}}})}} = \frac{½\mathrm{var}(A)}{\sqrt{1/n}\,\mathrm{var}(P)}, \text{ kui } \ldots$$

Õvede vaheline korrelatsioon

Poolõved:

$$r_\mathrm{HS} = \frac{\mathrm{cov(HS)}}{\sqrt{\mathrm{var}(P) \times \mathrm{var}(P)}} = \frac{¼\mathrm{var}(A)}{\mathrm{var}(P)} = ¼h^2 \;\Rightarrow\; \boxed{h^2 = 4 \times r_\mathrm{HS}}$$

Täisõved (=kahemunarakukaksikud):

$$r_\mathrm{FS} (= r_\mathrm{DZ}) = \frac{\mathrm{cov(FS)}}{\sqrt{\mathrm{var}(P) \times \mathrm{var}(P)}} = \frac{½\mathrm{var}(A) + ¼\mathrm{var}(D) + \mathrm{var}(E_C)}{\mathrm{var}(P)} \geq ½h^2$$

Ühemunarakukaksikud:

$$r_\mathrm{MZ} = \frac{\mathrm{cov(MZ)}}{\sqrt{\mathrm{var}(P) \times \mathrm{var}(P)}} = \frac{\mathrm{var}(G) + \mathrm{var}(E_C)}{\mathrm{var}(P)} \geq h_w^2$$

Kaksikute uuringuis:

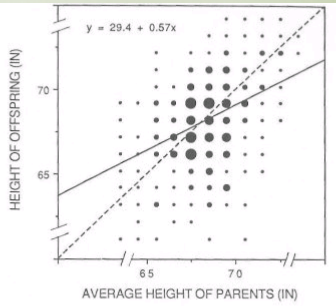
$$\boxed{h_w^2 = \frac{\mathrm{var}(G)}{\mathrm{var}(P)} \approx 2(r_\mathrm{MZ} - r_\mathrm{DZ})}$$

$$c^2 = \frac{\mathrm{var}(E_C)}{\mathrm{var}(P)} \approx 2r_\mathrm{DZ} - r_\mathrm{MZ}$$

Vanemate-järglaste vaheline regressioon

Sir Francis Galton'i (1889) uuring esitamaks järglaste pikkust (928 indiviidi) vanemate keskmise pikkuse kaudu (205 vanempaari).

Wikipedia



Regressioonikordaja on alati < 1,

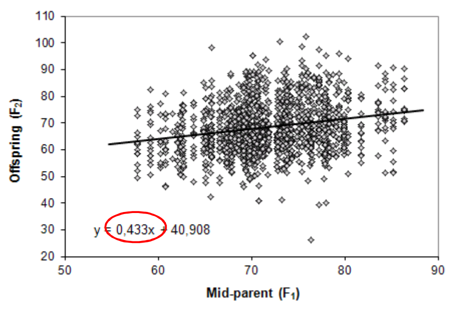
- st et järglaste fenotüüp on keskmiselt populatsiooni keskmisele lähemal (võrreldes vanemate fenotüübiga).
- *Regression* – tagasilangus, regressioon

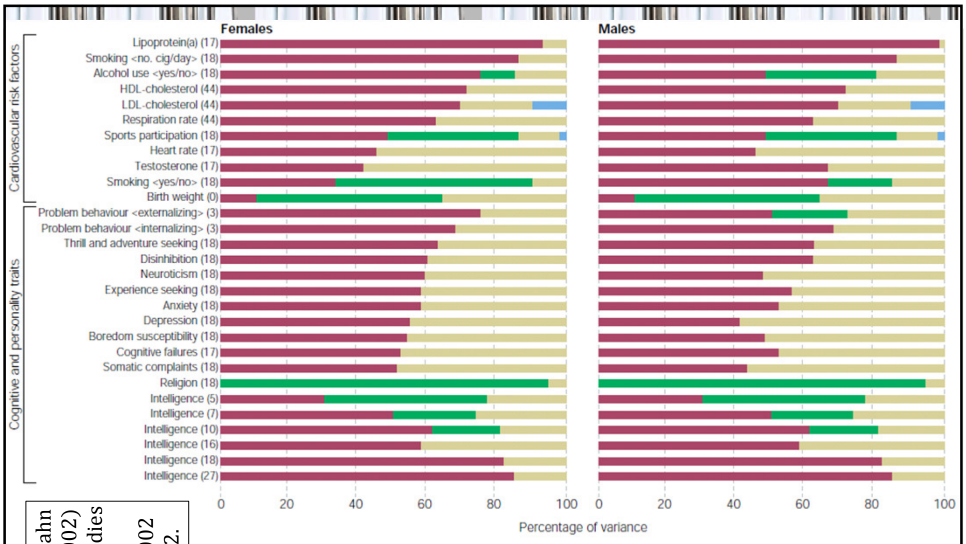
Vanemate-järglaste vaheline regressioon

Näide. Uurimisobjekt: võsavaksik; uuritav tunnus: nuku mass.

