Introduction to Biometrical Genetics {in the classical twin design}

Conor Dolan
&
Elizabeth Prom-Wormley

Boulder 2020

Slide acknowledgements (fonts all over the place and inconsistent color coding):

Manuel Ferreira, **Pak Sham**, ***Shaun Purcell***, Sarah Medland, and **Sophie van der Sluis**

Outline

Slides 3 -14: What is it about essentially + some basic statistics

Slides 15 – 18: Basic genetic terms

Slides 19 – 28: How a QTL contributes to phenotypic variance

Slides 30 – 37: How a QTL contributes to phenotypic variance

Slides 39 – 51: Genetic variance as a source of phenotypic covariance

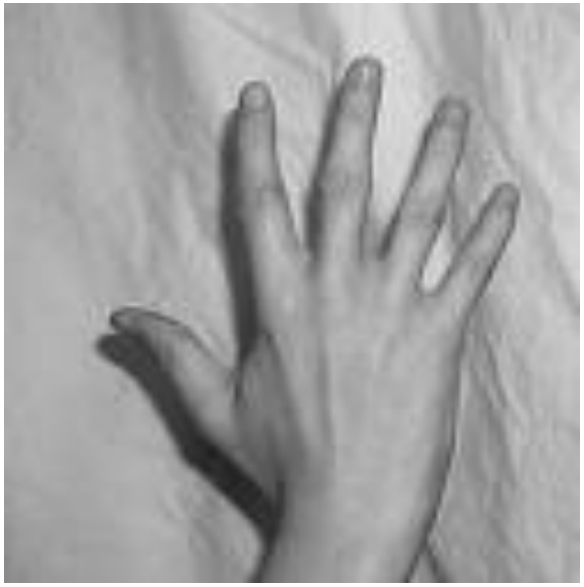
Slides 52 - 67 : Genetic variance as a source of phenotypic covariance

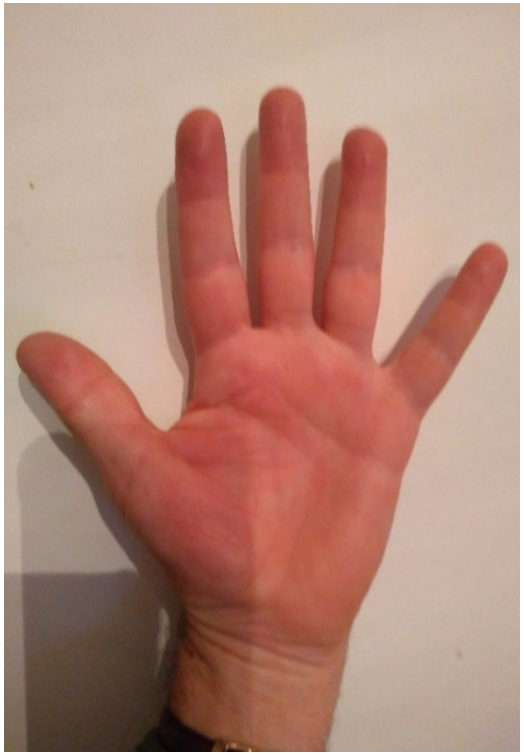
Slides 68 - 76: Not part of this talk

What are we on about when we talk about genetic influences?

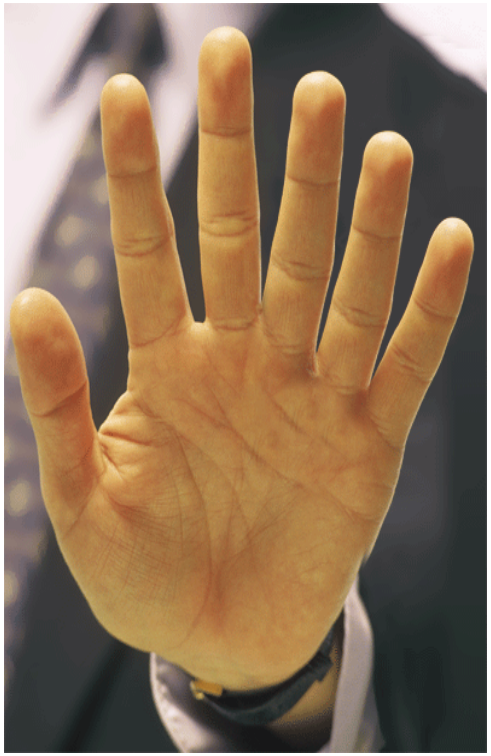
"Having 5 fingers genetically determined"

"*DNA includes a blueprint to build a hand*"

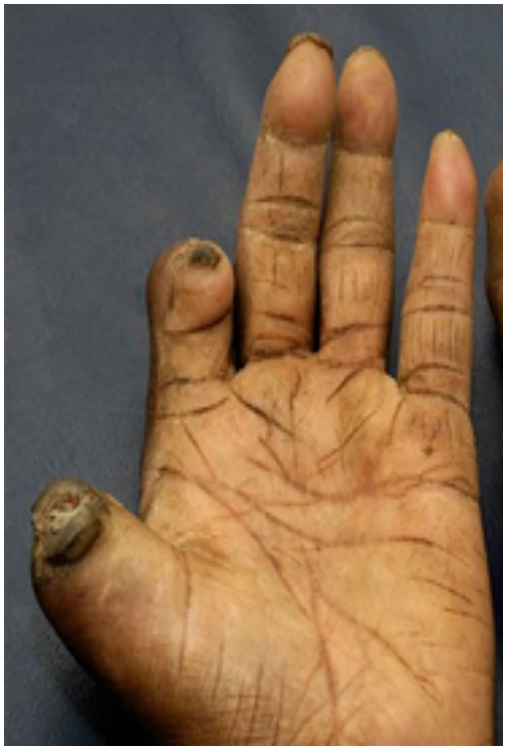




normal



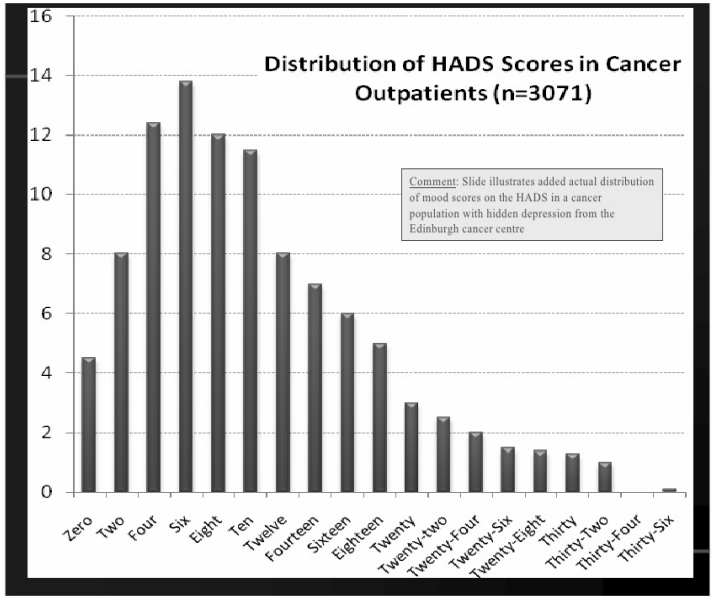
polydactyly

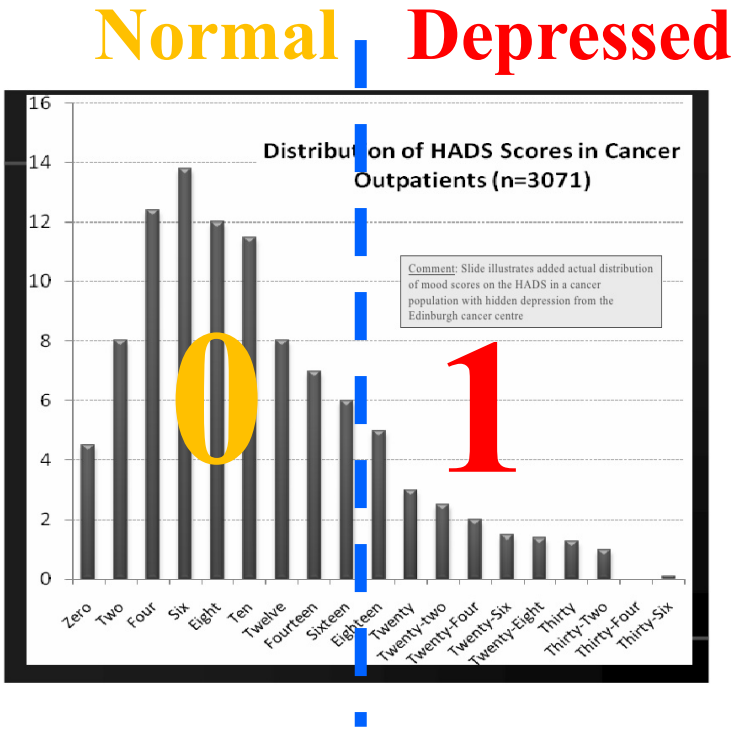


leprosy

phenotypic difference **6 – 5 = +1** with a genetic cause
(related to genetic difference - mutation)
phenotypic difference **3 – 5 = -2** with an environmental cause
(related to environmental difference – bacterium)

Phenotype**: continuously varying, genetically complex.**
e.g. (ideally) normally distributed
e.g., binary (dichotomous, 0-1 coded) phenotype
(based on continuous phenotype; liability threshold model).





The phenotype is a **quantitative** trait, a **metric** trait, a **complex** trait

Genetically complex:

Individual differences in the phenotype are subject to the effects of many genes of small effects, a.k.a. **polygenes, minor genes**. How many? Hundreds (Educational Attainment, Height) … Thousands….?

Phenotypic individual differences are attributable to genetic individual differences in a large number of polygenes, a.k.a. **QTLs (quantitative trait loci).**

Polygenicity implies phenotypic continuous distributions

People differ phenotypically
Q. How to quantify individual differences?

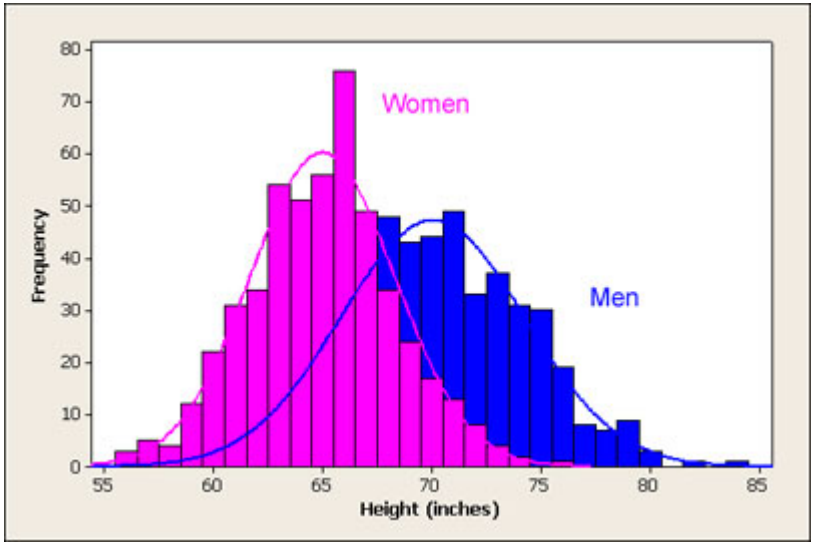
Variance: s^2, σ^2, σ^2_X, var(X), V_X

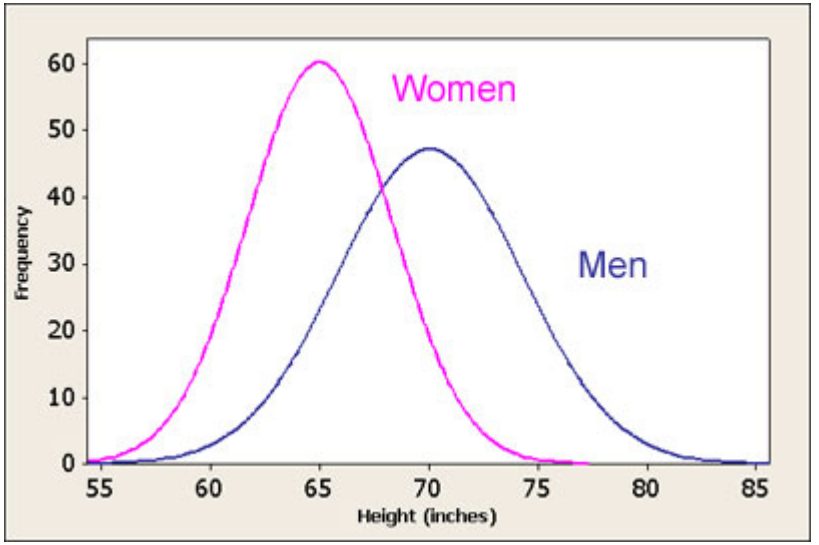
mean (X) $$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

variance (X) $$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2$$

x_i is the phenotypic value of person i (i=1,...,N)

height in inches - sex differences in the distribution
how? sex differences in mean and in variance.





Some continuously distributed phenotypes are approximately normally distributed e.g., height, IQ.

Means, Variances and Covariances

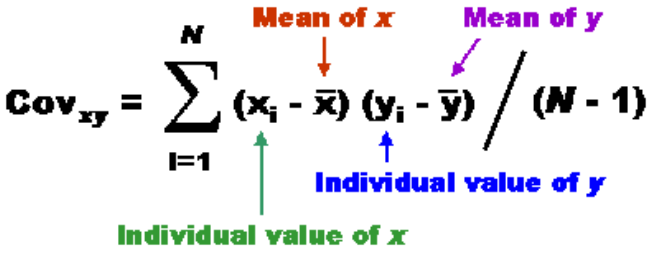
$$\mu = E(X) = \sum_i x_i f(x_i)$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$Var(X) = E(X-\mu)^2 = \sum_i (x_i-\mu)^2 f(x_i)$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i-\mu)^2$$

$$Cov(X,Y) = E(X-\mu_X)(Y-\mu_Y) = \sum_i (x_i-\mu_X)(y_i-\mu_Y) f(x_i,y_i)$$



We need the covariance: express the phenotypic relatedness among family members

Important to understand!

