Biometrical Genetics

2014 International Workshop on Statistical Genetic Methods

for Human Complex Traits

Boulder, CO.

Lindon Eaves,

VIPBG, Richmond VA.

March 2014

Biometrical Genetics

How do genes contribute to statistics (e.g. means, variances, skewness, kurtosis)?

Some Literature:

Mather K (1949) *Biometrical Genetics: the Study of Continuous Variation.* London UK: Methuen.

Mather K, Jinks JL (1982) *Biometrical Genetics: the Study of Continuous Variation (3rd Ed.).* London UK: Chapman Hall.

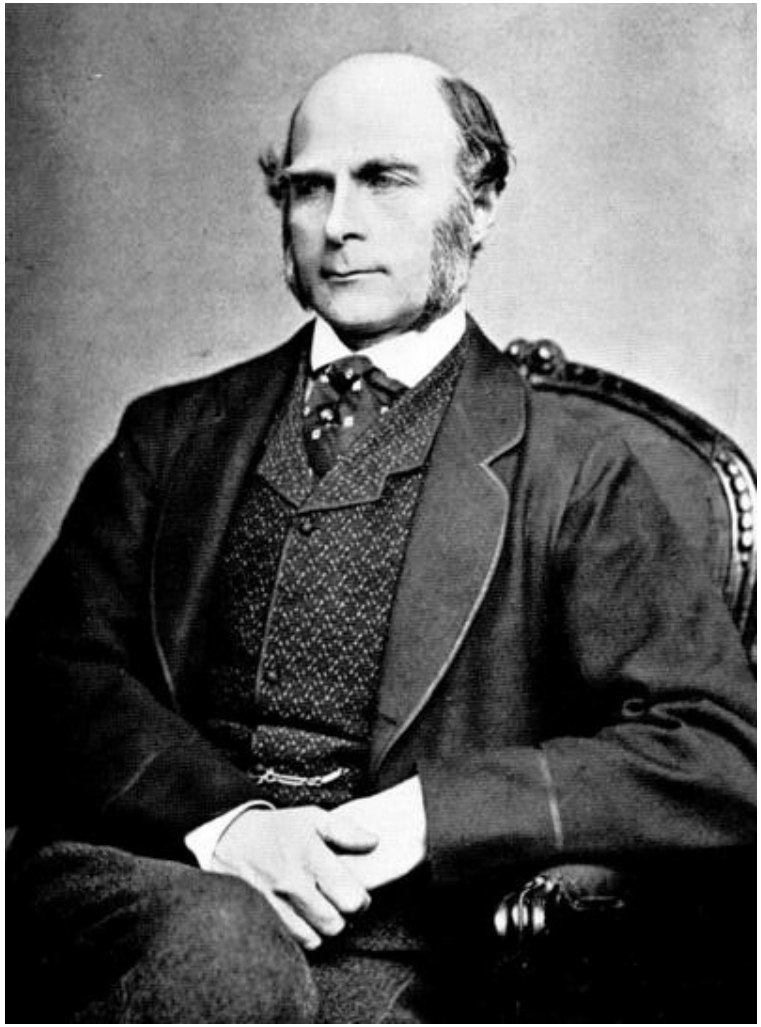
Jinks JL, Fulker DW (1970): Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychol Bull* 73(5):311-349.

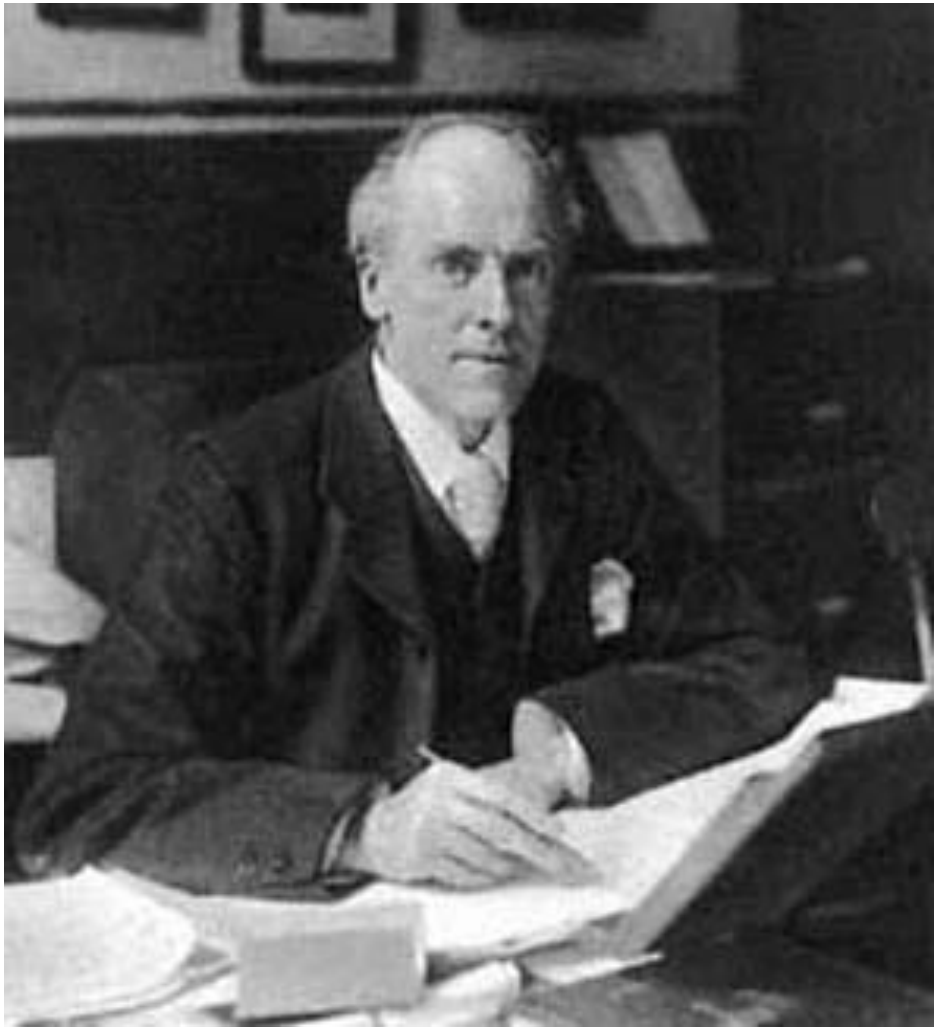
Kearsy MJ, Pooni HS (1996) *The Genetic Analysis of Quantitative Traits*. London UK: Chapman Hall.

Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*, 4th Ed. Harlow, UK: Addison Wesley Longman.

Neale MC, Cardon LR (1992). *Methodology for Genetic Studies of Twins and Families*. Ch 3. Dordrecht: Kluwer Academic Publisher. (See revised ed. Neale and Maes, pdf on VIPBG website)

Requires synthesis of two intellectual traditions…..





“Mendelian” Crosses with Quantitative Traits

Mendelian Basis of Continuous Variation? Experimental Breeding Experiments

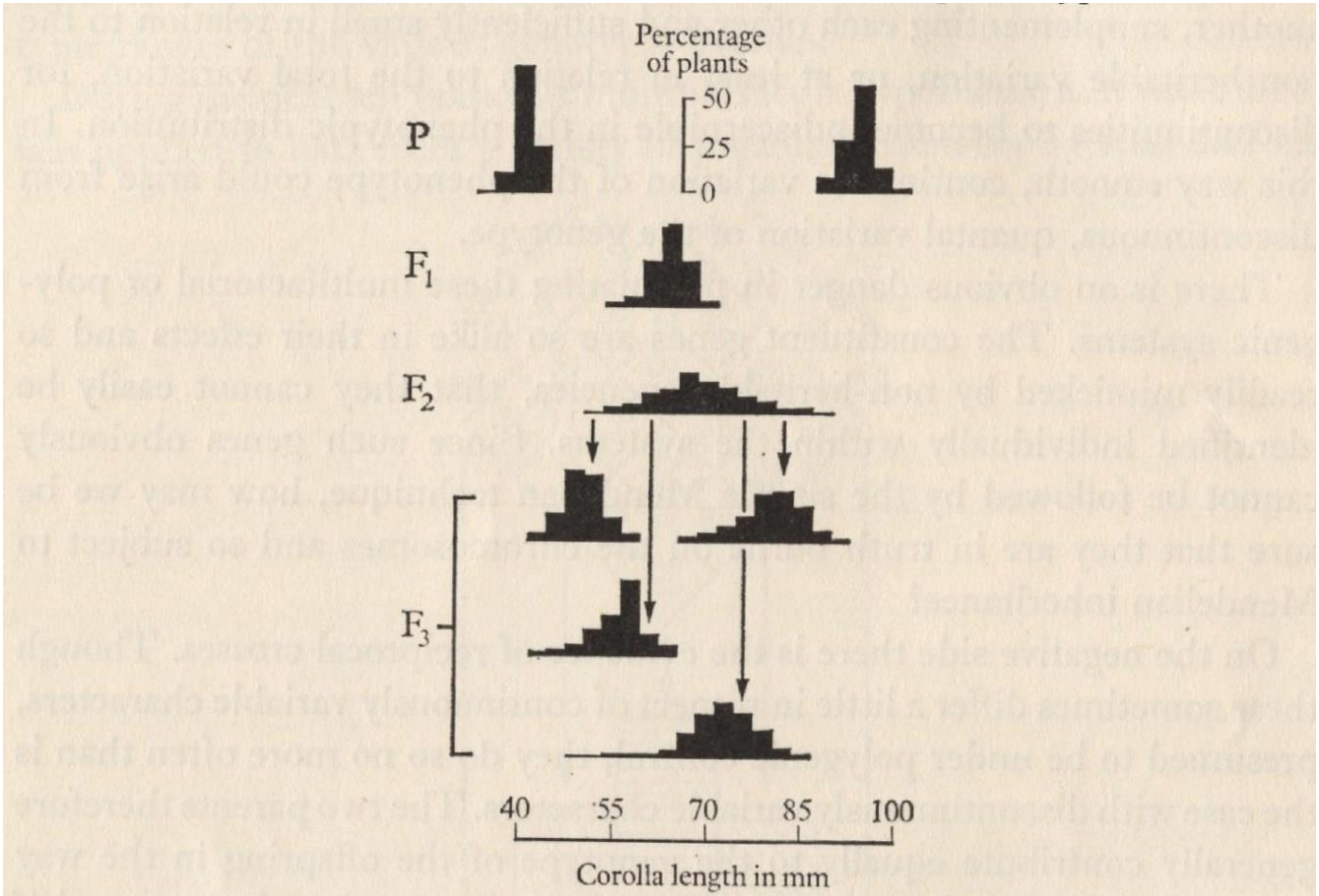
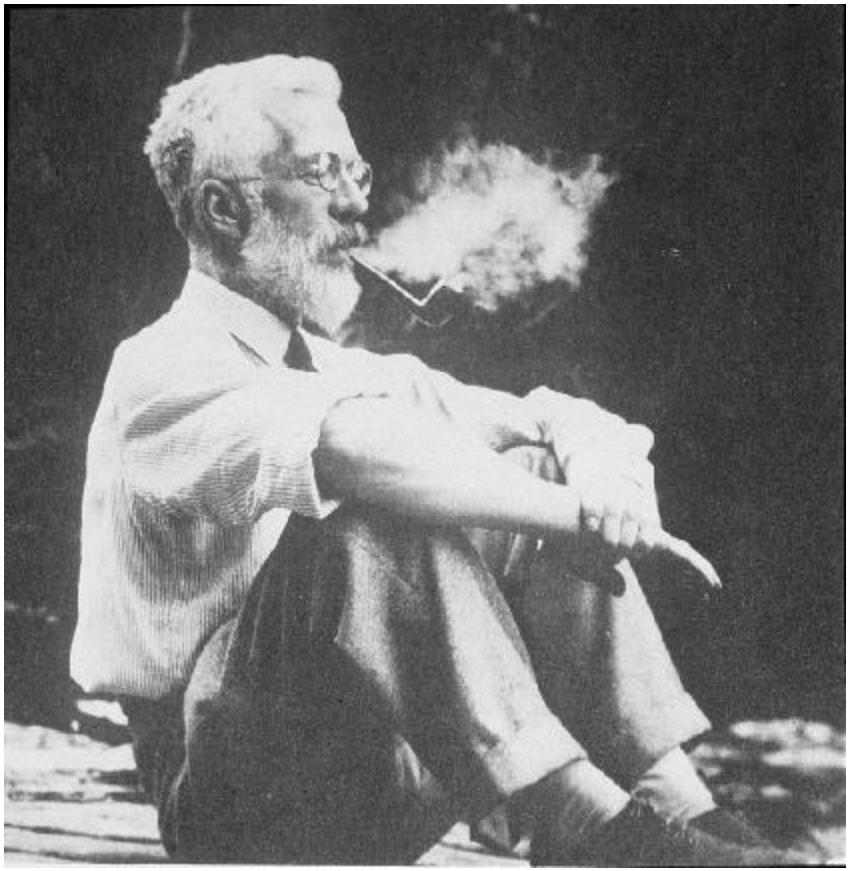


Figure 2 The inheritance of corolla length in *Nicotiana longiflora* (East 1915).