$F_1 \rightarrow F_2$	h^2
$(S+D)/2 \rightarrow O$	0,433
$S \rightarrow O$	0,581
$D \rightarrow O$	0,449
$(S+D)/2 \rightarrow O_m$	0,322
$S \rightarrow O_m$	0,472
$D \rightarrow O_m$	0,346
$(S+D)/2 \rightarrow O_f$	0,569
$S \rightarrow O_f$	0,721
$D \rightarrow O_f$	0,559

S – *sire* (isa), D – *dam* (ema),
O_m – *male offspring* (isane järglane),
O_f – *female offspring* (emane järglane)

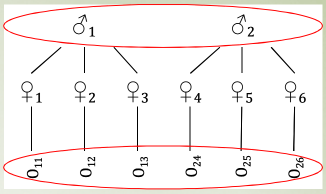




Boomsma D, Busjahn A, Peltonen L. (2002) Classical twin studies and beyond. *Nat Rev Genet.* 2002 Nov;3(11):872-82.

Figure 2 | **Examples of results from classical twin analysis.** Percentage of variances explained by genetic factors (purple), by shared environmental factors (green), by unique environmental influences (beige) and by differences in age (blue). The phenotypes were assessed in Dutch female and male twins (and in some cases also in their parents and siblings). The number in the brackets gives the modal age of the sample in years. Phenotypes include cardiovascular risk factors, and personality and cognitive traits. The heritability for a behavioural as well as cardiovascular risk factor, such as the number of cigarettes smoked per day, is nearly as high as for lipoprotein(a) levels. For personality traits and indices of psychopathology, heritability is ~50%, with a higher estimate for internalizing and externalizing problems in young children. Heritability of intelligence is age dependent and increases rapidly between 5 and 18 years of age. There are few differences in heritability between females and males. HDL, high-density lipoprotein; LDL, low-density lipoprotein. Modified with permission from REF. 23 © (1999) Australian Academic Press.

Poolõvede dispersioonanalüüs



$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + s_i + e_{ij}, \quad \sigma_y^2 = \sigma_s^2 + \sigma_e^2,$$

Et ühelt poolt

$$\operatorname{cov}(y_{ij}; y_{ij'}) = \operatorname{cov}(\mathrm{HS}) = \tfrac{1}{4}\operatorname{var}(A)$$

ja teiselt poolt

$$\operatorname{cov}(y_{ij}; y_{ij'}) = \operatorname{cov}(s_i + e_{ij}; s_i + e_{ij'}) = \operatorname{cov}(s_i; s_i) = \sigma_s^2$$

siis järelikult

$$\sigma_s^2 = \tfrac{1}{4}\operatorname{var}(A),$$

$$\sigma_e^2 = \tfrac{3}{4}\operatorname{var}(A) + \operatorname{var}(D) + \cdots + \operatorname{var}(E)$$

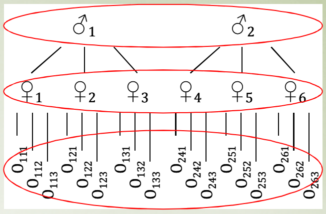
ja

$$\boxed{h^2 = 4\sigma_s^2 \big/ \left(\sigma_s^2 + \sigma_e^2\right)}$$

$$P_{ij} = \bar{P} + \overbrace{\tfrac{1}{2}A_{S_i}}^{s_i} + \overbrace{\tfrac{1}{2}A_{D_{ij}} + MS_{ij} + D_{ij} + \ldots + E_{ij}}^{e_{ij}}$$

$$\operatorname{var}(P) = \underbrace{\tfrac{1}{4}\operatorname{var}(A)}_{\sigma_s^2} + \underbrace{\tfrac{1}{4}\operatorname{var}(A) + \tfrac{1}{2}\operatorname{var}(A) + \operatorname{var}(D) + \ldots + \operatorname{var}(E)}_{\sigma_e^2}$$

