

Genetics 101: Intro to History and Theory

**Laura Almasy
John Blangero
Vincent Diego
And
Peter Kochunov**



Outline - revised

Part 1: THEORY

- I. Basic concepts of variance component models for genetics
- II. Heritability
- III. Covariates
- IV. Association
- V. A little bit about linkage

Break

Part 2: SOFTWARE

- I. SOLAR: download, registration, documentation, user support
- II. Pedigree file – what if you don't have families?
- III. Phenotype file – polygenic, heritability, covariates, normalization
- IV. SNP and map files – linkage disequilibrium, association analysis
- V. Just a little bit about linkage

Modern genetics starts with the work of Gregor Johann Mendel (1822-1884)



Mendelian Genetics: law of inheritance

- ✎ From 1856 to 1863, Mendel performed experimental crosses on some 29,000 pea plants.
- ✎ He published this work in 1866 in:
 - Experiments on Plant Hybridization
- ✎ He proposed what would come to be known as Mendel's Laws of Inheritance
 - The Law of Segregation or The First Law
 - The Law of Independent Assortment or The Second Law

Seed shape



Spherical



Dented

Seed color



Yellow



Green

Flower color

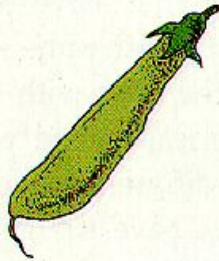


Purple

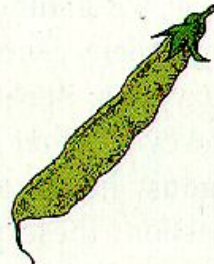


White

Pod shape

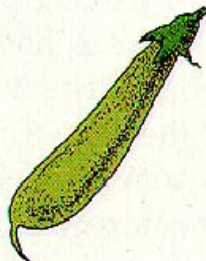


Inflated

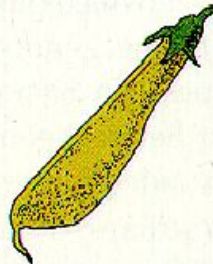


Constricted

Pod color



Green



Yellow

Flower position



Axial



Terminal

Stem height



Tall



Dwarf

Second school of thought: The Biometricians

- Rise of British biometry started by Francis Galton in the late 1800s.
- Chief scholars included Karl Pearson and Walter Weldon.
- In competition to Mendelian school of inheritance
- These camps were essentially addressing two different domains of genetics: simple and complex traits, respectively.

The Mendelians versus the Biometricians

- This controversy was resolved amicably in a seminal paper by Ronald A. Fisher:

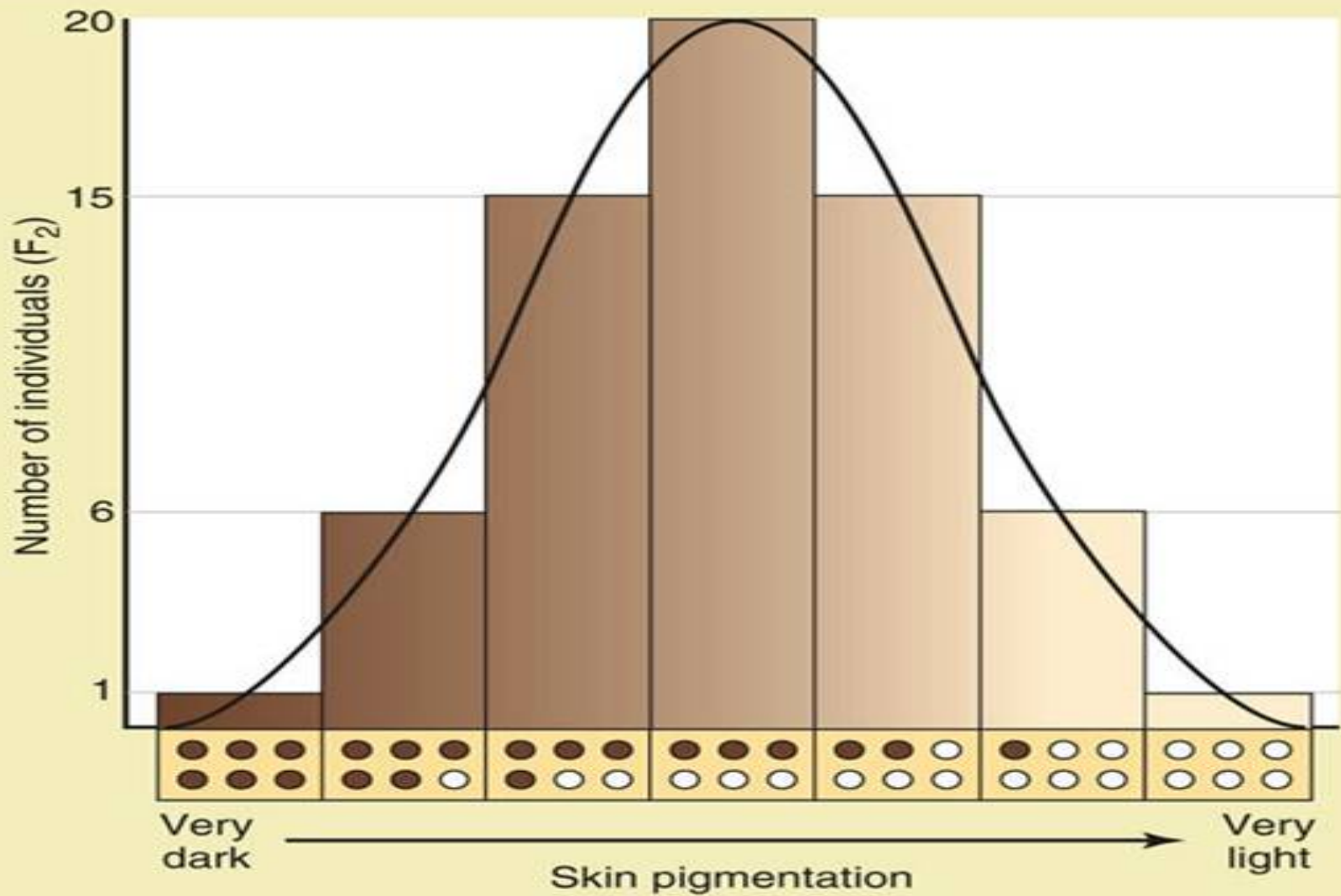
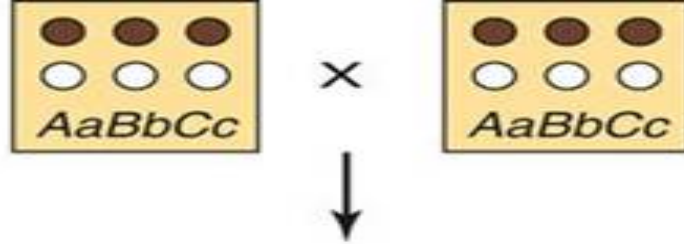
XV.—*The Correlation between Relatives on the Supposition of Mendelian Inheritance.* By R. A. Fisher, B.A. *Communicated by* Professor J. ARTHUR THOMSON. (With Four Figures in Text.)

(MS. received June 15, 1918. Read July 8, 1918. Issued separately October 1, 1918.)