Ronald Fisher (1890-1962)



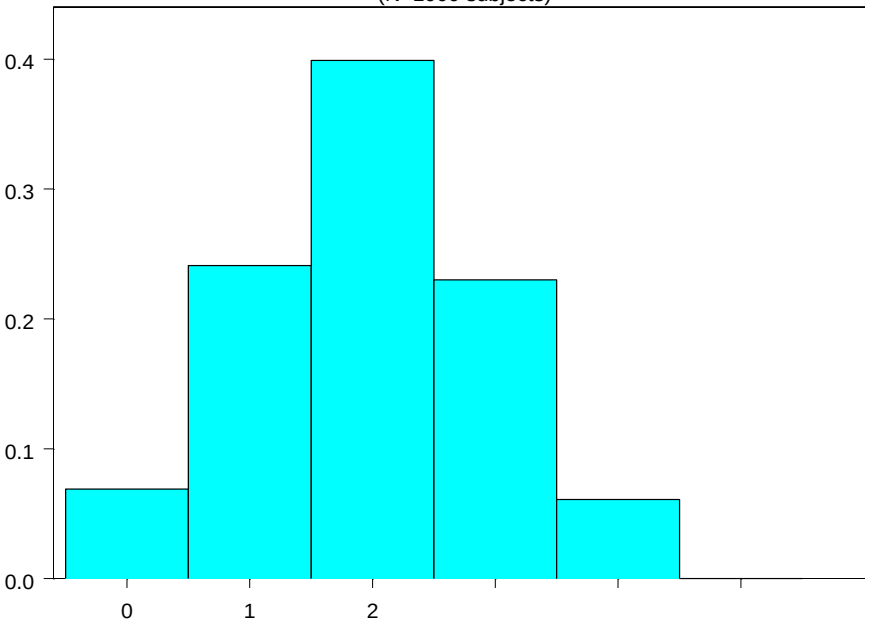
1918: The Correlation Between Relatives on the Supposition of Mendelian Inheritance
1921: Introduced concept of “likelihood”
1930: The Genetical Theory of Natural Selection
1935: The Design of Experiments

Fisher (1918): Basic Ideas

- Continuous variation caused by lots of genes ("polygenic inheritance")
- Each gene followed Mendel's laws
- Environment smoothed out genetic differences
- Genes may show different degrees of "dominance"
- Genes may have many forms ("mutliple alleles")
- Mating may not be random ("assortative mating")
- Showed that correlations obtained by e.g. Pearson and Lee were explained well by polygenic inheritance

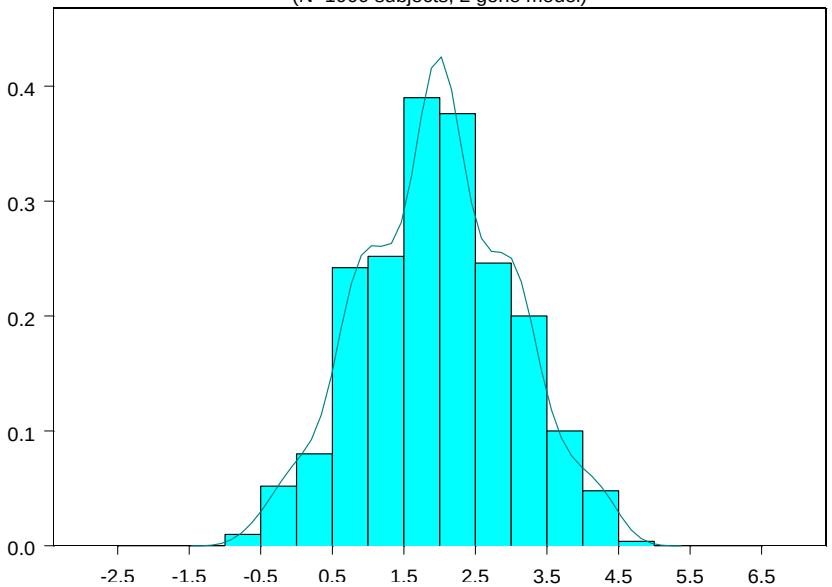
a. Distribution of scores produced by two genes

(N=1000 subjects)



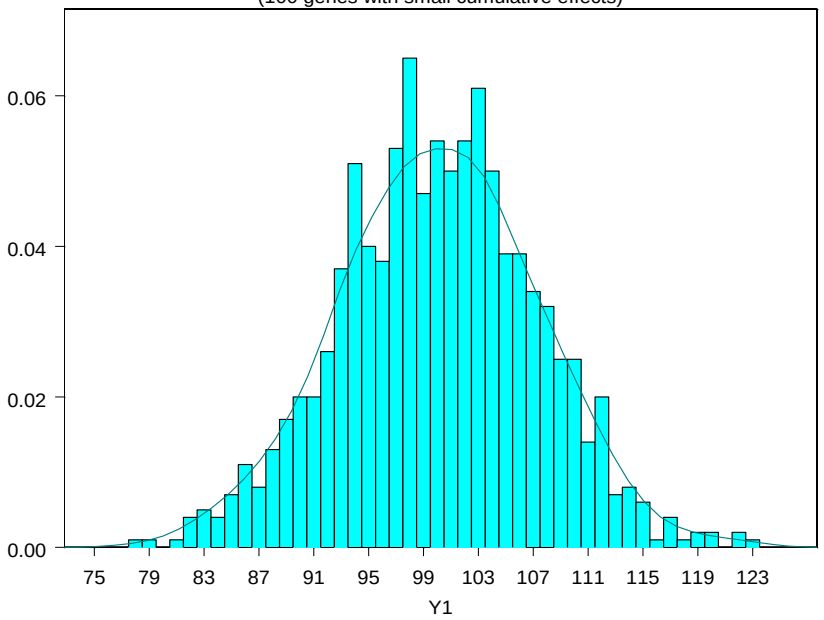
b. The "smoothing" effect of the environment

(N=1000 subjects, 2 gene model)



c. Continuous distribution of polygenic trait

(100 genes with small cumulative effects)



Kenneth Mather, FRS (1911-1990)

John Jinks, FRS (1929-1987)

Biometrical Genetics

Sir Kenneth Mather FRS
John L. Jinks FRS

Third Edition

“Biometrical Genetics”