Täisõvede dispersioonanalüüs



$$y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + s_i + d_{ij} + e_{ijk},$$

$$\sigma_y^2 = \sigma_s^2 + \sigma_d^2 + \sigma_e^2$$

$$\sigma_s^2 = \mathrm{cov(HS)} = \tfrac{1}{4}\mathrm{var}(A)$$

$$\sigma_e^2 = \sigma_y^2 - \underbrace{\mathrm{cov(FS)}}_{\frac{1}{2}\mathrm{var}(A)+\frac{1}{4}\mathrm{var}(D)+\mathrm{var}(E_C)} = \tfrac{1}{2}\mathrm{var}(A) + \tfrac{3}{4}\mathrm{var}(D) + \mathrm{var}(E)$$

$$\sigma_d^2 = \sigma_y^2 - \sigma_s^2 - \sigma_e^2$$
$$= \mathrm{cov(FS)} - \mathrm{cov(HS)} = \tfrac{1}{4}\mathrm{var}(A) + \tfrac{1}{4}\mathrm{var}(D) + \mathrm{var}(E_C)$$

Seega

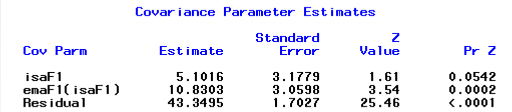
$$\boxed{h^2 = 4\sigma_s^2 \big/ \left(\sigma_s^2 + \sigma_d^2 + \sigma_e^2\right)}$$

$$h^2 \le 2\left(\sigma_s^2 + \sigma_d^2\right) \big/ \left(\sigma_s^2 + \sigma_d^2 + \sigma_e^2\right)$$

$c^2 = \left(\sigma_d^2 - \sigma_s^2\right)/\sigma_y^2$, kui $\mathrm{var}(D) = 0$; $d^2 = 4\left(\sigma_d^2 - \sigma_s^2\right)/\sigma_y^2$, kui $\mathrm{var}(E_C) = 0$; …

Täisõvede dispersioonanalüüs

Näide. Uurimisobjekt: võsavaksik; uuritav tunnus: nuku mass.



```
               Covariance Parameter Estimates

                                    Standard         Z
 Cov Parm            Estimate          Error     Value        Pr Z

 isaF1                 5.1016         3.1779      1.61      0.0542
 emaF1(isaF1)         10.8303         3.0598      3.54      0.0002
 Residual             43.3495         1.7027     25.46      <.0001
```

$$h^2 = 4\sigma_s^2 \big/ \left(\sigma_s^2 + \sigma_d^2 + \sigma_e^2\right)$$
$$= 4 \times 5{,}10 / (5{,}10 + 10{,}83 + 43{,}35) = 0{,}344$$

$$d^2 = 4\left(\sigma_d^2 - \sigma_s^2\right) \big/ \sigma_y^2$$
$$= 4 \times (10{,}83 - 5{,}10) / (5{,}10 + 10{,}83 + 43{,}35)$$
$$= 0{,}387$$

Mäletatavasti:

$F_1 \to F_2$	h^2
$(S+D)/2 \to O$	0,433
$S \to O$	0,581
$D \to O$	0,449
$(S+D)/2 \to O_m$	0,322
$S \to O_m$	0,472
$D \to O_m$	0,346
$(S+D)/2 \to O_f$	0,569
$S \to O_f$	0,721
$D \to O_f$	0,559

S – *sire* (isa), D – *dam* (ema),
O_m – *male offspring* (isane järglane),
O_f – *female offspring* (emane järglane)

Looma mudel

Geneetiline mudel	Statistiline mudel

$$P_i = \bar{P} + A_i + E_i$$

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + a_i + e_i$$

$$\sigma_y^2 = \sigma_a^2 + \sigma_e^2$$

$$\mathrm{var}\begin{pmatrix}\mathbf{a}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\mathbf{A}\sigma_a^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_N\sigma_e^2\end{pmatrix}$$

Näide:

```
(1) ---- [2] ---- (3)
       |       |
      [4] ---- (5)
           |
          (6)
```

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0.5 & 0 & 0.25\\ 0 & 1 & 0 & 0.5 & 0.5 & 0.5\\ 0 & 0 & 1 & 0 & 0.5 & 0.25\\ 0.5 & 0.5 & 0 & 1 & 0.25 & 0.625\\ 0 & 0.5 & 0.5 & 0.25 & 1 & 0.625\\ 0.25 & 0.5 & 0.25 & 0.625 & 0.625 & 1.125 \end{pmatrix}$$

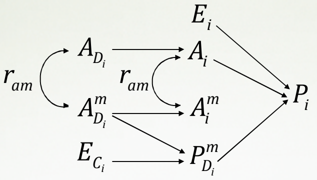
Maatriks **A**
(**aditiivgeneetilise suguluse maatriks**)