1,1,2,2,3,4,5,5,6,6

mean = (1+1+2+2+3+4+5+5+6+6)/10
= 36/12 = 3.5

f(1) = 2/10 = .2	.2*1 +
f(2) = 2/10 = .2	.2*2 +
f(3) = 1/10 = .1	.1*3 +
f(4) = 1/10 = .1	.1*4 +
f(5) = 2/10 = .2	.2*5 +
f(6) = 2/10 = .2	.2*6

	3.5

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\mu = E(X) = \sum_i x_i f(x_i)$$

1,1,2,2,2,3,4,5,5,5,6,6

mean = 3.5

$$\mu = E(X) = \sum_i x_i f(x_i)$$

f(1) = 2/10 = .2 .2*(1-3.5)^2 +
f(2) = 2/10 = .2 .2*(2-3.5)^2 +
f(3) = 1/10 = .1 .1*(3-3.5)^2 +
f(4) = 1/10 = .1 .1*(4-3.5)^2 +
f(5) = 2/10 = .2 .2*(5-3.5)^2 +
f(6) = 2/10 = .2 .2*(6-3.5)^2

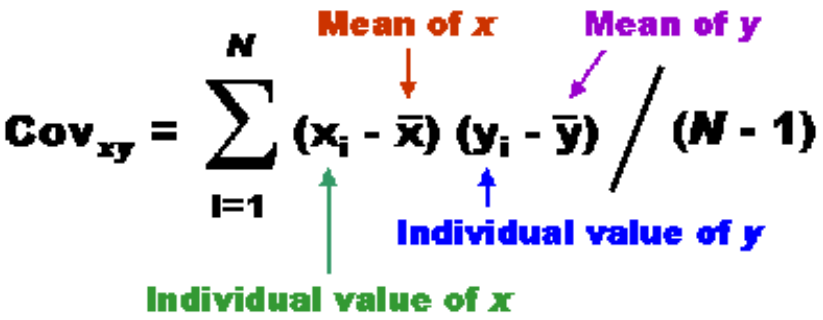
$$Var(X) = E(X-\mu)^2$$
$$= \sum_i (x_i - \mu)^2 f(x_i)$$

variance = 3.45
standard deviation (stdev) = √variance
stdev = √3.45 = 1.857

covariance

$$Cov(X,Y) = E(X - \mu_X)(Y - \mu_Y)$$

$$\sum_i (x_i - \mu_X)(y_i - \mu_Y) f(x_i, y_i)$$



correlation

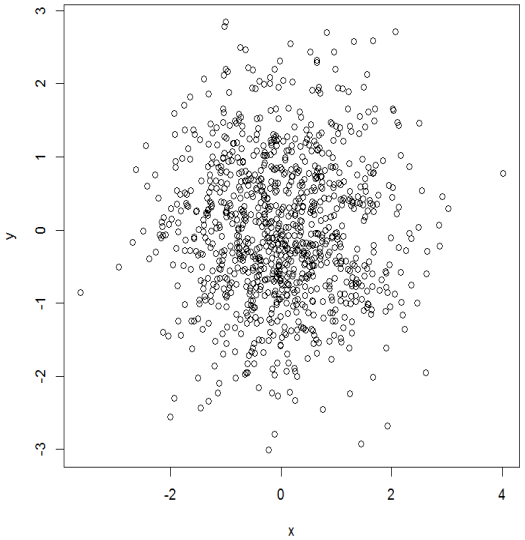
Cor(X,Y) = Cov(X,Y) / √ [Var(X)*var(Y)] =
= Cov(X,Y) / [stdev(X)*stdev(Y)]

Cor(X,Y) is – stand-alone - interpretable

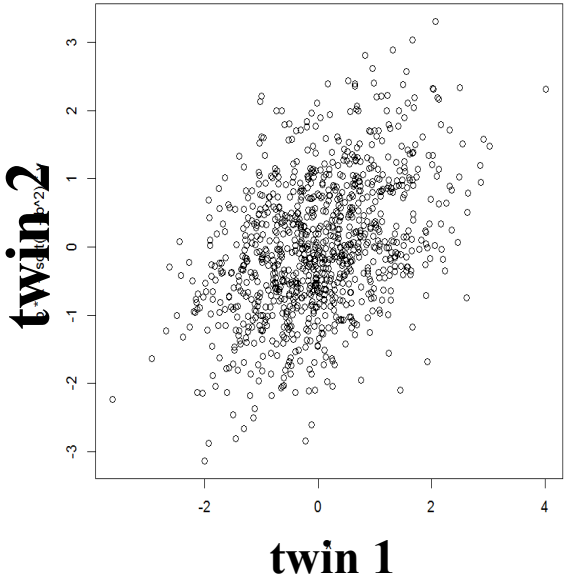
MZ covariance is 291.... uninterpretable
MZ correlation is .80 .... interpretable

Linear association between continuous variables: covariance or Pearson Product Moment (PPM) Correlation Coefficient, **r**.

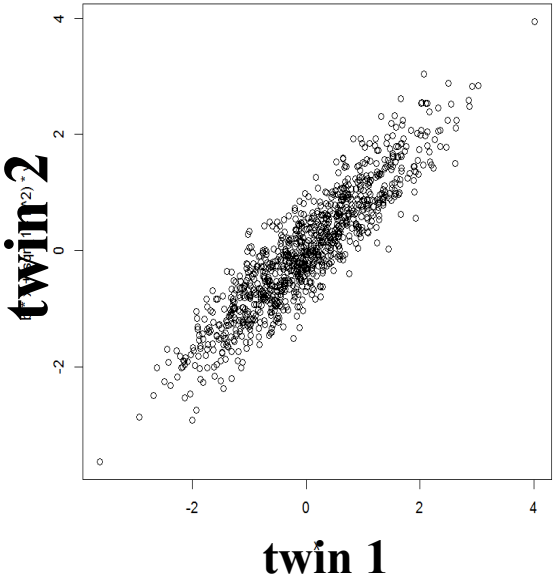
r = 0.00



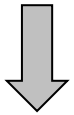
DZ r = .40



MZ r = .90



To what extent, and how, are
individual differences in genetic makeup, and
individual differences in environmental factors, related to
phenotypic (observed) ***individual differences*** *?*



To what extent, and how, do
individual differences in genotypes, and
individual differences in environmental factors, explain
phenotypic (observed) ***variance****?*

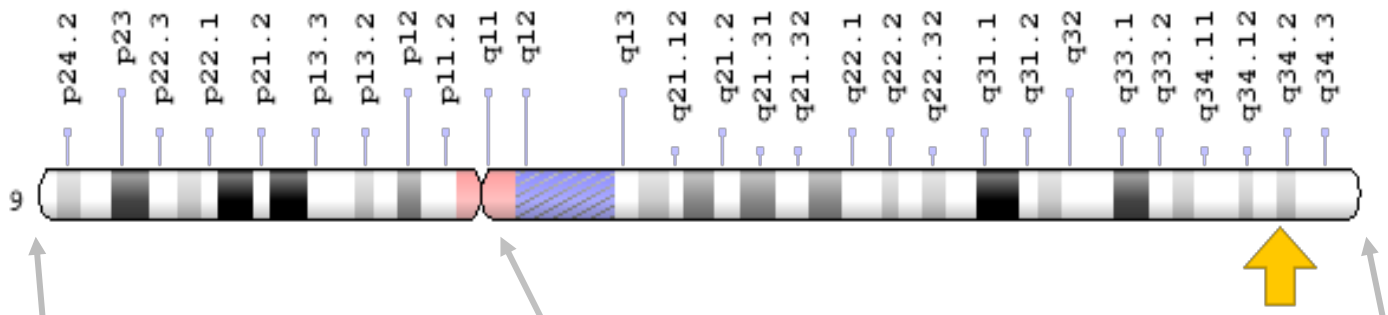
$$Var(X) = E(X-\mu)^2$$
$$= \sum_i (x_i - \mu)^2 f(x_i)$$

terminology

- **QTL Quantative trait locus**: a sequence of DNA base pairs (may be a SNP "snip": single base pair). a.k.a. genetic variant
- **Autosomal locus**: the site of the QTL on a chromosome (22 pairs + XY). Humans are dipoid (22 pairs autosomal chromosomes + sex chromosomes XY or XX). An autosomal locus is located on one of the 22 pairs.
- **Allele**: an alternative form of a gene at a locus
- **Genotype**: the combination of alleles at a particular locus
- **Complex phenotype**: an observed characteristic, which displays individual differences (in part due to differences at many loci... how many?)

3 alleles A-B-O (blood group)

Locus: autosomal chromosome 9, long arm (q), position 34.2



locus
(of allele A,B, or O)

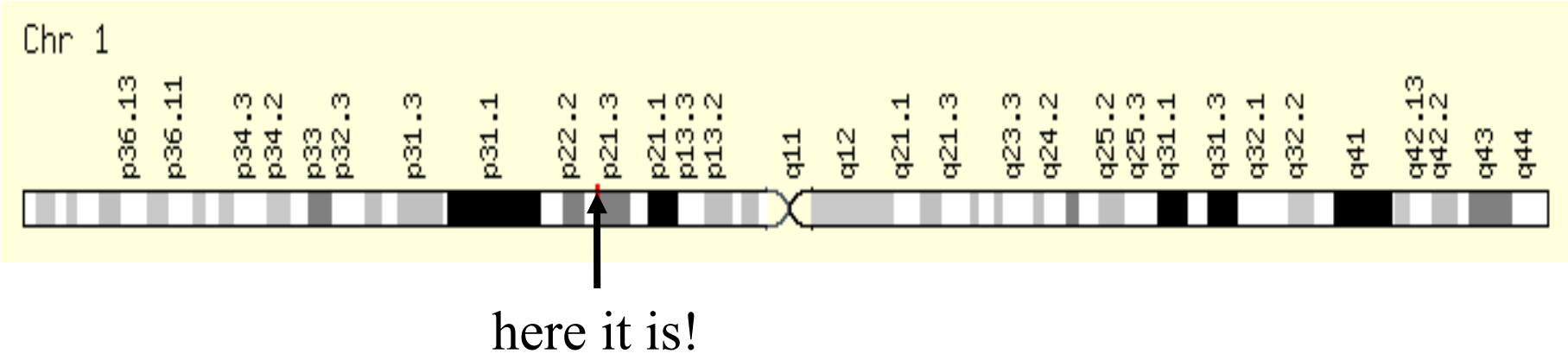
This is a member of a pair (autosomal chromosomes come in pairs).

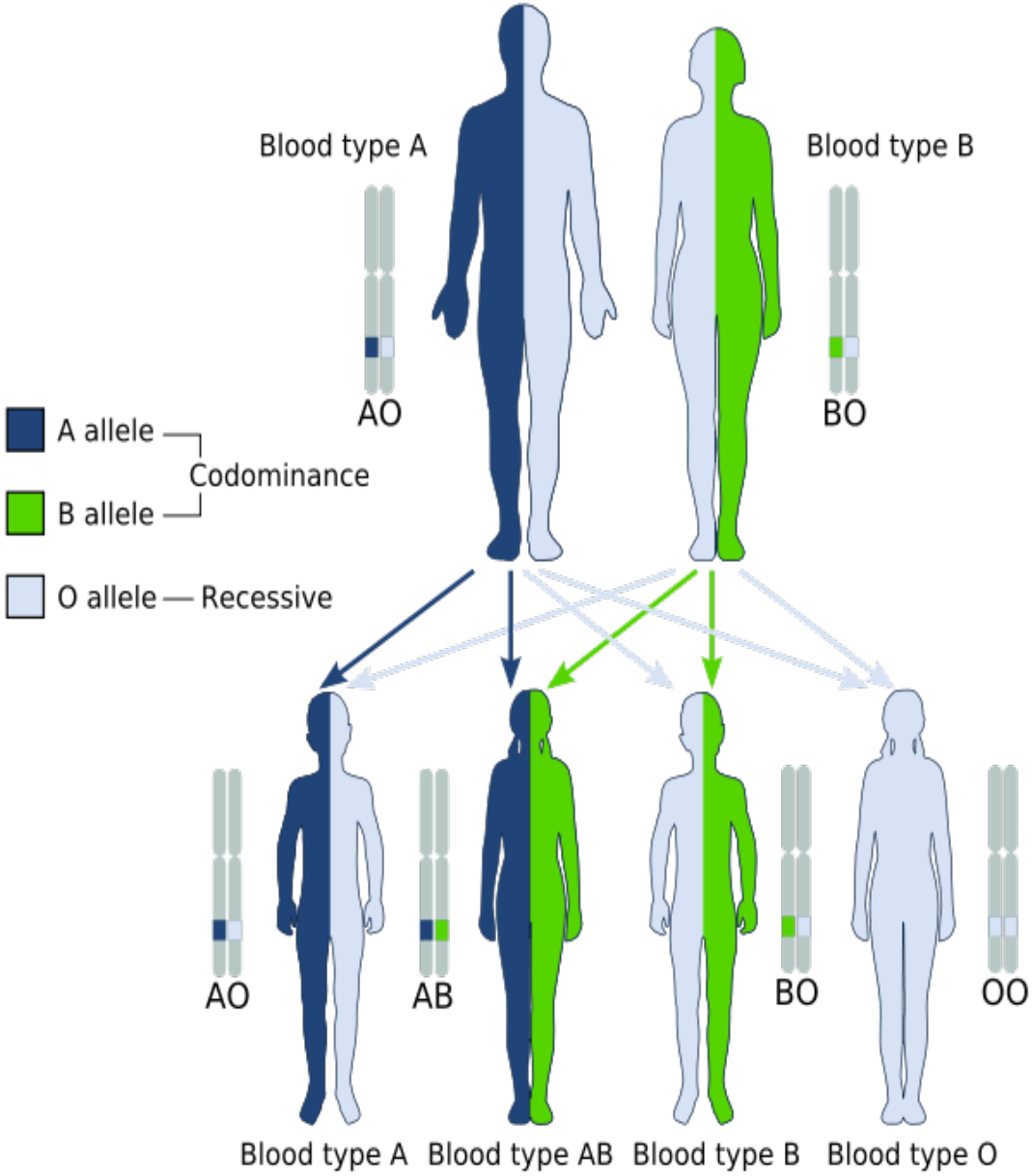
Example of a QTL: FNBP1L gene

The FNBP1L gene has been associated with intelligence in two studies:

- *Mol. Psychiatry 2012* **16** (10), 996-1005
- *Mol. Psychiatry 2011* **19**(2): 2538.