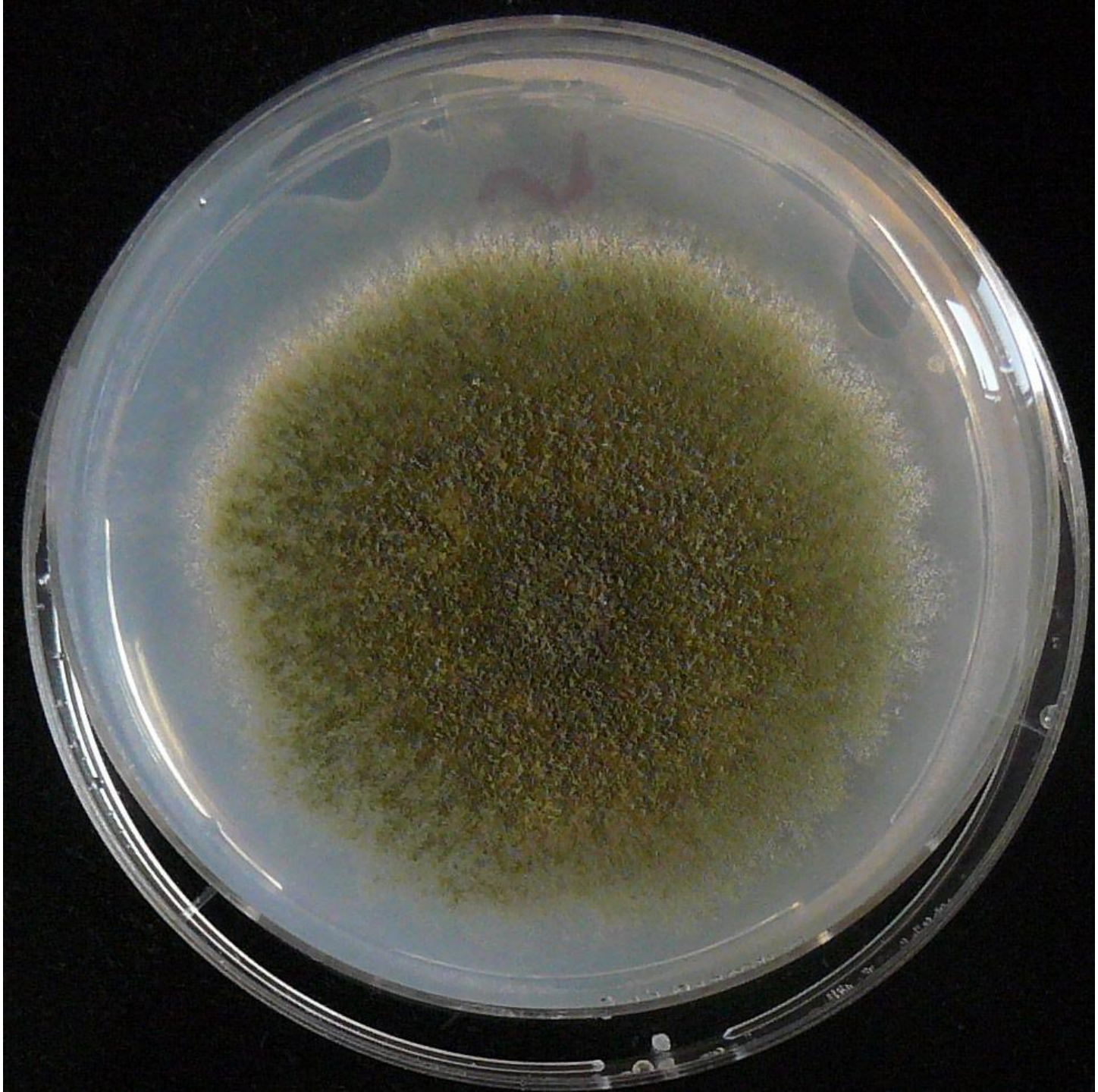
- Parsimonious specification of genetic influences in terms of effects and frequencies of individual genes (“model-building”)
- “Sensitivity to the environment” (GxE) is a phenotype like any other and analyzed with similar models
- rGE modeled by specifying genetic effects on environment e.g. effects of sibling and maternal genotype on home environment
- Systematic approach to choosing between different interpretations of the same data (“model-fitting”) e.g. effects of maternal genotype

Biometrical Genetics

- Worked out on experimental organisms
- Experimental manipulation of genotype – inbreeding and crossing
- Experimental control of environment – measurement and randomization
- Large, powerful, randomized genetic studies reveal subtleties of genetic systems – dominance, epistasis, linkage, GxE, environmental effects of genes, number of genes, genetic correlation, development…..

Model organisms

SCALE!!!!!



A “Good” Model

- Fits the data
- Explains a lot of different data in terms of relatively few theoretical constructs
- Predicts and embraces new data without substantial modification or post-hoc explanation (“fudging”)

See e.g. Lakatos I, Musgrave A (1970, Eds.) *“Criticism and the Growth of Knowledge”* Cambridge: Cambridge U.P.

Also: Urbach, P (1974) Progress and degeneration in the IQ debate. *Brit. J. Phil. Sci.* **25:***235-259.*