- omab üht rida ja üht veergu iga sugupuus sisalduva indiviidi tarvis;
- elemendid on aditiivgeneetilise suguluse kordajad ($2f_{XY}$) + inbriiding

$$\boxed{h^2 = \sigma_a^2 / \sigma_y^2}$$

Maternaalse efektiga looma mudel

Geneetiline mudel	Statistiline mudel



$$y_{ijk} = \mathbf{x}_i'\boldsymbol{\beta} + a_i + m_j + c_k + e_{ijk}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Za} + \mathbf{Wm} + \mathbf{Sc} + \mathbf{e}$$

$$\sigma_y^2 = \sigma_a^2 + \sigma_m^2 + 2\sigma_{am} + \sigma_c^2 + \sigma_e^2$$

$$\mathrm{var}\begin{pmatrix}\mathbf{a}\\ \mathbf{m}\\ \mathbf{c}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{am} & \mathbf{0} & \mathbf{0}\\ \mathbf{A}\sigma_{am} & \mathbf{A}\sigma_m^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_b\sigma_c^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_N\sigma_e^2\end{pmatrix}$$

Päritavus: $\boxed{h^2 = \sigma_a^2 / \sigma_y^2}$

Maternaalne päritavus: $\boxed{m^2 = \sigma_m^2 / \sigma_y^2}$

Püsiva keskkonnamõju osakaal: $c^2 = \sigma_c^2 / \sigma_y^2$

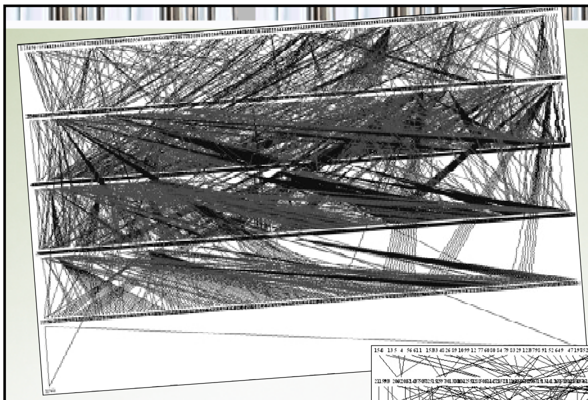
Otsese ja maternaalse geneetilise efekti vaheline korrelatsioon:

$$r_{am} = \sigma_{am} \Big/ \sqrt{\sigma_a^2 \cdot \sigma_m^2}$$

Geneetiliste parameetrite hinnangud sõltuvalt põlvnemisandmete hulgast

Näide.

Eesti mustapealiste lammaste analüüs (2006)



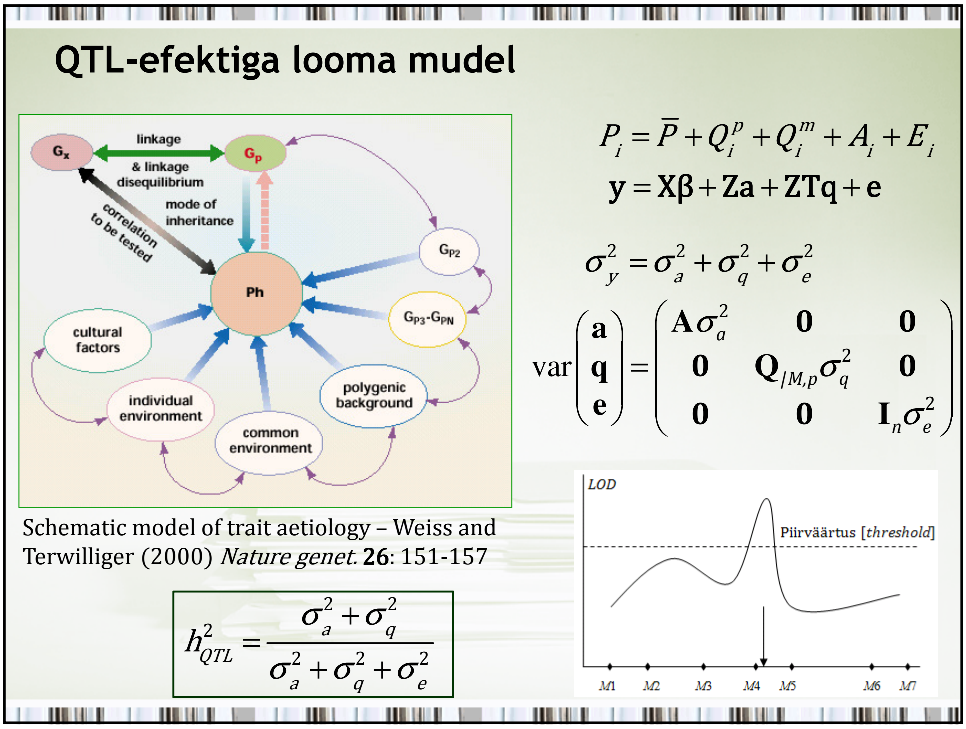
1886 looma
(üksnes vanemad ja vana-vanemad;
inbriidsete loomade arv – 116)