This gene is on chromosome 1 (1p22,1), and it comprises 106531 bases (106.5Kb). Within this gene the SNP rs236330 specifically is associated with intelligence.





A-B-O locus
chr 9 location 9q34.2

Mendelian inheritance

The law of segregation

Consider a single diallelic locus with alleles A and a

Set up the model to relate the locus (A-a) to the phenotypic variance.

How does the locus contribute to phenotypic individual differences?

Population level

1. Allele frequencies (QTL: diallelic autosomal)

- A single autosomal locus, with two alleles
  - Biallelic a.k.a. diallelic

- Alleles *A* and *a*
  - Frequency of *A* is *p*
  - Frequency of *a* is *q* = 1 – *p*

} frequencies in the population

- Every individual inherits two alleles
  - A genotype is the combination of the two alleles
  - e.g. *AA*, *aa* (the homozygotes) or *Aa* (the heterozygote)
    - * what are the genotype frequencies?

Biometrical model for single biallelic

QTL

- Biallelic locus
  - Genotypes: *AA, Aa, aa*
  - Genotype frequencies: p^2, $2pq$, q^2

Genotype frequencies (Random mating)

Father's gametes (sperm) \ Mother's gametes (egg)	*A* (*p*)	*a* (*q*)
A (*p*)	*AA* (p^2)	*Aa* (*pq*)
a (*q*)	*aA* (*qp*)	*aa* (q^2)

Hardy-Weinberg Equilibrium frequencies

$P(AA) = p^2$

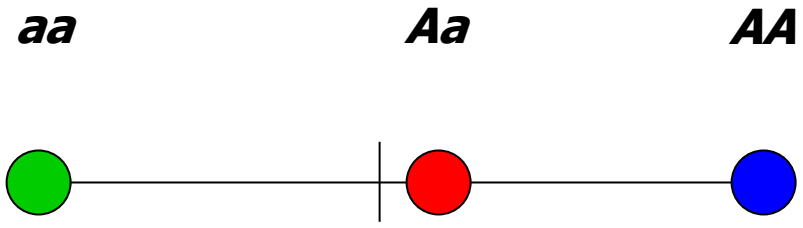
$P(Aa) = 2pq$

$P(aa) = q^2$

$p^2 + 2pq + q^2 = 1$

Phenotype level: contribution to continuous variation

Biometric Model



$-a$ $\quad d \quad +a$ ← Genotypic effect

$\mu - a \quad \mu + d \quad \mu + a$ ← phenotypie means within each genotype (aa, Aa, AA) .....conditional on genotype

Q: Phenotypic mean conditional on genotype means what?

A: Take all ***aa*** individuals and calculate their mean phenotypic value: $\mu - a$ (the phenotypic mean **conditional** on genotype ***aa***)

Biometrical model for single biallelic

QTL

1. Contribution of the QTL to the Mean

Genotypes	AA	Aa	aa
Effect, x	$\mu + a$	$\mu + d$	$\mu - a$
Frequencies, $f(x)$	p^2	$2pq$	q^2

$(\mu+a)(p^2) + (\mu+d)(2pq) + (\mu-a)(q^2) =$
$\mu + a(p^2) + d(2pq) - a(q^2) =$
$\mu + a(p-q) + 2pqd$

see slide 11!

the unconditional mean $\mu + a(p-q) + 2pqd = \mu + m$
contribution of the QTL $m = a(p-q) + 2pqd$

Biometrical model for single biallelic

QTL

2. Contribution of the QTL to the Variance (X)

Genotypes	AA	Aa	aa
Effect (x)	$\mu + a$	$\mu + d$	$\mu - a$
Frequencies, f(x)	p^2	$2pq$	q^2

$$m = a(p-q) + 2pqd$$

$$s^2_{Ph\_QTL} = (a-m)^2p^2 + (d-m)^2 2pq + (-a-m)^2q^2$$

see slide 12!

Q: WAIT!!! What happened to μ?

$$s^2_{Ph\_QTL} = (a\text{-}m)^2p^2 + (d\text{-}m)^2 2pq + (-a\text{-}m)^2q^2$$

actually

$$((\mu + \mathbf{a})–(\mu+m))^2p^2 + ((\mu + \mathbf{d})–(\mu+m))^2 2pq + ((\mu\text{-}\mathbf{a})–(\mu +m))^2q^2$$

$$((\mu + \mathbf{a})–(\mu+m)) = (\mu + \mathbf{a}–\mu\text{-}m) = (\mathbf{a}\text{-}m)$$

A: μ cancels out.

Biometrical model for single biallelic

QTL

$$s^2_{Ph\_QTL} = (a-m)^2p^2 + (d-m)^2 2pq + (-a-m)^2q^2$$

$$= \underline{2pq[a+(q-p)d]^2} + \underline{(2pqd)^2}$$

$$= s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)}$$

Additive or linear effects give rise to variance component
$s^2_{Ph\_QTL(A)} = 2*pq[a+(q-p)d]^2$ (additive genetic variance)

Dominance or *within local allelic interaction* effects give rise to variance component
$s^2_{Ph\_QTL(D)} = (2pqd)^2$ (dominance variance)

Biometrical model for single biallelic

QTL

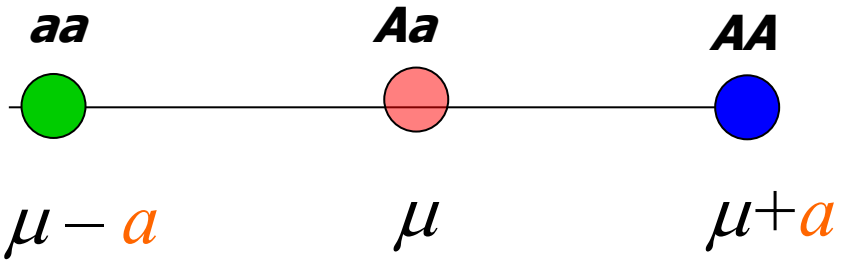
$$s^2_{Ph\_QTL} = (a-m)^2p^2 + (d-m)^2 2pq + (-a-m)^2q^2$$

$$= \underline{2pq[a+(q-p)d]^2} + \underline{(2pqd)^2}$$

$$= s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)}$$

Additive effects: $s^2_{Ph\_QTL(A)} = 2*pq[a]^2$

Dominance effects: $s^2_{Ph\_QTL(D)} = 0 \quad (d=0)$



Biometrical model for single biallelic

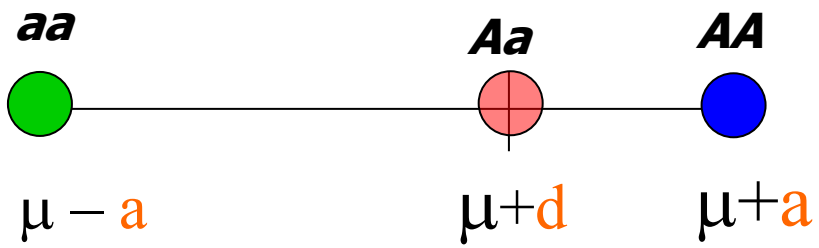
QTL $s^2_{Ph\_QTL} = (a\text{-}m)^2p^2 + (d\text{-}m)^2 2pq + (-a\text{-}m)^2q^2$

$= 2pq[a+(q\text{-}p)d]^2 + (2pqd)^2$

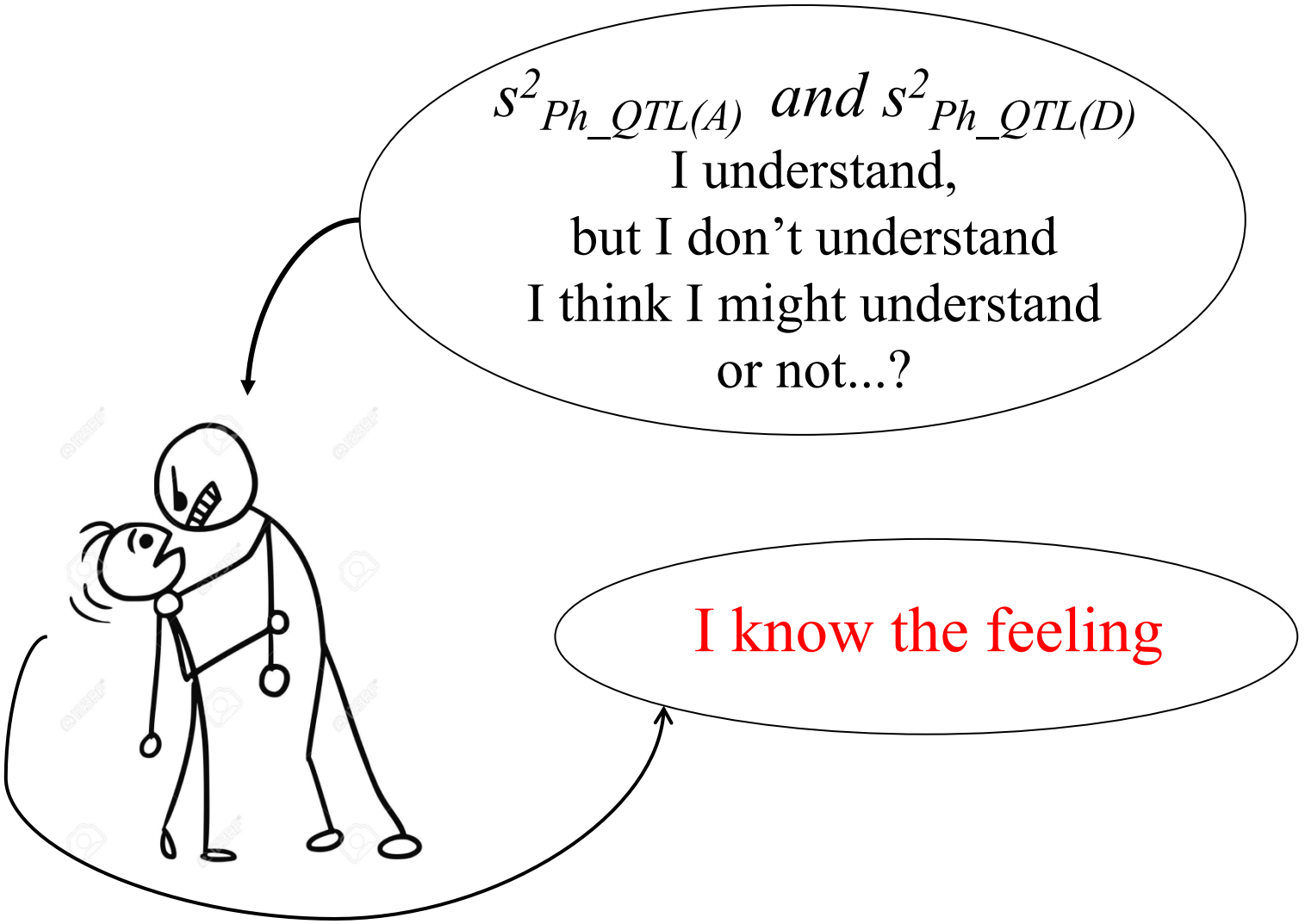
$= s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)}$

Additive effects: $s^2_{Ph\_QTL(A)} = 2*pq[a+(q\text{-}p)d]^2$

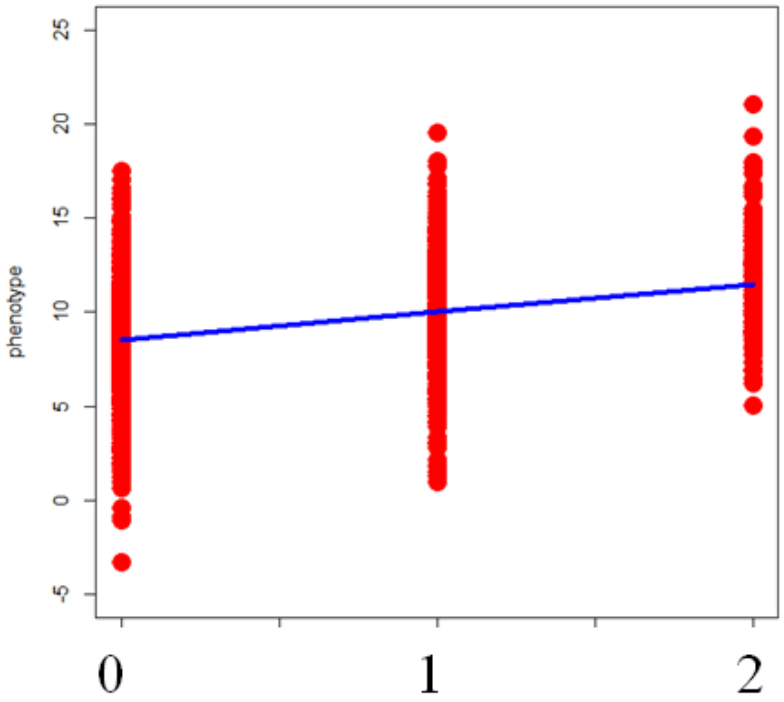
Dominance effects: $s^2_{Ph\_QTL(D)} = (2pqd)^2$



Q: what if **d** =0 and **a** =0?


$s^2_{Ph\_QTL(A)}$ and $s^2_{Ph\_QTL(D)}$
I understand,
but I don't understand
I think I might understand
or not...?
I know the feeling

Suppose we measure the QTL and the phenotype and regress X on QTL.The scatterplot of the data
(aa coded 0; Aa coded 1; AA coded 2 - call it $\mathbf{QTL_A}$).



we ask:

how much of the phenotypic variance is explained by the predictor (QTL_A)?

In the following slides we look at the regression lines only (not plotting the residuals – just to avoid clutter).