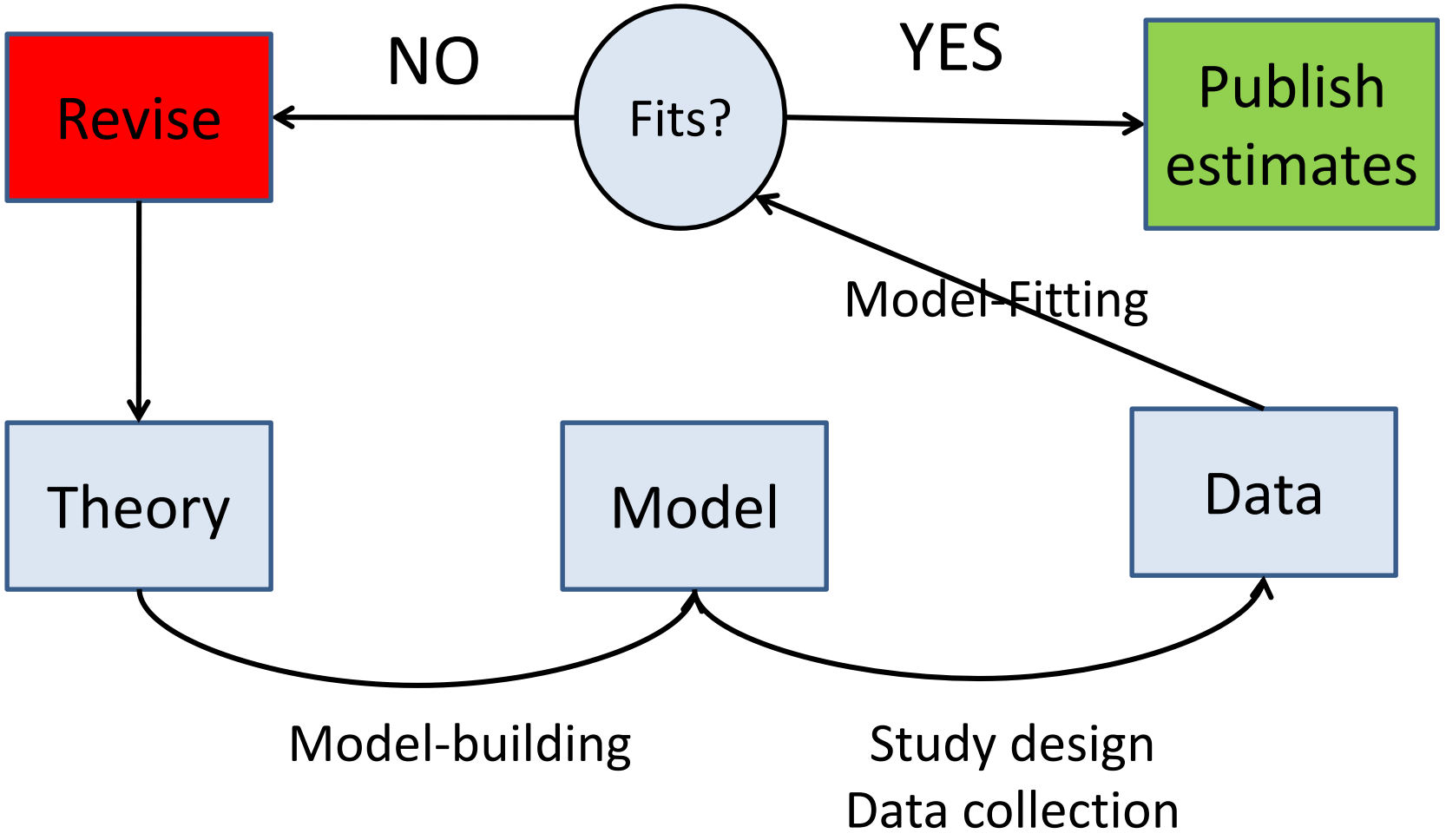
“Sociologists are like foootbol team:

zey play ze game, lose, zen shout

‘goals don’t count’”

Imre Lakatos, c. 1972.

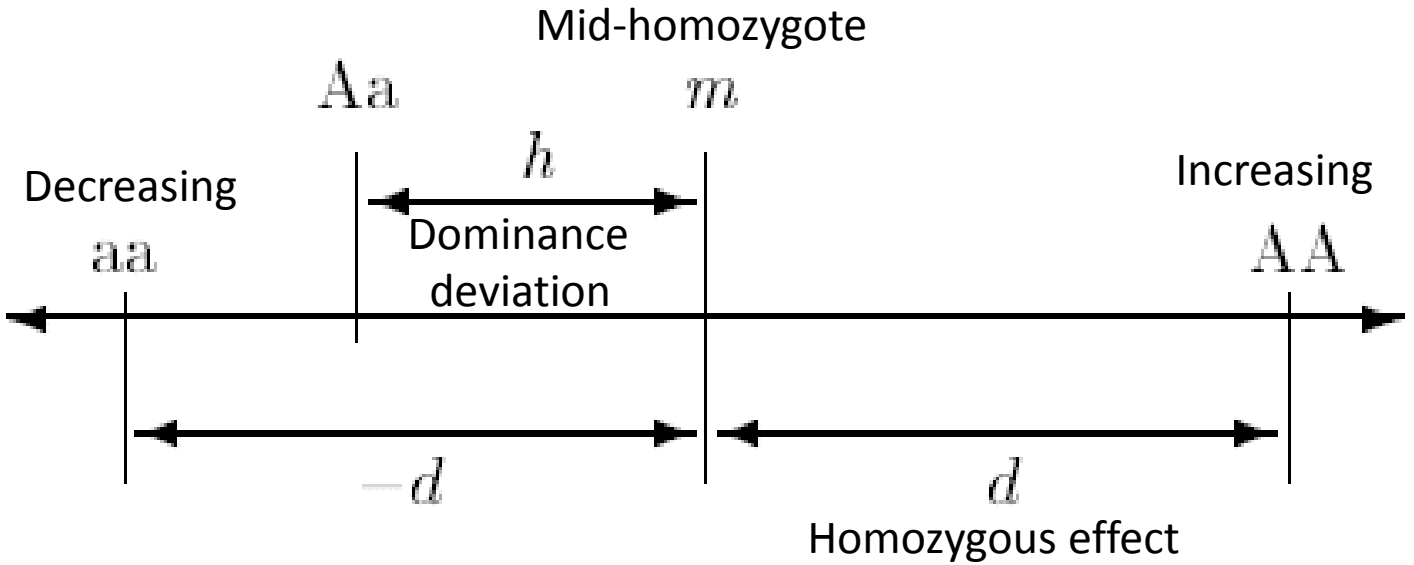
“The Logic of Scientific Discovery”



Assumptions (Initially)

- Autosomal inheritance
- No epistasis
- No sex-dependent gene expression
- Random mating
- Genes of relatives (e.g. mothers) do not affect phenotype directly
- No GxE (see Mather and Jinks for GxE)
- No G-E correlation
- Simple model for environment
- Effects of selection/mutation too small to affect result.

Basic Model for Effects of a Single Gene on a Quantitative Trait



Derivation of Genotype Frequencies
"Hardy-Weinberg Equilibrium"

		Male Gametes	
		$u\ A$	$v\ a$
Female Gametes	$u\ A$	$u^2 AA$	$uvAa$
	$v\ a$	$uvAa$	$v^2 aa$

Genotype Frequencies in Randomly Mating Population

Genotypes	AA	Aa	aa
Frequency	u^2	$2uv$	v^2

"Hardy-Weinberg Equilibrium" frequencies

What is the mean expected to be?