2451 looma (kõik teadaolevad esivanemad; inbriidsete loomade arv – 636)

Geneetiliste parameetrite hinnangud sõltuvalt mudelist ja põlvnemisandmete hulgast

Eesti mustapealiste lammaste võõrutusmass. Mudel: $y_{ijk} = \mathbf{x}_i'\boldsymbol{\beta} + a_i + m_j + c_k + e_{ijk}$

	Põlvnemis-andmed	Hinnatud geneetilised parameetrid				
		$\hat{\sigma}_e^2$	$\hat{h}^2$	$\hat{m}^2$	$\hat{r}_{am}$	$\hat{c}^2$
Mudel 1	isad	48,67	0,256	-	-	-
Mudel 2	isad+emad	44,82	0,274	-	-	-
Mudel 3	≤2 eellasgen.	43,03	0,162	-	-	-
Mudel 4		41,67	0,118	-	-	0,065
Mudel 5		41,74	0,139	0,043	0,446	-
Mudel 6		41,30	0,127	0,012	0,922	0,048
Mudel 7	≤12 eellasgen.	43,00	0,165	-	-	-
Mudel 8		41,62	0,129	-	-	0,066
Mudel 9		41,97	0,125	0,055	0,087	-
Mudel 10		41,66	0,118	0,024	0,260	0,041



QTL-efektiga looma mudel

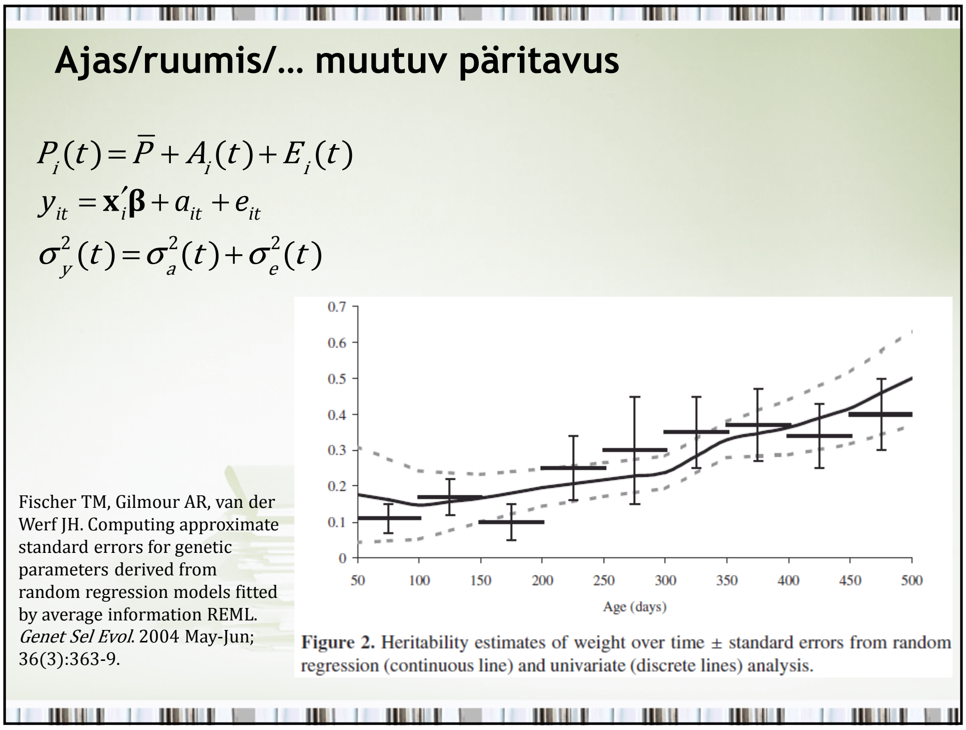
Schematic model of trait aetiology – Weiss and Terwilliger (2000) *Nature genet.* **26**: 151-157

$$P_i = \bar{P} + Q_i^p + Q_i^m + A_i + E_i$$
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Za} + \mathbf{ZTq} + \mathbf{e}$$

$$\sigma_y^2 = \sigma_a^2 + \sigma_q^2 + \sigma_e^2$$

$$\mathrm{var}\begin{pmatrix}\mathbf{a}\\ \mathbf{q}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\mathbf{A}\sigma_a^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{|M,p}\sigma_q^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_n\sigma_e^2\end{pmatrix}$$

$$\boxed{h_{QTL}^2 = \frac{\sigma_a^2 + \sigma_q^2}{\sigma_a^2 + \sigma_q^2 + \sigma_e^2}}$$