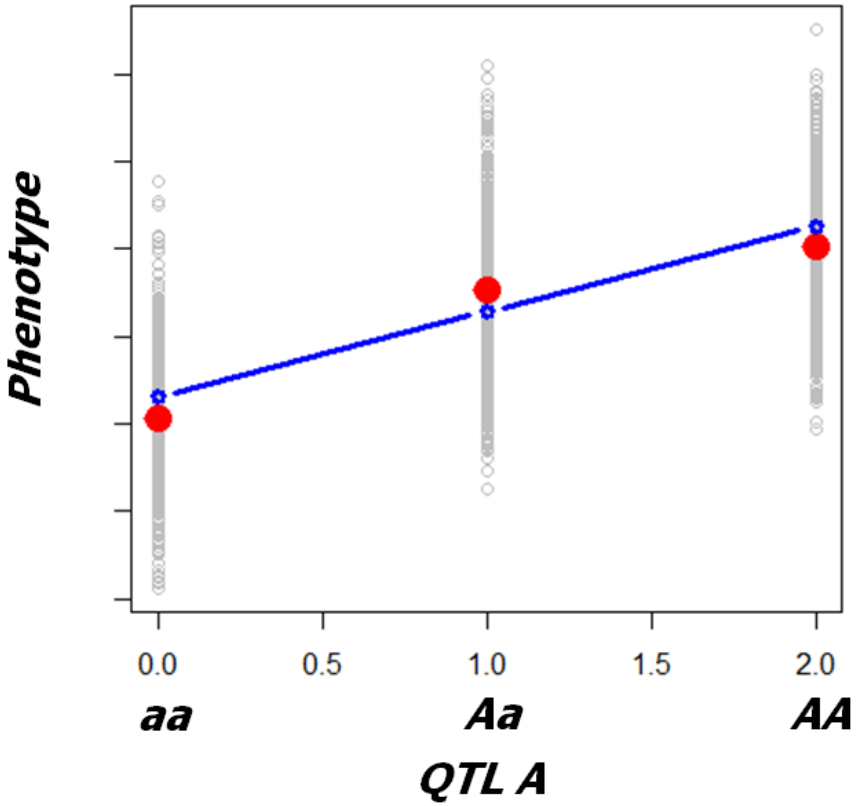
Linear regression model $y_i = \mathbf{a_0} + \mathbf{a_1}*x_i + e_i$

x = predictor (variable) ... here: QTL_A, values: aa (0), Aa (1) , AA (2)
y = dependent (variable) .... here: phenotype (ph)
e = residual (variable) ....
a_0 = intercept (parameter often denoted b_0)
a_1 = slope or regression coefficient (parameter often denoted b_1)

variance of y equals $a_1^2*s^2_x + s^2_e$
variance explained $a^2*s^2_x$
standard effect size: $R^2 = \{a^2 * s^2_x\} / \{a^2* s^2_x + s^2_e\}$

$y_{predicted} = a_0+a_1*x$ $e_{estimated} = y - y\ y_{predicted}$
$var(y_{predicted}) = a_1^2 *var(x)$ $var(e)$

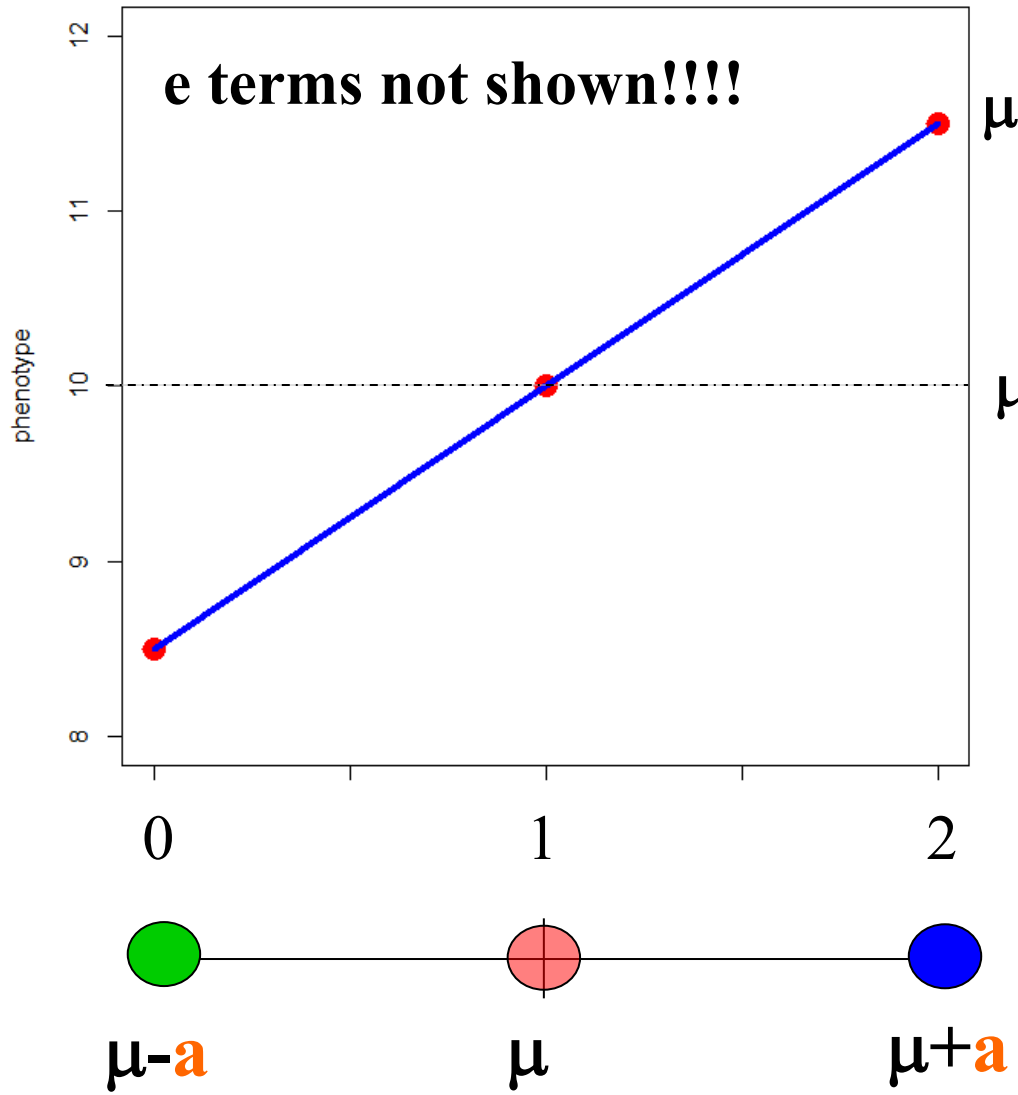
Linear regression model $pheno_i = \mathbf{a_0} + \mathbf{a_1}*QTL_{Ai} + e_i$



variance of pheno $a_1^2 * s^2_{QTL_A} + s^2_e$

variance explained $a_1^2 * s^2_{QTL_A}$

Warning!!! Next slides without residual (error) terms



regression model

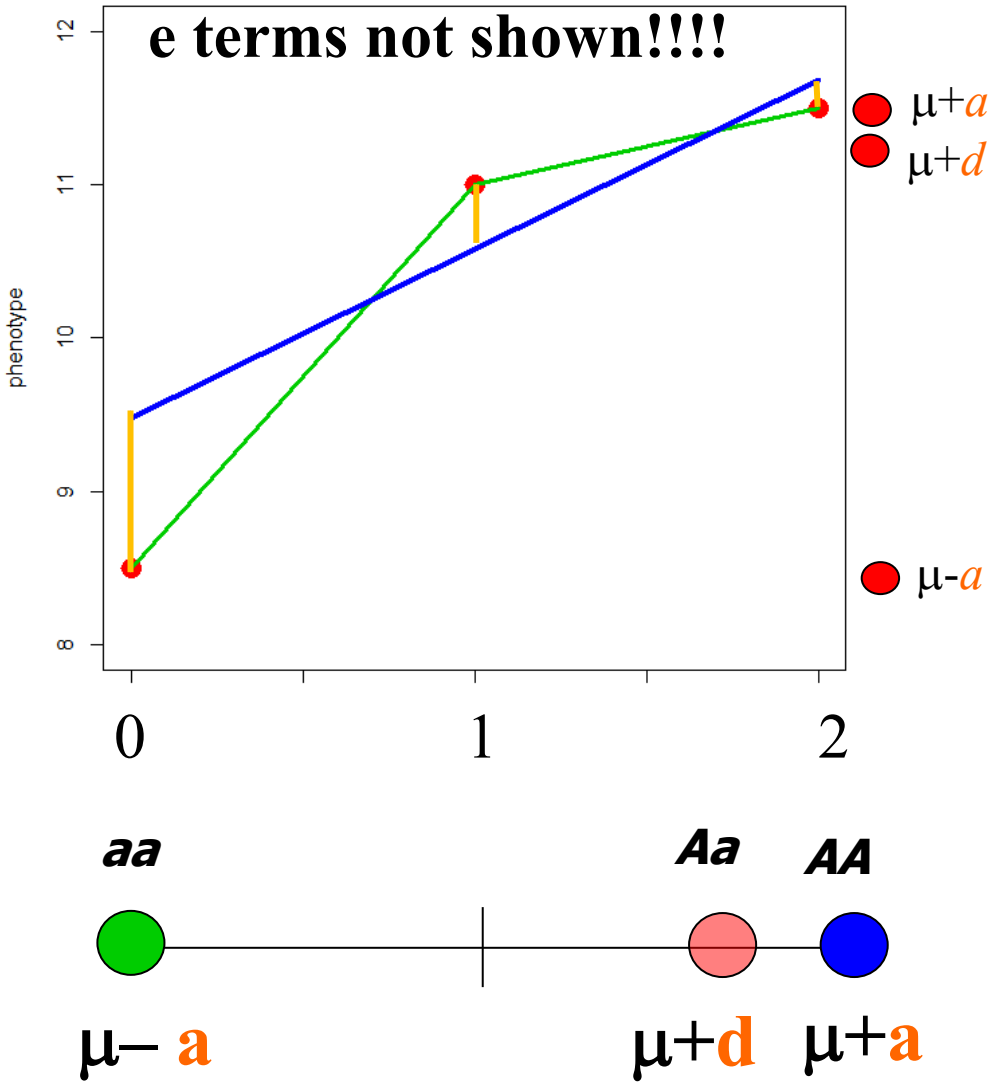
$ph_i = a_0 + a_1 * QTL_{Ai} + e_i$

$s^2_{Ph\_QTL(A)} = 2 * pq[a+(q-p)d]^2$

$s^2_{Ph\_QTL(A)} = a_1^2 * s^2_{QTL\,A}$

variance of pheno $a_1^2 * s^2_{QTL_A} + s^2_e = 2 * pq[a+(q-p)d]^2 + s^2_e$

variance explained $a_1^2 * s^2_{QTL_A} = \quad = 2 * pq[a+(q-p)d]^2$



Explained variance (blue line):

$s^2_{Ph\_QTL(A)} = 2 * pq[a+(q-p)d]^2$

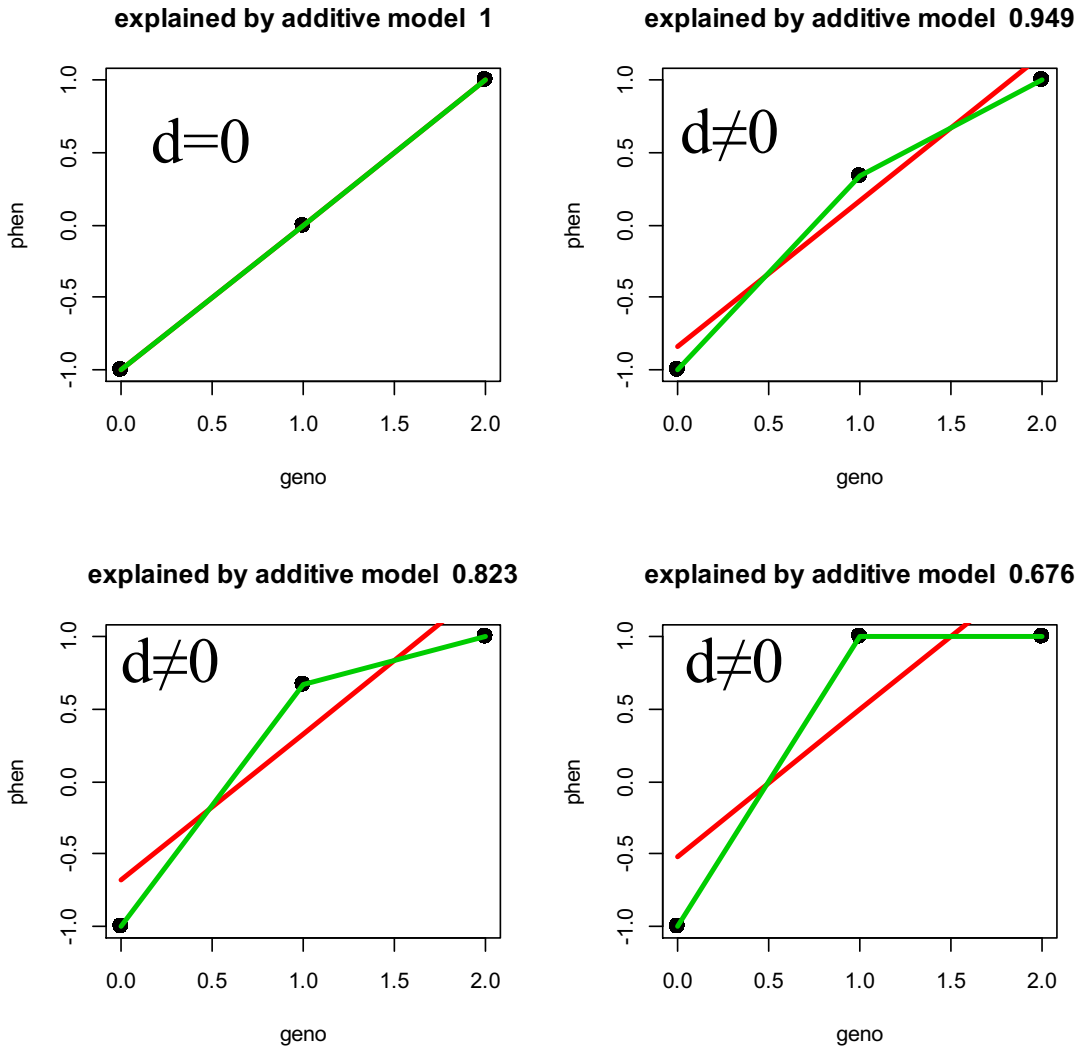
$s^2_{Ph\_QTL(A)} = a_1^2 * s^2_{QTL\ A}$

Not explained

$s^2_{Ph\_QTL(D)} = (2pqd)^2$

Important to note:

s^2_e includes $s^2_{Ph\_QTL(D)}$



$s^2_{Ph\_QTL(A)}$ *always greater than zero (given* $d \neq 0$ & $a>0$*)*

$s^2_{Ph\_QTL(D)}$ *can be zero (additive model* $d=0$*)*

What about the dominance variance? Can we estimate that?

regression model $ph_i = a_0 + a_1*QTL_{Ai} + d_1*QTL_{Di} + e_i$

genotype	QTL_A	QTL_D	p=.5
AA	2	4*p-2	0
Aa (aA)	1	2*p	1
aa	0	0	0

$$s^2_{Ph} = a_1^2*s^2_{QTL_A} + d_1^2*s^2_{QTL_D} + s^2_e$$

$$s^2_{Ph\_QTL(A)} = a_1^2*s^2_{QTL_A} = 2*pq[a+(q-p)d]^2$$
$$s^2_{Ph\_QTL(D)} = d_1^2*s^2_{QTL_D} = (2pqd)^2$$

Dominance deviation can **μ+d** (positive) or **μ-d** (negative)
Q: If we know the value of $s^2_{QTL_D}$ do we know the sign of the dominance deviation?