Genotypes	AA	Aa	aa
Frequency	u^2	$2uv$	v^2
Genotypic effect	d	h	$-d$

Note: Effects measured from mid-homozygote (“m”)

$$\begin{aligned} \mu &= u^2 d + 2uvh - v^2 d \\ &= (u - v)d + 2uvh \end{aligned}$$

With equal allele frequencies (easier!) put u=v= ½

Genotype (i)	AA	Aa	aa
Frequency (f)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Genotypic effect (x)	d	h	$-d$

And the mean is expected to be….

$$\begin{aligned} \mu_A &= \sum f_i x_i \\ &= \frac{1}{4}d + \frac{1}{2}h - \frac{1}{4}d \\ &= \frac{1}{2}h \end{aligned}$$

How does A/a affect the variance?

Equal allele frequencies u=v= ½

$$
\begin{aligned}
\sigma_A^2 &= \sum f_i (x_i - \mu_A)^2 \\
&= \frac{1}{4}(d - \frac{1}{2}h)^2 + \frac{1}{2}(h - \frac{1}{2}h)^2 + \frac{1}{4}(-d - \frac{1}{2}h)^2 \\
&= \frac{1}{4}d^2 - \frac{1}{4}dh + \frac{1}{16}h^2 + \frac{1}{8}h^2 + \frac{1}{4}d^2 + \frac{1}{4}dh + \frac{1}{16}h^2 \\
&= \frac{1}{2}d^2 + \frac{1}{4}h^2
\end{aligned}
$$

Dominance component

Additive component

Q: What happens with lots of genes?

A: The effects of the individual genes add up.

IF… the genes are independent
("linkage equilibrium")
Requires random mating, complete admixture

So:

$$\mu = \frac{1}{2}\sum_{i=1}^{k} h_i \ ,$$

$$\begin{aligned} \sigma^2 &= \frac{1}{2}\sum_{i=1}^{k} d_i^2 + \frac{1}{4}\sum_{i=1}^{k} h_i^2 \\ &= V_A + V_D \ . \end{aligned}$$

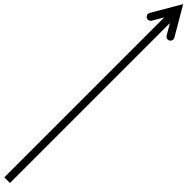
Additive Genetic Variance

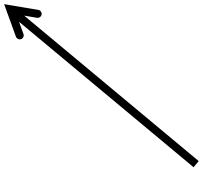
Dominance Genetic Variance

Additive and Dominance Components: Unequal allele frequencies.

Can show (see e.g. Mather, 1949)

$$
\begin{aligned}
\sigma^2 &= u^2d^2 + 2uvh^2 + v^2d^2 - [(u-v)d + 2uvh]^2 \\
&= u^2d^2 + 2uvh^2 + v^2d^2 - [(u-v)^2d^2 + 4uvdh(u-v) + 4u^2v^2h^2] \\
&= u^2d^2 + 2uvh^2 + v^2d^2 - [(u^2 - 2uv - v^2)d^2 + 4uvdh(u-v) + 4u^2v^2h^2] \\
&= 2uv[d^2 + 2(v-u)dh + (1-2uv)h^2] \\
&= 2uv[d^2 + 2(v-u)dh + (v-u)h^2 + 2uvh^2] \\
&= 2uv[d + (v-u)h]^2 + 4u^2v^2h^2 \; . \qquad (2.13)
\end{aligned}
$$





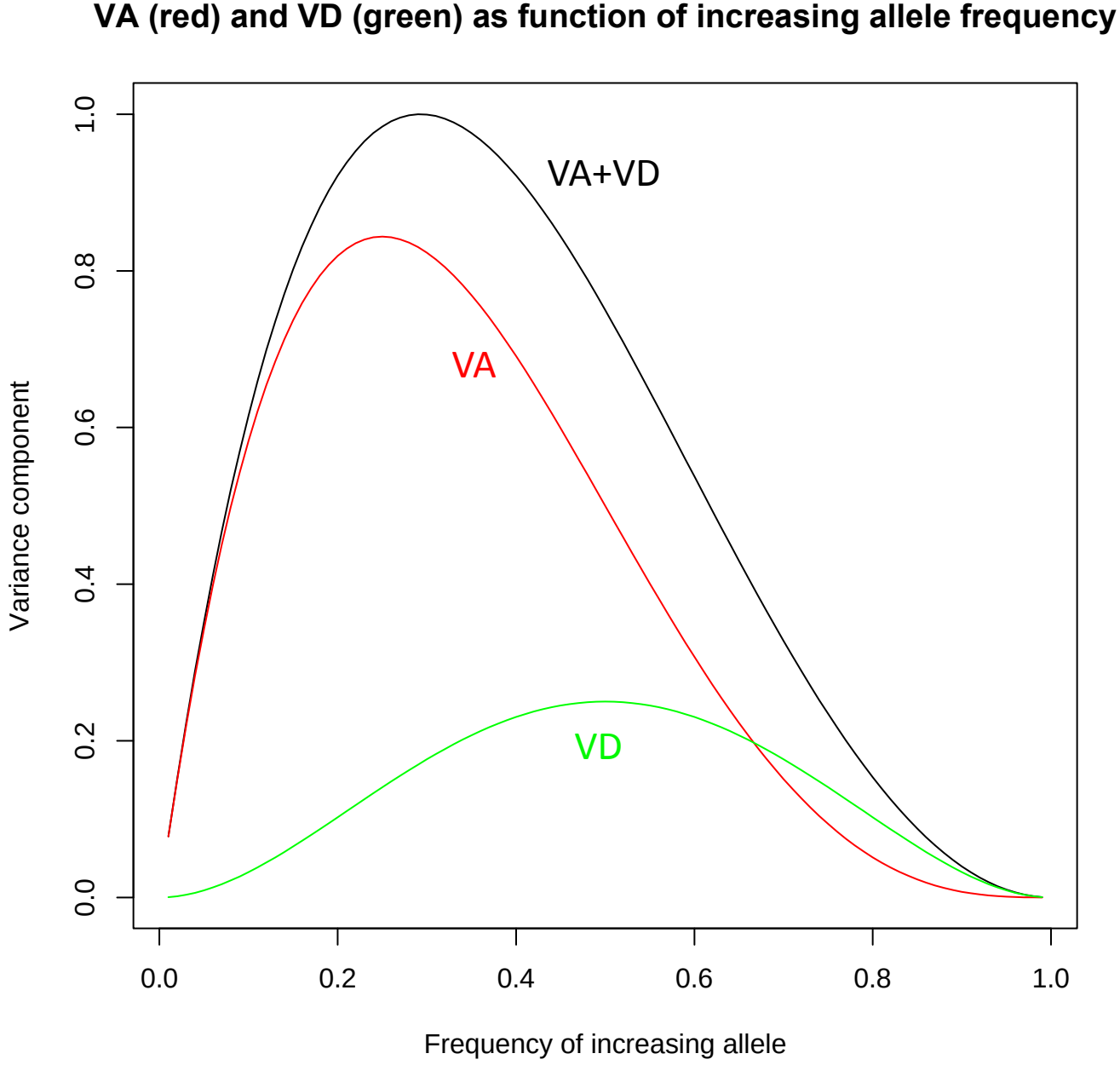
Q: What happens when u=v?