Ajas/ruumis/… muutuv päritavus

$$P_i(t) = \bar{P} + A_i(t) + E_i(t)$$
$$y_{it} = \mathbf{x}_i'\boldsymbol{\beta} + a_{it} + e_{it}$$
$$\sigma_y^2(t) = \sigma_a^2(t) + \sigma_e^2(t)$$

Fischer TM, Gilmour AR, van der Werf JH. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. *Genet Sel Evol.* 2004 May-Jun; 36(3):363-9.

Figure 2. Heritability estimates of weight over time ± standard errors from random regression (continuous line) and univariate (discrete lines) analysis.

Ajas/ruumis/… muutuv päritavus

$$a_t = a_0 + a_1 \times t, \quad \operatorname{cov}(a_t, a_s)$$

$$= \operatorname{cov}(a_0 + a_1 t, a_0 + a_1 s)$$

$$= \operatorname{var}(a_0) + t\operatorname{cov}(a_1, a_0) + s\operatorname{cov}(a_0, a_1) + st\operatorname{var}(a_1)$$

$$= \begin{pmatrix} 1 & t \\ 1 & s \end{pmatrix} \operatorname{var}\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ t & s \end{pmatrix} = \mathbf{\Phi K \Phi}^{\mathrm{T}}$$

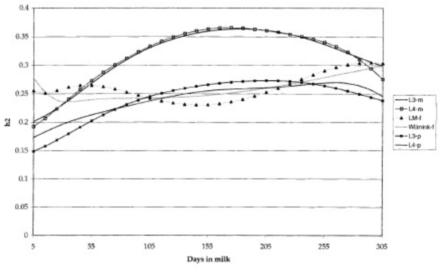


Figure 1. Heritability (h^2) for milk (m), fat (f), and protein (p) yield as a function of DIM. Third (L3) and fourth (L4) order Legendre polynomials and a function proposed by Lidauer and Mantysaari (1999) (LM) functions were fitted in an animal model for milk and protein, and the LM and the Wilmink function were fitted for fat (Modified from Jakobsen, 2000).

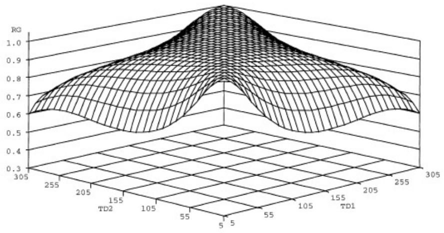
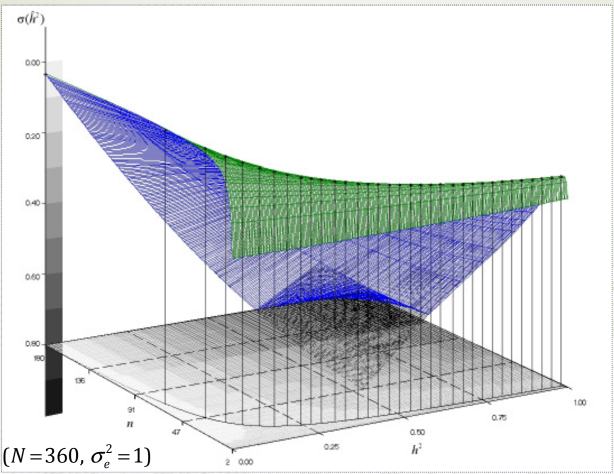


Figure 2. Genetic correlations (r_g) between test-day milk yields on one test day (TD_1) and another test day (TD_2) using a 4 order Legendre polynomial on 8075 Holstein cows (Modified from Jakobsen, 2000).

Jensen J. Genetic evaluation of dairy cattle using test-day models. *J Dairy Sci.* 2001 Dec; 84(12):2803-12.