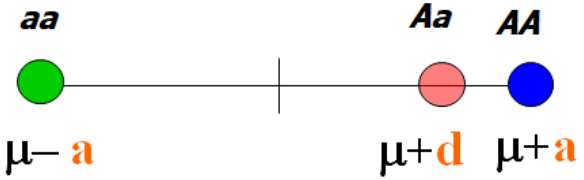
regression model $ph_i = a_0 + a_1*QTL_{Ai} + d_1*QTL_{Di} + e_i$

$$s^2_{Ph} = a_1^2*s^2_{QTL_A} + d_1^2*s^2_{QTL_D} + s^2_e$$

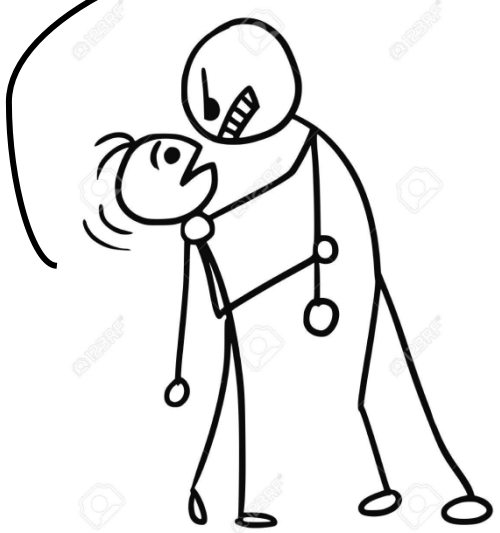
$2*pq[a+(q-p)d]^2$ $(2pqd)^2$

Dominance deviation can **μ+d** (positive) or **μ-d** (negative)

Q: If we know the value of $s^2_{Ph\_QTL(D)}$ do we know the sign of the dominance deviation?



I haven’t measured any QTLs!
What am I supposed to do?



Thank you!
Good question

Remember slide 13 ? Of course you do!

covariance

$$Cov(X,Y) = E(X-\mu_X)(Y-\mu_Y)$$

$$\sum_i (x_i-\mu_X)(y_i-\mu_Y) f(x_i,y_i)$$

correlation

$$Cor(X,Y) = Cov(X,Y) / \sqrt{[Var(X)*var(Y)]} =$$

$$= Cov(X,Y) / [stdev(X)*stdev(Y)]$$

Q: How does locus A-a contribute to the phenotypic covariance among family members?

A: Depends on the exact relationship

Biometrical model for single biallelic QTL

3. Contribution of the QTL to the Cov (X,Y) -

$$Cov(X,Y) = \sum_i (x_i - \mu_X)(y_i - \mu_Y) f(x_i, y_i)$$

person 1 (x_i)

person 2 (y_i)	AA (a-m)	Aa (d-m)	aa (-a-m)
AA (a-m)	(a-m)^2	(a-m) (d-m)	(a-m) (-a-m)
Aa (d-m)	(a-m) (d-m)	(d-m)^2	(d-m) (-a-m)
aa (-a-m)	(a-m) (-a-m)	(d-m) (-a-m)	(-a-m)^2

$m = a(p-q) + 2pqd$

Q: What about the $f(x_i, y_i)$?

Biometrical model for single biallelic

QTL

3A. Contribution of the QTL to the Cov (X,Y) – MZ twins

$$Cov(X,Y) = \sum_i (x_i - \mu_X)(y_i - \mu_Y) f(x_i, y_i)$$

	AA (a-m)	Aa (d-m)	aa (-a-m)
AA (a-m)	$p^2(a\text{-}m)^2$	0 (a-m) (d-m)	0 (a-m) (-a-m)
Aa (d-m)	0 (a-m) (d-m)	$2pq(d\text{-}m)^2$	0 (d-m) (-a-m)
aa (-a-m)	0 (a-m) (-a-m)	0 (d-m) (-a-m)	$q^2(-a\text{-}m)^2$

$$\begin{aligned} Cov(X_i, Y_j) &= (a\text{-}m)^2p^2 + (d\text{-}m)^2 2pq + (-a\text{-}m)^2q^2 \\ &= 2pq[a+(q-p)d]^2 + (2pqd)^2 \\ &= s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)} \end{aligned}$$

Biometrical model for single biallelic

QTL

3B. Contribution of the QTL to the Cov (X,Y) – Parent-Offspring

$$Cov(X,Y) = \sum_i (x_i - \mu_X)(y_i - \mu_Y) f(x_i, y_i)$$

		parent		
		AA (a-m)	**Aa** (d-m)	**aa** (-a-m)
child	**AA** (a-m)	$p^3(a\text{-}m)^2$	p^2q (a-m) (d-m)	0 (a-m) (-a-m)
	Aa (d-m)	p^2q (a-m) (d-m)	pq $(d\text{-}m)^2$	pq^2 (d-m) (-a-m)
	aa (-a-m)	0 (a-m) (-a-m)	pq^2 (d-m) (-a-m)	$q^3(-a\text{-}m)^2$

given an ***AA*** parent, an ***AA*** offspring can come from either ***AA*** x ***AA*** or ***AA*** x ***Aa*** parental random mating types

AA x ***AA*** will occur $p^2 \times p^2 = p^4$
and have ***AA*** offspring Prob(**AA**)=1

AA x ***Aa*** will occur $p^2 \times 2pq = 2p^3q$
and have ***AA*** offspring Prob(**AA**)=0.5
and have ***Aa*** offspring Prob(**Aa**)=0.5

AA x ***aa*** Not relevant (offspring Aa)

Therefore, P(***AA*** parent & ***AA*** offspring) $= p^4 + .5*2*p^3q$
$= p^3(p+q)$
$= p^3$

So can be complicated, but can also be simple ….

Parent

Offspring	AA (a-m)	Aa (d-m)	aa (-a-m)
AA (a-m)	$p^3(a\text{-}m)^2$	$p^2q(a\text{-}m)\ (d\text{-}m)$	$0\ (a\text{-}m)\ (-a\text{-}m)$
Aa (d-m)	$p^2q(a\text{-}m)\ (d\text{-}m)$	$pq\ (d\text{-}m)^2$	$pq^2\ (d\text{-}m)\ (-a\text{-}m)$
aa (-a-m)	$0\ (a\text{-}m)\ (-a\text{-}m)$	$pq^2\ (d\text{-}m)\ (-a\text{-}m)$	$q^3(-a\text{-}m)^2$

why zero probability {0}?

Biometrical model for single biallelic QTL

3B. Contribution of the QTL to the Cov (*X*,*Y*) – Parent-Offspring

Offspring (Y) \ Parent (X)	**AA** (a-m)	**Aa** (d-m)	**aa** (-a-m)
AA (a-m)	$p^3(a-m)^2$	p^2q (a-m) (d-m)	0 (a-m) (-a-m)
Aa (d-m)	p^2q (a-m) (d-m)	pq $(d-m)^2$	pq^2 (d-m) (-a-m)
aa (-a-m)	0 (a-m) (-a-m)	pq^2 (d-m) (-a-m)	$q^3(-a-m)^2$

$$\text{Cov}(X_i, Y_i) = (a-m)^2p^3 + \ldots + (-a-m)^2q^3$$

$$= pq[a+(q-p)d]^2 = \tfrac{1}{2}\, s^2_{QTL(A)}$$

Biometrical model for single biallelic QTL

3C. Contribution of the QTL to the Cov (X,Y) – Unrelated individuc

		p^2 AA (a-m)	$2pq$ Aa (d-m)	q^2 aa (-a-m)
p^2	AA (a-m)	$p^4(a\text{-}m)^2$	$2p^3q(a\text{-}m)(d\text{-}m)$	$p^2q^2(a\text{-}m)(-a\text{-}m)$
$2pq$	Aa (d-m)	$2p^3q(a\text{-}m)(d\text{-}m)$	$4p^2q^2(d\text{-}m)^2$	$2pq^3(d\text{-}m)(-a\text{-}m)$
q^2	aa (-a-m)	$p^2q^2(a\text{-}m)(-a\text{-}m)$	$2pq^3(d\text{-}m)(-a\text{-}m)$	$q^4(-a\text{-}m)^2$

$$\text{Cov}(X_i, Y_j) = (a\text{-}m)^2p^4 + \ldots + (-a\text{-}m)^2q^4$$

$$= 0$$

Note if mating is random - the spousal correlation is zero.
Mother and father are **Unrelated individuals** !

Follow same method for full sibs and DZ twins

Derive genotype frequences ....

s1	s2	eff	eff	freq	frequency (p(A)=p, p(a)=q=1-p)
AA	AA	a	a	r1	p**4+p**3*q+p**2*q**2/4
aa	aa	-a	-a	r2	p**2*q**2/4+p*q**3+q**4
Aa	Aa	d	d	r3	p**3*q+3*p**2*q**2+p*q**3
AA	Aa	a	d	r4	p**3*q+p**2*q**2/2
Aa	AA	d	a	r4	p**3*q+p**2*q**2/2
Aa	aa	d	-a	r5	p**2*q**2/2+p*q**3
aa	Aa	-a	d	r5	p**2*q**2/2+p*q**3
AA	aa	a	-a	r6	p**2*q**2/4
aa	AA	-a	a	r6	p**2*q**2/4

Biometrical model for single biallelic QTL

3B. Contribution of the QTL to the Cov (X,Y) – DZ twins

DZ twin 2 \ DZ twin 1	AA (a-m)	Aa (d-m)	aa (-a-m)
AA (a-m)	r1 $(a-m)^2$	r4 (a-m) (d-m)	r6 (a-m) (-a-m)
Aa (d-m)	r4 (a-m) (d-m)	r2 $(d-m)^2$	r5 (d-m) (-a-m)
aa (-a-m)	r6 (a-m) (-a-m)	r5 (d-m) (-a-m)	r3 $(-a-m)^2$

$$\text{Cov}(X_i, X_j) = (a-m)^2 r1 + \ldots + (-a-m)^2 r3$$

$$= \tfrac{1}{2}\, 2pq[a+(q-p)d]^2 + \tfrac{1}{4}(2pqd)^2 = \tfrac{1}{2}\, s^2_{QTL(A)} + \tfrac{1}{4}\, s^2_{QTL(D)}$$

Genetic variance *shared* contributes to the phenotypic covariance

	$s^2_{Ph\_QTL(A)}$	$s^2_{Ph\_QTL(D)}$
Unrelateds	0	0
Parent - child	½	0
full (DZ) sibs	½	¼
MZ twins	1	1

Q: So how does this help to estimate $s^2_{Ph\_QTL(A)}$ & $s^2_{Ph\_QTL(D)}$?
A: Come back this afternoon!