V_A

V_D

Bottom line:

With unequal allele frequencies can still separate V_A and V_D but their definitions change


VA (red) and VD (green) as function of increasing allele frequency
VA+VD
VA
VD
Variance component
0.0
0.2
0.4
0.6
0.8
1.0
Frequency of increasing allele

What about the environment???

Two main sources of environment

- Individual experiences – not shared with siblings:

$$V_E$$

- "Family" environment – shared with siblings:

$$V_C$$

So: the TOTAL variance (Genes + Environment) is:

$$V_P = V_A + V_D + V_E + V_C$$

“Heritability”

“Broad” heritability:

$h^2_b=(V_A+V_D)/V_P$

Proportion of total variance explained by genes

“Narrow” heritability:

$h^2_n=V_A/V_P$

Proportion of total variance explained by additive (homozygous) genetic effects (predicts response to selection – Fisher, 1930)

So far: have looked at effects on total variance…

How do V_A and V_D affect the correlations between relatives?

Contribution of genes to correlation between relatives (r):

$r = C/V_P$

Where C=Covariance between relative pairs

"C" depends of kind of relationship (sibling, parent-offspring, MZ twin etc)

But can also be expressed in terms of V_A and V_D

Approach

1. For a given relationship, work out expected frequencies of each type of pair (AA, aa etc.)
2. Write phenotypes of each type of relative
3. Compute cross-products of phenotypes of members of type of pair
4. Each cross-product by the corresponding frequency
5. Add the result of “4” across all pair types

The answer is the covariance you want (if you have done the algebra right!)

For equal allele frequencies….

Table 2.2: Genetic covariance components for MZ, DZ, and Unrelated siblings with equal gene frequencies at a single locus ($u = v = \frac{1}{2}$).

Genotype Pair	Effect					Frequency		
	x_{1i}	x_{2i}	$x_{1i} - \mu_1$	$x_{2i} - \mu_2$	$(x_{1i} - \mu_1)(x_{2i} - \mu_2)$	MZ	DZ	U
AA, AA	d	d	$d - \frac{1}{2}h$	$d - \frac{1}{2}h$	$d^2 - dh + \frac{1}{4}h^2$	$\frac{1}{4}$	$\frac{9}{64}$	$\frac{1}{16}$
AA, Aa	d	h	$d - \frac{1}{2}h$	$\frac{1}{2}h$	$\frac{1}{2}dh - \frac{1}{4}h^2$	-	$\frac{3}{32}$	$\frac{1}{8}$
AA, aa	d	$-d$	$d - \frac{1}{2}h$	$-d - \frac{1}{2}h$	$-d^2 + \frac{1}{4}h^2$	-	$\frac{1}{64}$	$\frac{1}{16}$
Aa, AA	h	d	$\frac{1}{2}h$	$d - \frac{1}{2}h$	$\frac{1}{2}dh - \frac{1}{4}h^2$	-	$\frac{3}{32}$	$\frac{1}{8}$
Aa, Aa	h	h	$\frac{1}{2}h$	$\frac{1}{2}h$	$\frac{1}{4}h^2$	$\frac{1}{2}$	$\frac{5}{16}$	$\frac{1}{4}$
Aa, aa	h	$-d$	$\frac{1}{2}h$	$-d - \frac{1}{2}h$	$-\frac{1}{2}dh - \frac{1}{4}h^2$	-	$\frac{3}{32}$	$\frac{1}{8}$
aa, AA	$-d$	d	$-d - \frac{1}{2}h$	$d - \frac{1}{2}h$	$-d^2 + \frac{1}{4}h^2$	-	$\frac{1}{64}$	$\frac{1}{16}$
aa, Aa	$-d$	h	$-d - \frac{1}{2}h$	$\frac{1}{2}h$	$-\frac{1}{2}dh - \frac{1}{4}h^2$	-	$\frac{3}{32}$	$\frac{1}{8}$
aa, aa	$-d$	$-d$	$-d - \frac{1}{2}h$	$-d - \frac{1}{2}h$	$d^2 + dh + \frac{1}{4}h^2$	$\frac{1}{4}$	$\frac{9}{64}$	$\frac{1}{16}$

$\mu_{x_1} = \mu_{x_2} = \frac{1}{2}h$ in all cases; genetic covariance $= \sum_i f_i(x_{1i} - \mu_1)(x_{2i} - \mu_2)$

Contribution of one gene to covariance:

$$\begin{aligned}
\mathrm{Cov(MZ)} &= d^2(\frac{1}{4}+\frac{1}{4}) + dh(-\frac{1}{4}+\frac{1}{4}) + \frac{1}{4}h^2(\frac{1}{4}+\frac{2}{4}+\frac{1}{4}) \\
&= \frac{1}{2}d^2 + \frac{1}{4}h^2 \ ,
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov(DZ)} &= d^2(\frac{9}{64}-\frac{1}{64}-\frac{1}{64}+\frac{9}{64}) \\
&+ dh(-\frac{9}{64}+\frac{3}{64}+\frac{3}{64}-\frac{3}{64}-\frac{3}{64}+\frac{9}{64}) \\
&+ \frac{1}{4}h^2(\frac{9}{64}-\frac{6}{64}+\frac{1}{64}-\frac{6}{64}+\frac{20}{64}-\frac{6}{64}+\frac{1}{64}-\frac{6}{64}+\frac{9}{64}) \\
&= \frac{1}{4}d^2 + \frac{1}{16}h^2
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov(U)} &= d^2(\frac{1}{16}-\frac{1}{16}-\frac{1}{16}+\frac{1}{16}) \\
&= dh(-\frac{1}{16}+\frac{1}{16}+\frac{1}{16}-\frac{1}{16}-\frac{1}{16}+\frac{1}{16}) \\
&= \frac{1}{4}h^2(\frac{1}{16}-\frac{2}{16}+\frac{1}{16}-\frac{2}{16}+\frac{4}{16}-\frac{2}{16}+\frac{1}{16}-\frac{2}{16}+\frac{1}{16}) \\
&= 0
\end{aligned}$$

Notice that terms in d^2 and h^2 are separated – but their coefficients change as a function of relationship