Optimaalne disain

- Regressioonanalüüs vanemate keskmise baasil ($h^2 = b_{\bar{\mathrm{P}} \to \mathrm{O}}$).
- Korrelatsioonanalüüs poolõvede baasil ($h^2 = 4 \times r_{\mathrm{HS}}$)



$(N = 360,\ \sigma_e^2 = 1)$

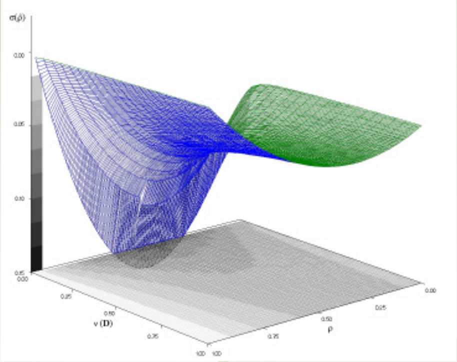
- Poolõvede dispersioonanalüüsil

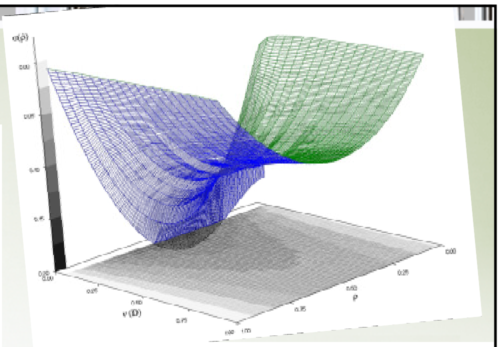
$$\operatorname{var}(\hat{h}^2) \to \min,$$

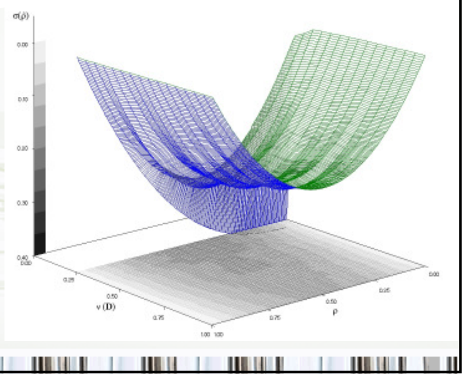
$$\text{kui} \quad n = \frac{N(\tau+1)+1}{N\tau+2}$$

$$\left(\tau = \sigma_s^2 / \sigma_e^2\right)$$

Sõltuvus andmete tasakaalulisusest



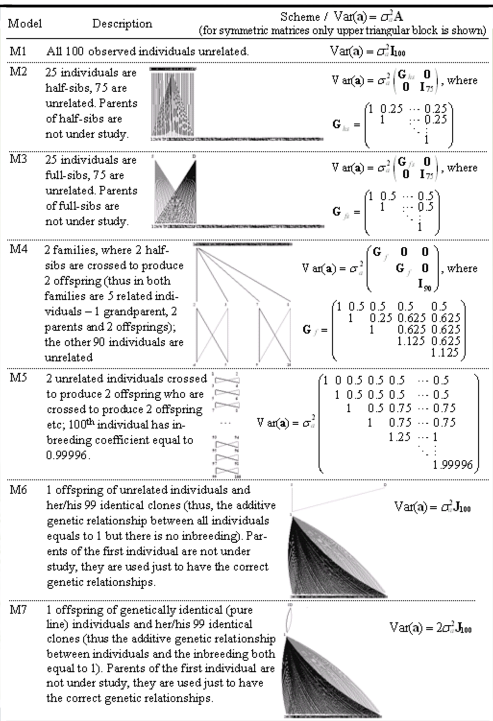




Intraklasskorrelatsioonikordaja (ρ) (=päritavuskoefitsiendi) hinnangute täpsus poolõvede analüüsil sõltuvalt andmete tasakaalulisusest (gruppide/isade arv $a = 90, 20, 4$; $N = 360$, $\sigma_e^2 = 1$)

Geneetiliste parameetrite hinnangute täpsus sõltuvalt sugulussidemete tugevusest

$\mathrm{Var}(\hat{\sigma}^2_{a,REML})$ x10		
$h^2 = 0.8$	$h^2 = 0.25$	$h^2 = 0.01$
5.0505	0.3591	0.2061
4.4824	0.3502	0.2060
3.5541	0.3372	0.2059
4.2435	0.3542	0.2061
0.2534	0.2131	0.2027
0.2020	0.2020	0.2020
0.2038	0.1995	0.2020