Covariance matrix (2x2) in MZ twins

	MZ1	MZ2
MZ1	s^2_{Ph1} (variance)	$s^2_{Ph1,Ph2}$ (covariance)
MZ2	$s^2_{Ph1,Ph2}$ (covariance)	s^2_{Ph2} (variance)

	MZ1	MZ2
MZ1	$s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)} + s^2_{rest}$	$s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)}$
MZ2	$s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)}$	$s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)} + s^2_{rest}$

	DZ1	DZ2
DZ1	s^2_{Ph1}	$s^2_{Ph1,Ph2}$
DZ1	$s^2_{Ph1,Ph2}$	s^2_{Ph2}

	DZ1	DZ2
DZ1	$s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)} + s^2_{rest}$	$\frac{1}{2} s^2_{Ph\_QTL(A)} + \frac{1}{4} s^2_{Ph\_QTL(D)}$
DZ1	$\frac{1}{2} s^2_{Ph\_QTL(A)} + \frac{1}{4} s^2_{Ph\_QTL(D)}$	$s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)} + s^2_{rest}$

$$s^2_{Ph} = 2pq[a+(q-p)d]^2 + (2pqd)^2 + \text{residual variance}$$

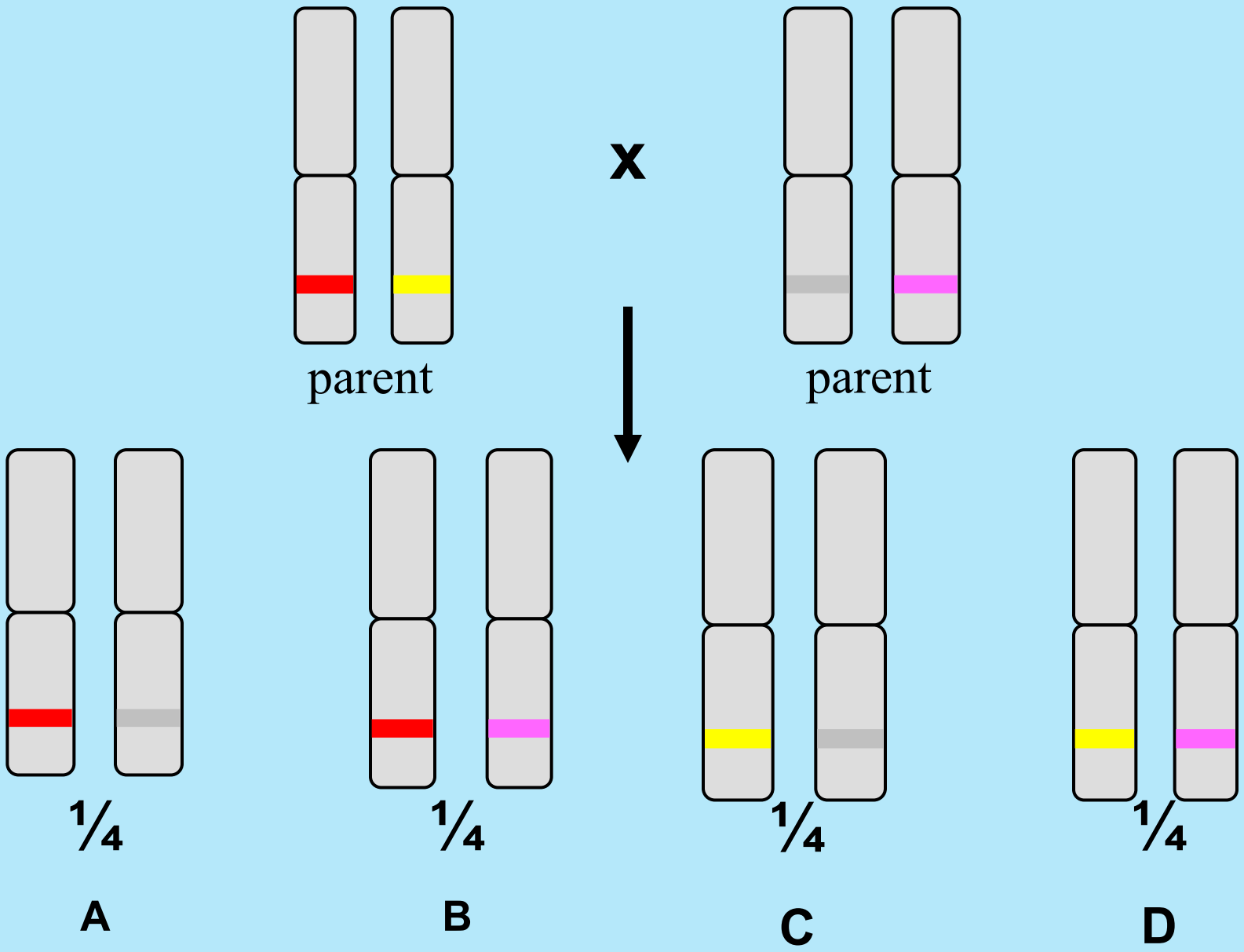
1: Genetic variance is due to individual differences in genotype
2: Genotype depends on alleles
3: Alleles are passed on from parents to offspring
4: Relatives share genetic variance, because they share alleles
5: Shared genetic variance contributes to phenotypic covariance

Offspring (DZ twins) share genetic variance, because they share alleles
Parents and Offspring share genetic variance, because they share alleles
Monozygotic (identical) twins share genetic variance, because they share alleles

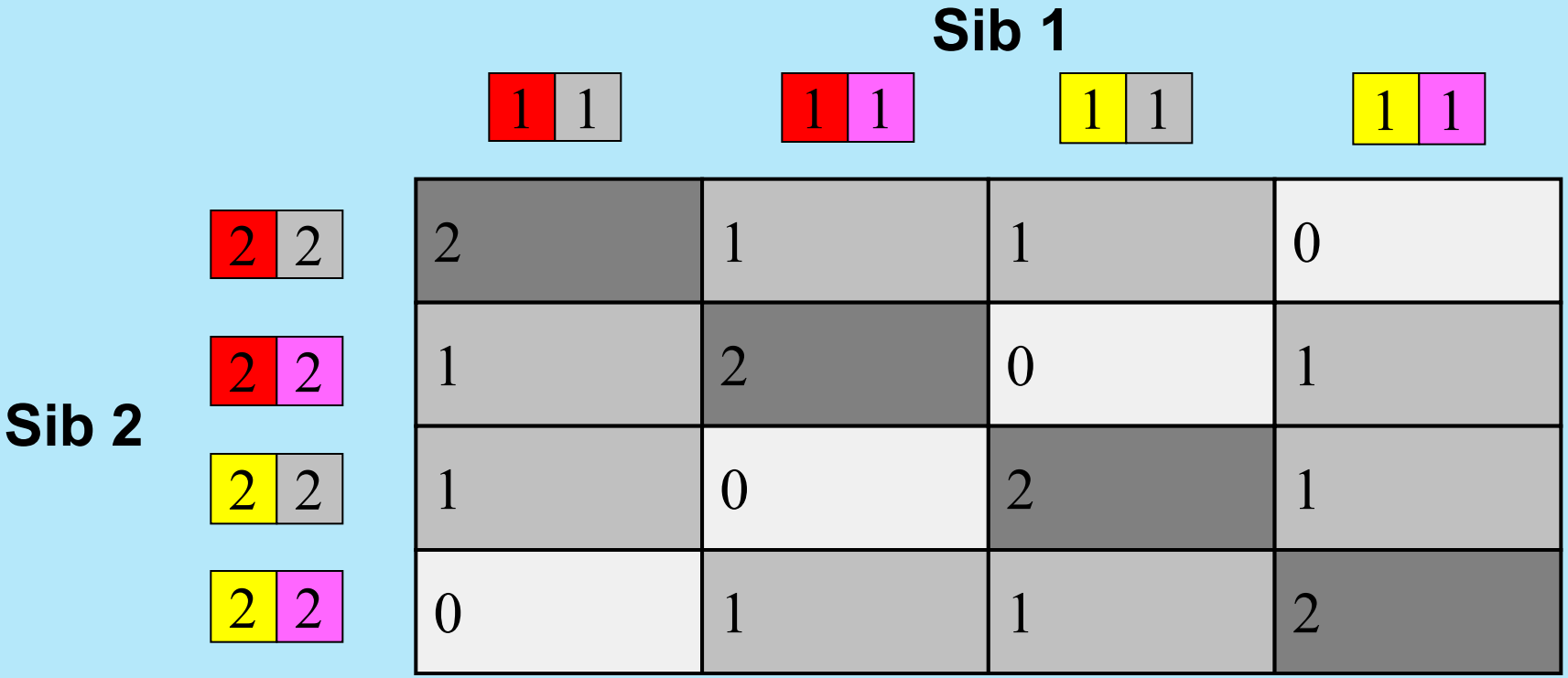
If I know the proportion of alleles they share at locus,
I'll will know the contribution of the locus to the phenotypic covariance ...

Concept of allele sharing IBD .... **I**DENTICALLY **B**Y **D**ESCENT

Segregation and identity-by-descent (IBD) in sibpairs



IDENTITY BY DESCENT (IBD) DZs

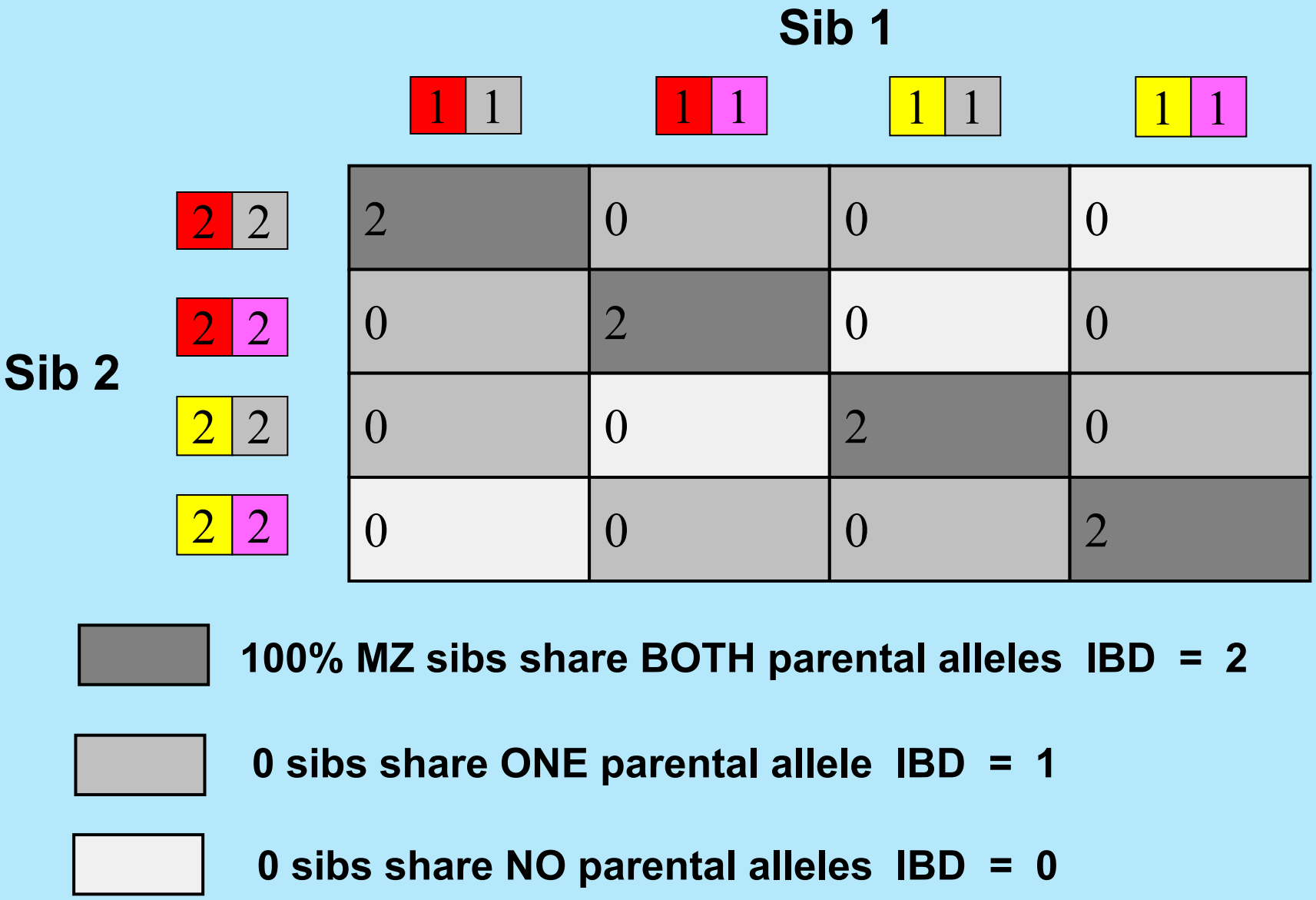


4/16 = 1/4 sibs share BOTH parental alleles IBD = 2

8/16 = 1/2 sibs share ONE parental allele IBD = 1

4/16 = 1/4 sibs share NO parental alleles IBD = 0

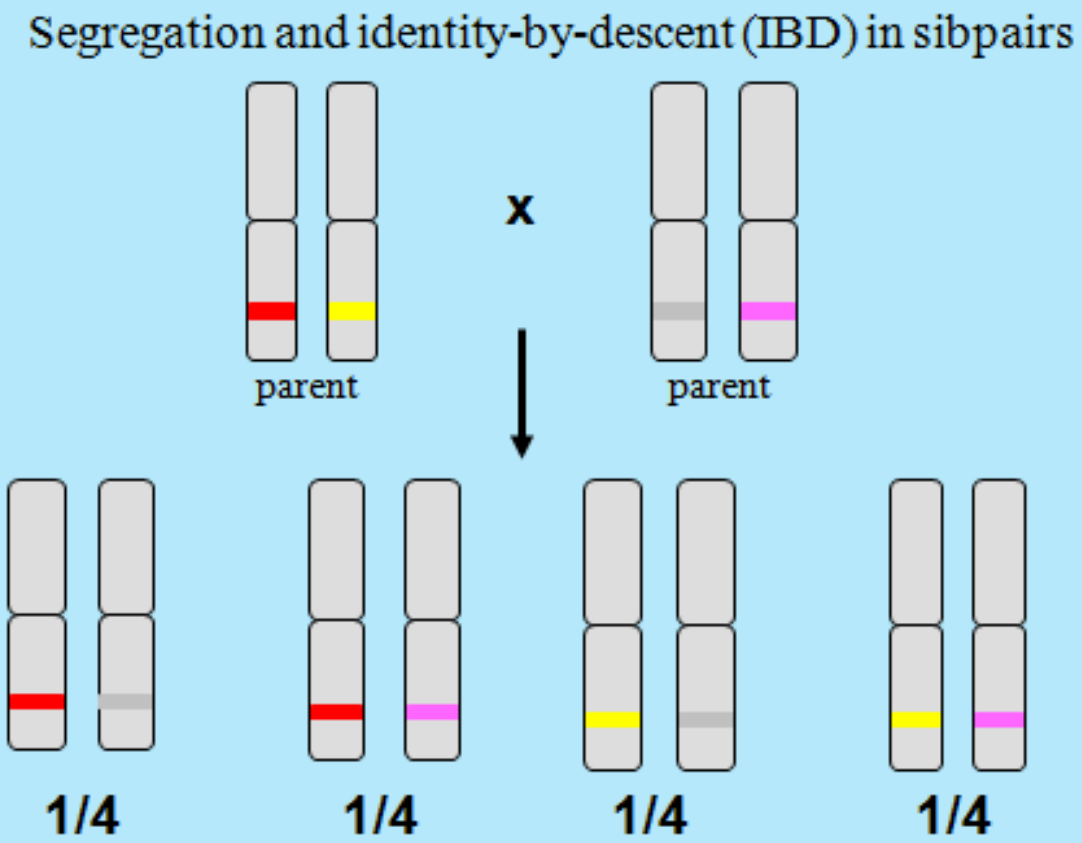
IDENTITY BY DESCENT (IBD) MZs



100% MZ sibs share BOTH parental alleles IBD = 2

0 sibs share ONE parental allele IBD = 1

0 sibs share NO parental alleles IBD = 0



What about parent offsping?
many alleles do they share IBD?
(decending from the grandparent)

(2 alleles IBD)	(1 allele IBD)	(0 alleles IBD)
MZ twins	Parent- Offspring (P-O)	Unrelateds
Cov(MZ)	Cov(P-O)	Cov(Unrelateds)
$s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)}$	$\frac{1}{2}\, s^2_{Ph\_QTL(A)}$	0

slide 43

slide 47

slide 43

Note: spouses given random mating

(2 alleles IBD)	(1 allele IBD)	(0 alleles IBD)
MZ twins	Parent- Offspring (P-O)	Unrelateds
Cov(MZ)	Cov(P-O)	Cov(Unrelateds)
.25 DZ twins	**.50 DZ twins**	**.25 DZ twins**
$s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)}$	$\frac{1}{2}\, s^2_{Ph\_QTL(A)}$	0

average DZ genetic variance sharing (based on IBD):

$.25*(s^2_{Ph\_QTL(A)} + s^2_{Ph\_QTL(D)}) + .50*(\frac{1}{2}s^2_{Ph\_QTL(A)}) + .25*\ 0 =$

$.5*s^2_{Ph\_QTL(A)} + .25*\ s^2_{Ph\_QTL(D)}$ ← slide 50

$s^2_{Ph\_QTL_A} = 2pq[a+(q-p)d]^2$ $s^2_{Ph\_QTL_D} = (2pqd)^2$

IBD=0	**0**	**0**	Unrelated
IBD=1	**½**	**0**	Parent - Offspring
IBD=2	**1**	**1**	MZ twins
IBD=0	**0**	**0**	25% (¼) DZ twins
IBD=1	**½**	**0**	50% (½) DZ twins
IBD=2	**1**	**1**	25% (¼) DZ twins
average	0*¼+ ½ * ½ +1*¼ = ½	0*¼+0* ½ +1*¼ = ¼	

proportion of alleles shared IBD

probability of sharing 2 alleles IBD

Q: Why do twins have to be IBD=2 to shared dominance variance? (prob(IBD=2) = 1)?