Can add over all genes to get total contribution to covariance

$$Cov(MZ) = V_A + V_D$$

$$Cov(DZ) = \tfrac{1}{2}V_A + \tfrac{1}{4}V_D$$

$$Cov(U) = 0$$

Can use the same approach for other relationships

Contributions of V_A and V_D to covariances between relatives (ignoring environment)

Relationship	Contribution to Covariance	
	V_A	V_D
Total variance	1	1
Sibling (DZ twin)	½	¼
MZ twin	1	1
Half-sibling	¼	0
First cousin	$^1/_8$	0
Parent-offspring	½	0
Avuncular	¼	0
Grand-parent	$^1/_8$	0
Unrelated	0	0

Adding effects of Environment

$V_P = V_A + V_D + V_E + V_C$

$Cov(MZ) = V_A + V_D + V_C$

$Cov(DZ) = \frac{1}{2}V_A + \frac{1}{4}V_D + V_C$

$Cov(UT) = V_C$

Etc.

To get the expected correlations

Just divided expectations by expected total variance

Results are proportional contributions of V_A, V_D etc. to total variance

Practice (paper and pencil)

- Set “d” = 1
- Pick an “h” (e.g. h=-1.0,-0.5,0,0.5,1.0)
- Pick a frequency for the increasing (A) allele (e.g. u=0.25, 0.5, 0.75)
- Work out μ, VA and VD
- Tabulate on board

Substitute in algebra:
Get your own parameter values

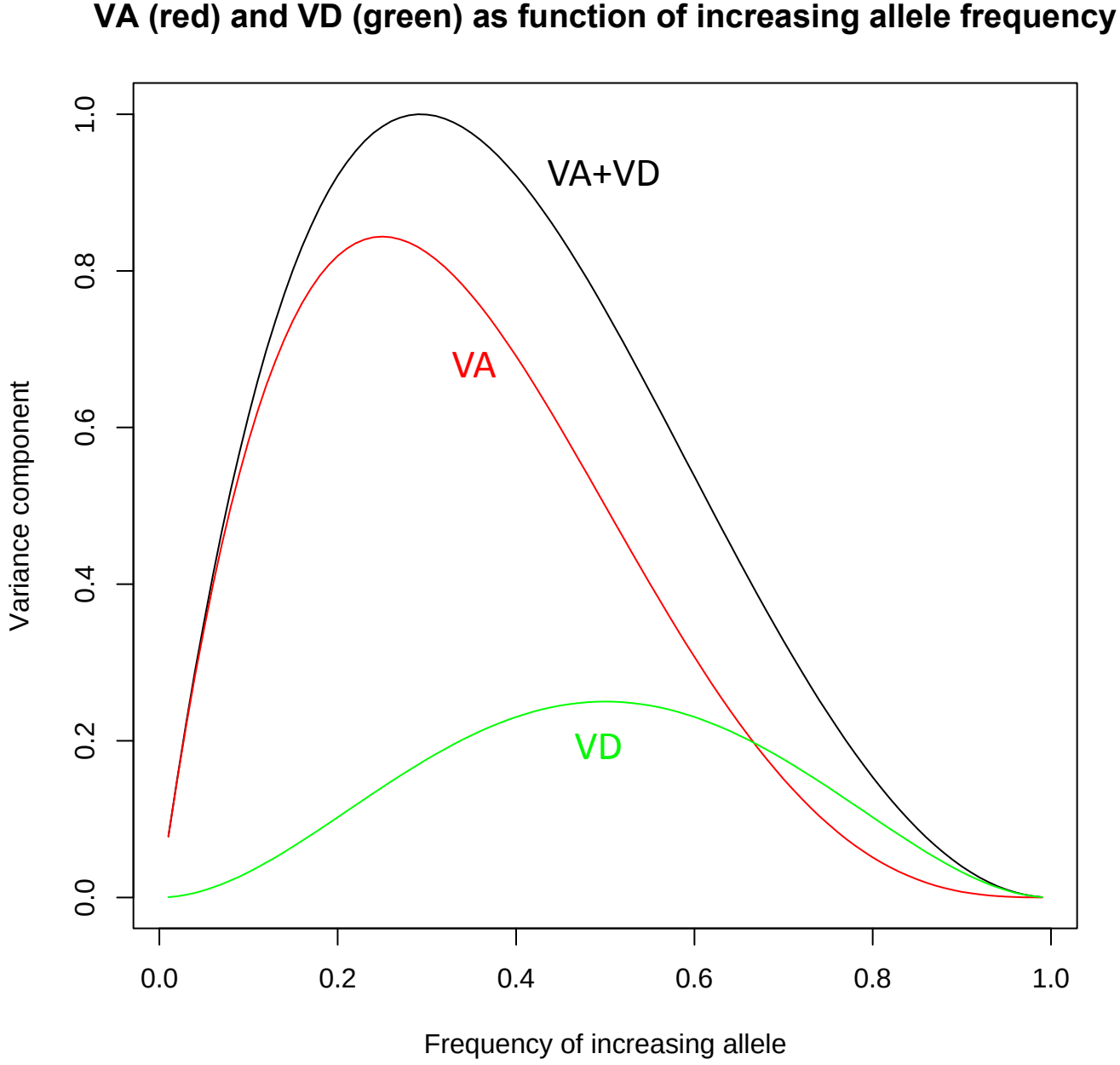
$$\mu = (u-v)d + 2uvh$$

$$V_A = 2uv[d + (v-u)h]^2$$

$$V_D = 4u^2v^2h^2$$

Plotting Effect of Allele frequency on Genetic Variance Components ("R")

```
d<-1                          #  Homozygous effect ("additive")
h<-1                          #  Heterozygous deviation ("dominance")
u<-seq(0.01,0.99,by=.01)      # Vector of frequencies of increasing allele
v<-1-u                        # Frequencies of decreasing allele
VA<-2*u*v*(d+(v-u)*h)^2       # Additive genetic variance
VD<-4*u*u*v*v*h*h             # Dominance genetic variance
VP<-VA+VD                     # Total (genetic) variance
#  Plot results
plot(u,VP,type="l",
main="VA (red) and VD (green) as function of increasing allele frequency",
xlab="Frequency of increasing allele",ylab="Variance component")
# Add line for VA
lines(u,VA,col="red")
# Add line for VD
lines(u,VD,col="green")
```


VA (red) and VD (green) as function of increasing allele frequency
VA+VD
VA
VD
Variance component
0.0
0.2
0.4
0.6
0.8
1.0
Frequency of increasing allele