Model	Description	Scheme / $\mathrm{Var}(\mathbf{a}) = \sigma_a^2\mathbf{A}$ (for symmetric matrices only upper triangular block is shown)
M1	All 100 observed individuals unrelated.	$\mathrm{Var}(\mathbf{a}) = \sigma_a^2\mathbf{I}_{100}$
M2	25 individuals are half-sibs, 75 are unrelated. Parents of half-sibs are not under study.	$\mathrm{Var}(\mathbf{a}) = \sigma_a^2\begin{pmatrix}\mathbf{G}_{hs} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{75}\end{pmatrix}$, where $\mathbf{G}_{hs} = \begin{pmatrix}1 & 0.25 & \cdots & 0.25\\ & 1 & \cdots & 0.25\\ & & \ddots & \vdots\\ & & & 1\end{pmatrix}$
M3	25 individuals are full-sibs, 75 are unrelated. Parents of full-sibs are not under study.	$\mathrm{Var}(\mathbf{a}) = \sigma_a^2\begin{pmatrix}\mathbf{G}_{fs} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{75}\end{pmatrix}$, where $\mathbf{G}_{fs} = \begin{pmatrix}1 & 0.5 & \cdots & 0.5\\ & 1 & \cdots & 0.5\\ & & \ddots & \vdots\\ & & & 1\end{pmatrix}$
M4	2 families, where 2 half-sibs are crossed to produce 2 offspring (thus in both families are 5 related individuals – 1 grandparent, 2 parents and 2 offsprings); the other 90 individuals are unrelated	$\mathrm{Var}(\mathbf{a}) = \sigma_a^2\begin{pmatrix}\mathbf{G}_f & \mathbf{0} & \mathbf{0}\\ & \mathbf{G}_f & \mathbf{0}\\ & & \mathbf{I}_{90}\end{pmatrix}$, where $\mathbf{G}_f = \begin{pmatrix}1 & 0.5 & 0.5 & 0.5 & 0.5\\ & 1 & 0.25 & 0.625 & 0.625\\ & & 1 & 0.625 & 0.625\\ & & & 1.125 & 0.625\\ & & & & 1.125\end{pmatrix}$
M5	2 unrelated individuals crossed to produce 2 offspring who are crossed to produce 2 offspring etc; 100th individual has inbreeding coefficient equal to 0.99996.	$\mathrm{Var}(\mathbf{a}) = \sigma_a^2\begin{pmatrix}1 & 0 & 0.5 & 0.5 & 0.5 & \cdots & 0.5\\ & 1 & 0.5 & 0.5 & 0.5 & \cdots & 0.5\\ & & 1 & 0.5 & 0.75 & \cdots & 0.75\\ & & & 1 & 0.75 & \cdots & 0.75\\ & & & & 1.25 & \cdots & 1\\ & & & & & \ddots & \vdots\\ & & & & & & 1.99996\end{pmatrix}$
M6	1 offspring of unrelated individuals and her/his 99 identical clones (thus, the additive genetic relationship between all individuals equals to 1 but there is no inbreeding). Parents of the first individual are not under study, they are used just to have the correct genetic relationships.	$\mathrm{Var}(\mathbf{a}) = \sigma_a^2\mathbf{J}_{100}$
M7	1 offspring of genetically identical (pure line) individuals and her/his 99 identical clones (thus the additive genetic relationship between individuals and the inbreeding both equal to 1). Parents of the first individual are not under study, they are used just to have the correct genetic relationships.	$\mathrm{Var}(\mathbf{a}) = 2\sigma_a^2\mathbf{J}_{100}$

Päritavuskoefitsiendi väärtus sõltuvalt kasutatavast mudelist

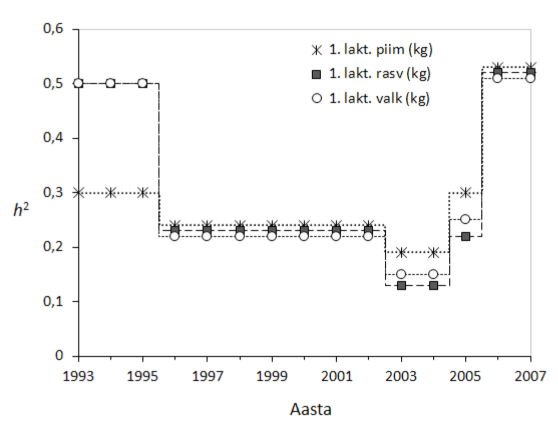
Jõudluskontrollikeskus

1993-1995 Isa mudel

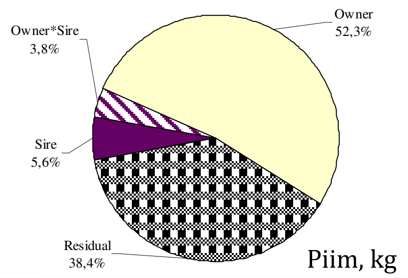
1996-2002 Looma mudel

2003-2005 Kontrollpäeva looma mudel

2006-… Juhuslike regressioonikordajatega kontrollpäeva looma mudel (*random regression test day animal model*)



Genotüübi-keskkonna interaktsioon



Eesti holsteini tõugu lehmade uuring 2003. aastal;

loomade arv – 13194;
isade arv – 231;
omanike arv – 1029