A: Because similaries due to dominance effects are related to genotype not individual alleles. You have to have the same genotype to shared dominance variance.

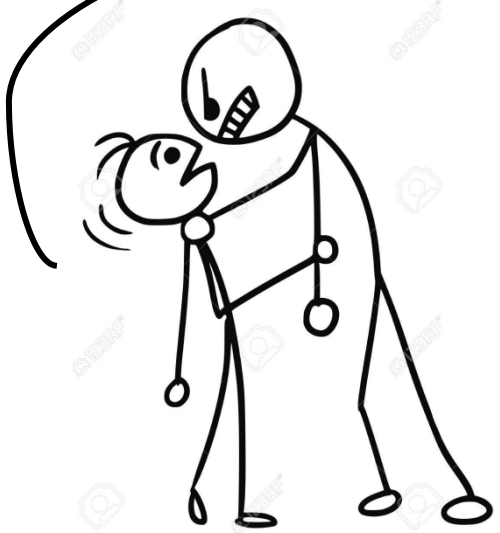
Q: Why does the (average) proportion of alleles shared IBD reflect shared additive genetic variance?

A: Because similaries due to additive effect are related to individual alleles. Sharing an allele implies sharing additive genetic variance.

Q: If I know MZ twin are IBD=2, do I know what actual alleles they have?

NO: IBD is about sharing alleles, but if not says nothing about the actual identity of the alleles. However, if relatives are IBD 2, you so know that they have the same alleles (AA and AA, Aa and Aa, or aa and aa).

But all this was about 1 QTL!
What if there are >1 or > 100?



Thank you!
Good question !

Linear regression model N QTLs (N > 1... N>1000)

$$\text{pheno}_i = \mathbf{a_0} + \mathbf{a_1}*\text{QTL}_{A1i} + \mathbf{a_2}*\text{QTL}_{A2i} + \ldots + \mathbf{a_N}*\text{QTL}_{ANi} + \mathbf{d_1}*\text{QTL}_{D1i} + \mathbf{a_2}*\text{QTL}_{D2i} + \ldots + \mathbf{d_N}*\text{QTL}_{DNi} + e_i$$

$$s^2_{Ph\_QTL(A)} = 2*p_1q_1[a_1+(q_1-p_1)d_1]^2 + 2*p_1q_1[a_1+(q_1-p_1)d_1]^2 + \ldots + 2*p_Nq_N[a_N+(q_N-p_N)d_N]^2$$

$$s^2_{Ph\_QTL(A)} = a_1^2*s^2_{QTLA1}+a_2^2*s^2_{QTLA2}+\ldots+a_N^2*s^2_{QTLAN}$$

$$s^2_{Ph\_QTL(D)} = (2p_1q_1d_1)^2 + (2p_2q_2d_2)^2 + \ldots + (2p_Nq_Nd_N)^2$$

$$s^2_{Ph\_QTL(D)} = d_1^2*s^2_{QTLD1} + d_2^2*s^2_{QTLD2} + \ldots + d_N^2*s^2_{QTLDN}$$

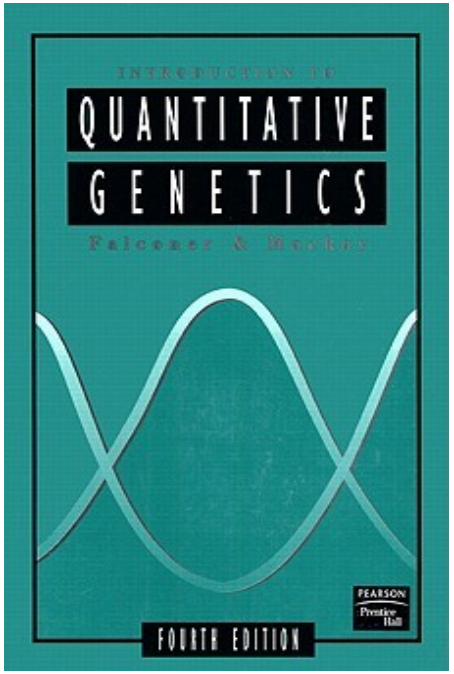
Covariance matrix (2x2) in DZ and MZ twins

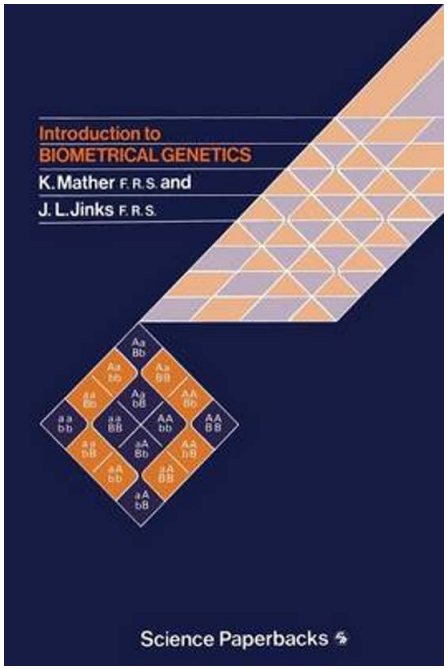
	MZ1	MZ2
MZ1	$s^2_A + s^2_D + s^2_E$	$s^2_A + s^2_D$
MZ2	$s^2_A + s^2_D$	$s^2_A + s^2_D + s^2_E$

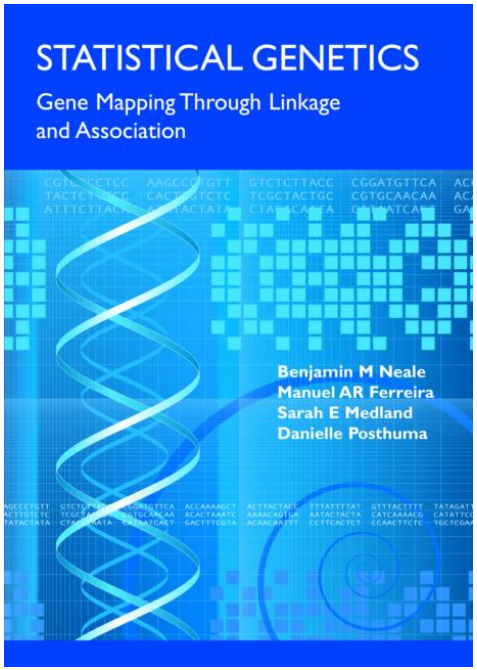
	DZ1	DZ2
DZ1	$s^2_A + s^2_D + s^2_E$	$\frac{1}{2}s^2_A + \frac{1}{4}s^2_D$
DZ2	$\frac{1}{2}s^2_A + \frac{1}{4}s^2_D$	$s^2_A + s^2_D + s^2_E$

Point of departure (more or less) for later on

Slide acknowledgement: Manuel Ferreira, Pak Sham, Shaun Purcell, Sarah Medland, and Sophie van der Sluis







Numerical (toy) example.

Suppose a phenotype subject to the influence of one QTL and environmental influences.

You observe the phenotype and the QTL in 500 individuals

I observe the phenotype S in 250 MZ and 250 DZ twin pairs

0 (aa) 1 (AA) 2 (AA)

0.236 (q^2) 0.526 ($2pq$) 0.238 (p^2)

variance of the phenotype s^2_{Ph} = 1.520

$\mathbf{a_0} + \mathbf{a_1}*\mathrm{QTL}_{Ai} + \mathrm{e}_i$

a_0 -0.561

a_1 1.111

Multiple R-squared: 0.386

$\mathbf{a_0} + \mathbf{a_1}*\mathrm{QTL}_{Ai} + \mathbf{d_1}*\mathrm{QTL}_{Di} + \mathrm{e}_i$

a_0 -1.10449

a_1 1.114

d_1 1.028

Multiple R-squared: 0.560

0.386 * 1.520 = 0.586 $\Longrightarrow$ $s^2_{Ph\_QTL_A} = 2pq[a+(q-p)d]^2$

(0.560-0.386)*1.520

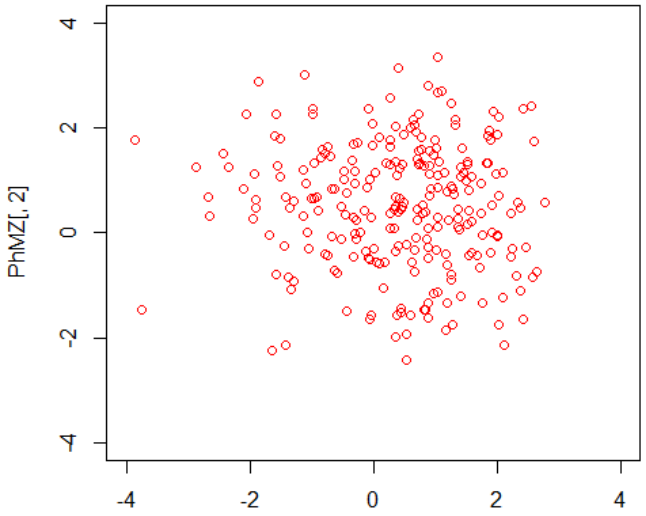
= 0.174*1.520 = 0.264 $\Longrightarrow$ $s^2_{Ph\_QTL_D} = (2pqd)^2$

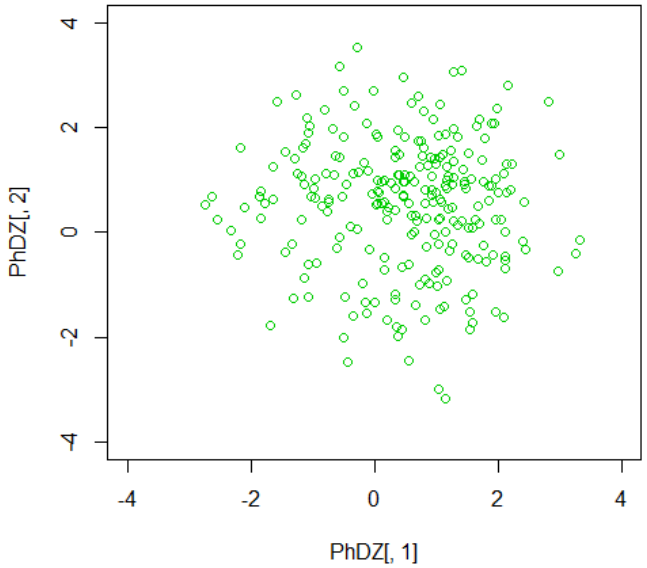
```
cov(PhMZ) = .525
          [,1]  [,2]
[1,] 1.466 0.736
[2,] 0.736 1.343

cov(PhDZ) = .192
          [,1]  [,2]
[1,] 1.559 0.311
[2,] 0.311 1.682
```

$0.736 = S^2_{Ph\_QTL_A} + S^2_{Ph\_QTL_D}$

$0.311 = \frac{1}{2} S^2_{Ph\_QTL_A} + \frac{1}{4} S^2_{Ph\_QTL_D}$



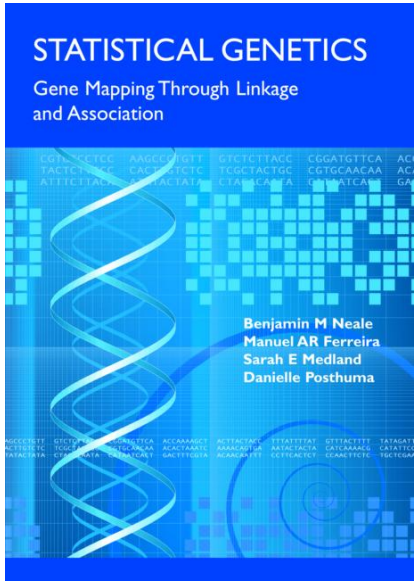


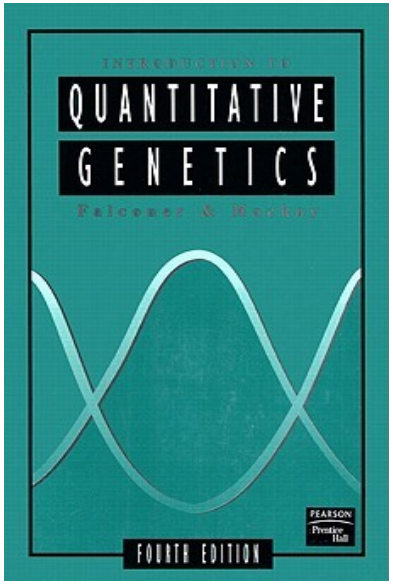
regression model vs biometric model

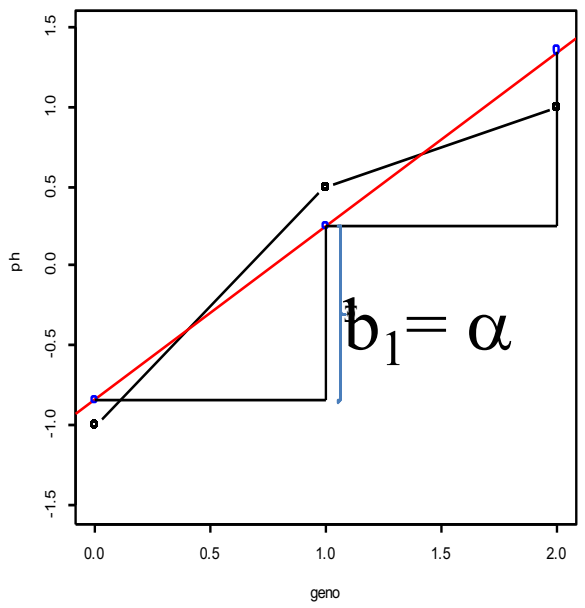
regression parameter a (henceforth b_1)

=

average effect of allele substitution







predicted values

b_0+b_1*0 (aa)

b_0+b_1*1 (Aa or aA)

b_0+b_1*2 (AA)

difference in regression model

$b_0+b_1*1 - (b_0+b_1*0) =$

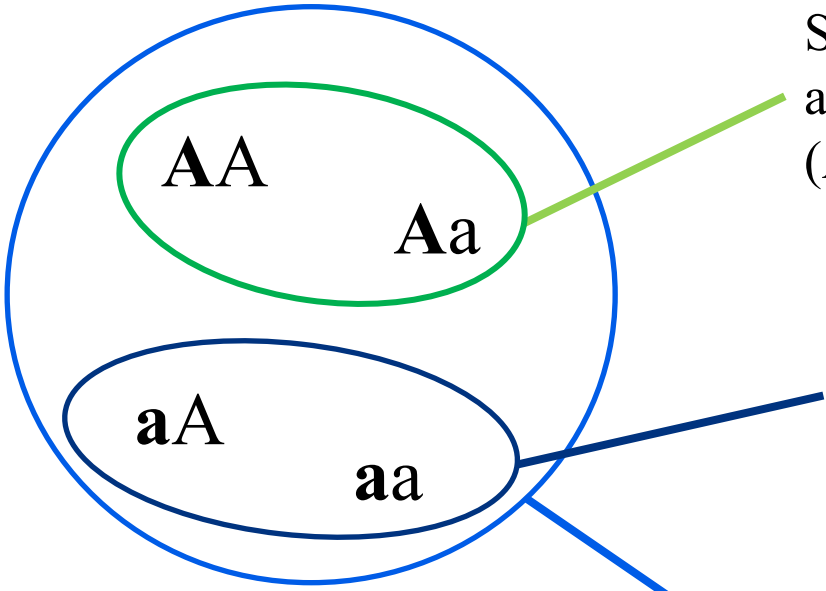
$b_0+b_1*2 - (b_0+b_1*1) = b_1$

b_1 is the average effect of substituting A for a (or vice versa)

The parameter b_1 in the regression model corresponds to a specific parameter in the biometric model, called α

Now: derive α from the biometric model.

α is the average effect (on the phenotype) of substituting allele A for allele a - how to derive this?

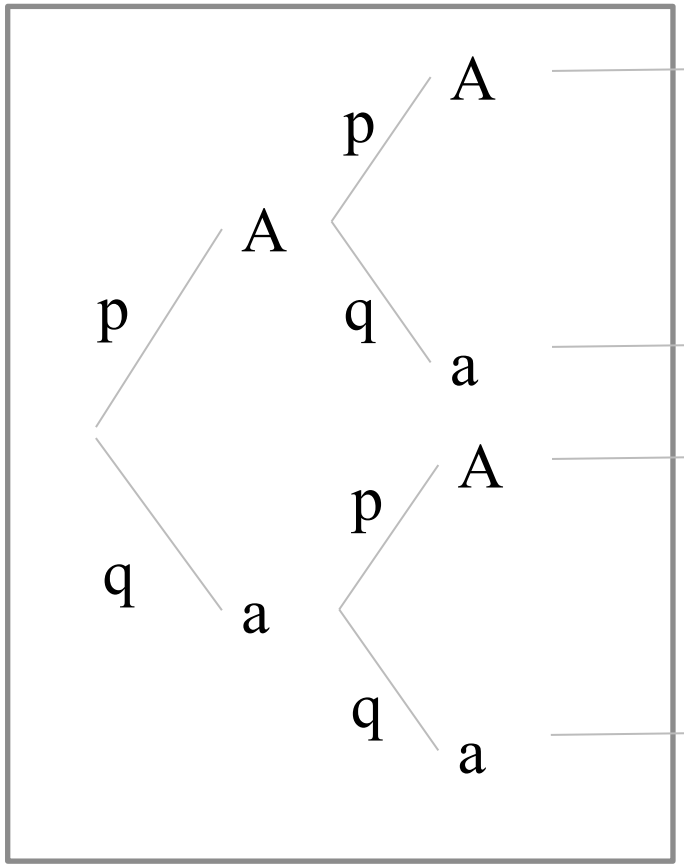


Subpopulation of individual with first allele **A**
(AA and Aa).

Subpopulation of individual with first allele **a**
(aA and aa).

Population of all individuals
(AA, Aa, aA, aa)

Population of all individuals (HWE)



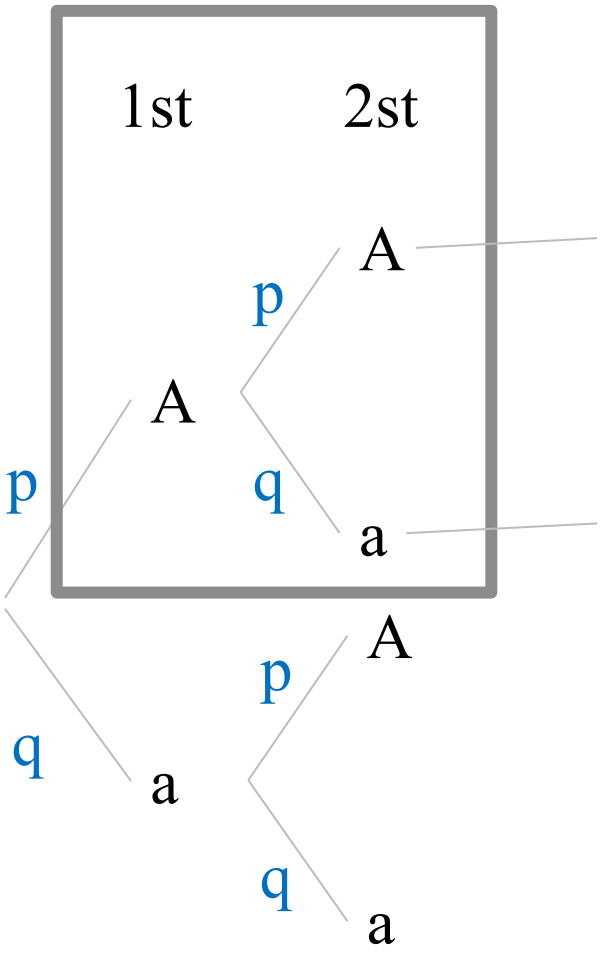
genotype AA; freq = (p*p); effect = a

genotype Aa; freq = (p*q); effect = d

genotype aA; freq = (q*p); effect = d

genotype aa; freq = (q*q); effect = -a

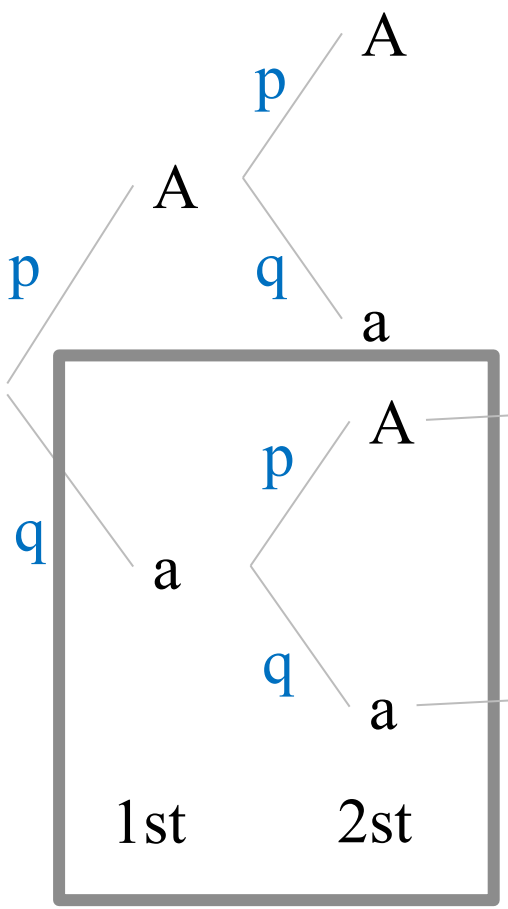
Subpopulation of individual with first allele **A**



conditional mean

α_1 = mean(1st allele=A) = p*a + q*d

Subpopulation of individual with first allele $\mathbf{A_2}$



aA (effect d, freq = p)

aa (effect -a, freq = q)

conditional mean (1st allele=a)

α_2 = mean(1st allele=a) = p*d + q*-a

average effect of allele substitution $\alpha = a + d(q-p)$

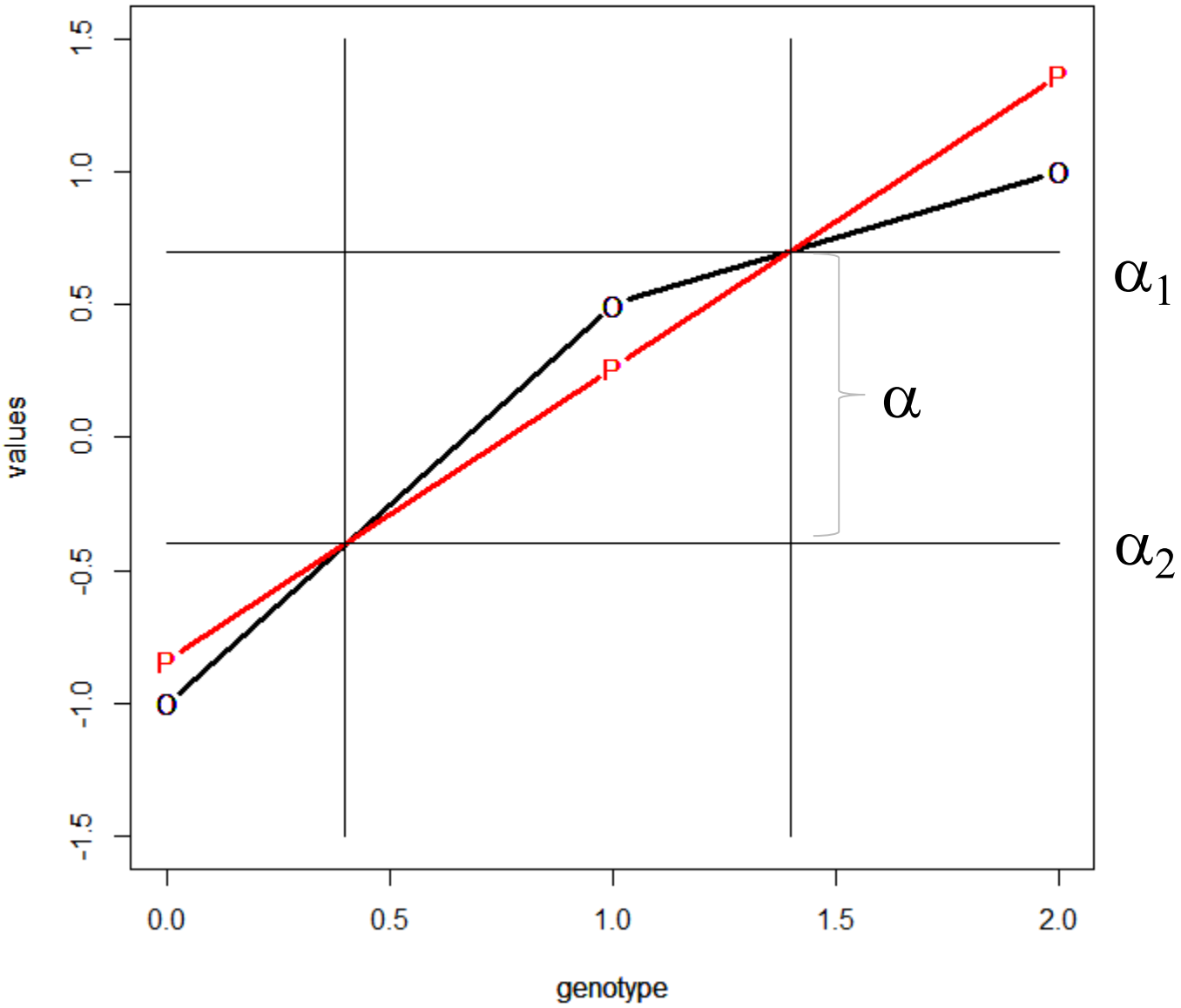
conditional mean α_1 = mean(1st=A) = (p*a + q*d)
conditional mean (1st=a) α_2 = mean(1st=a) = (p*d + q*-a)

difference α = average effect of allele substitution
$\alpha = \alpha_1 - \alpha_2$ = (p*a + q*d) - (p*d+q*-a) =
pa +qd -pd +qa =
pa +qa - pd + qd =
(p+q)a +d(q-p) = a + d(q-p)

b_1 is the average effect of substituting A for a (or vice versa)

$b_1 = \alpha = (a + d(q-p))$

parameter α derived from the biometric model



α defined in the regression model (b_1) and in the biometric model (α)

$$b_1 = \alpha = (a + d(q-p))$$