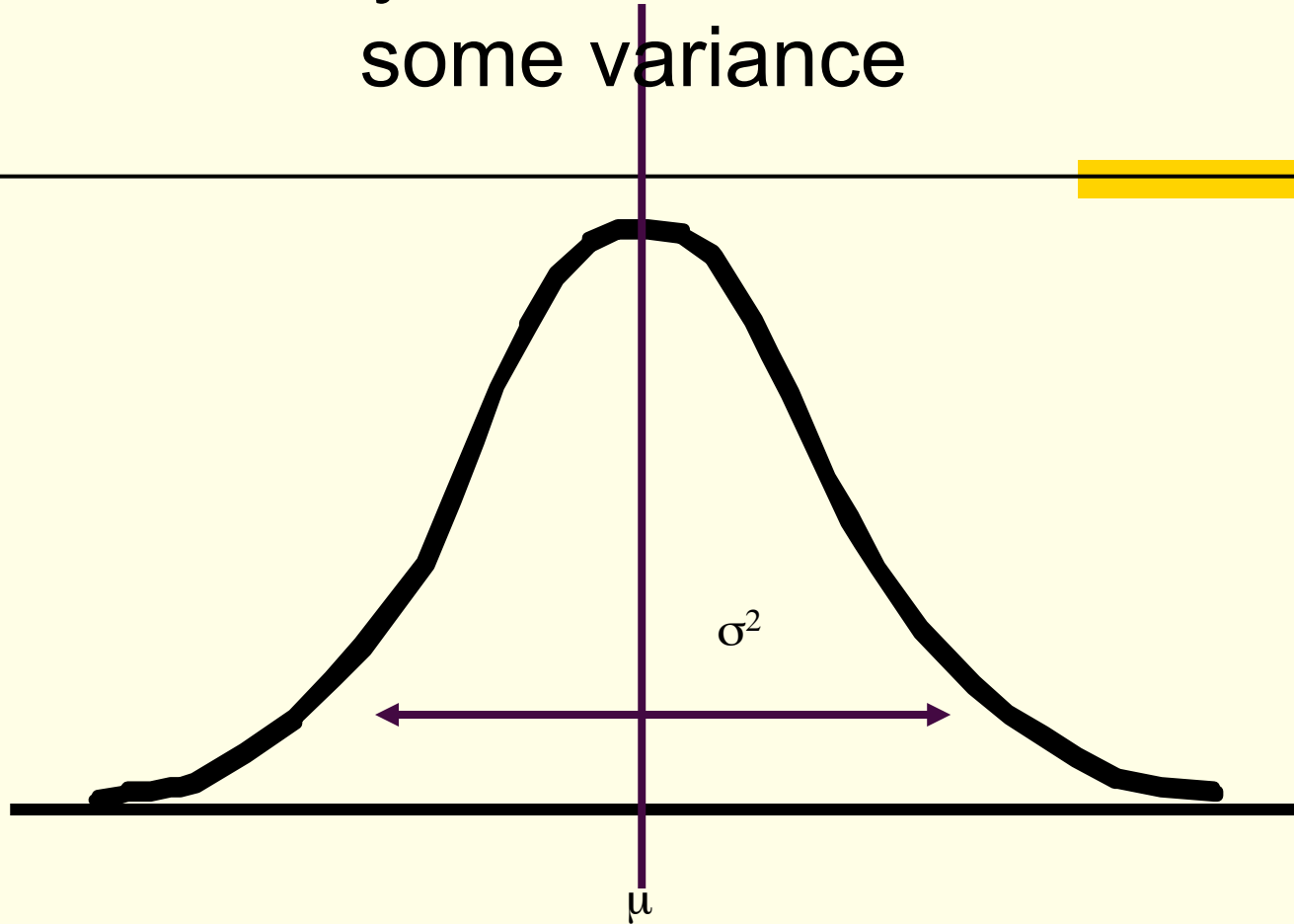
Transactions of the Royal Society of Edinburgh, 52: 399-433, (1918).

The Fisher's Bridge

- In his 1918 paper Fisher reconciled the Mendelian segregation of genes with the determination process of complex traits.
- His model assumed that complex traits were determined by a large number of genes (each one exhibiting Mendelian segregation) with small additive effects.
- The term “variance” was first coined here.
- The modern day linear model along with the use of the normal distribution trace back to this paper.



Normally distributed Trait with some variance



Variance Decomposition

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2$$

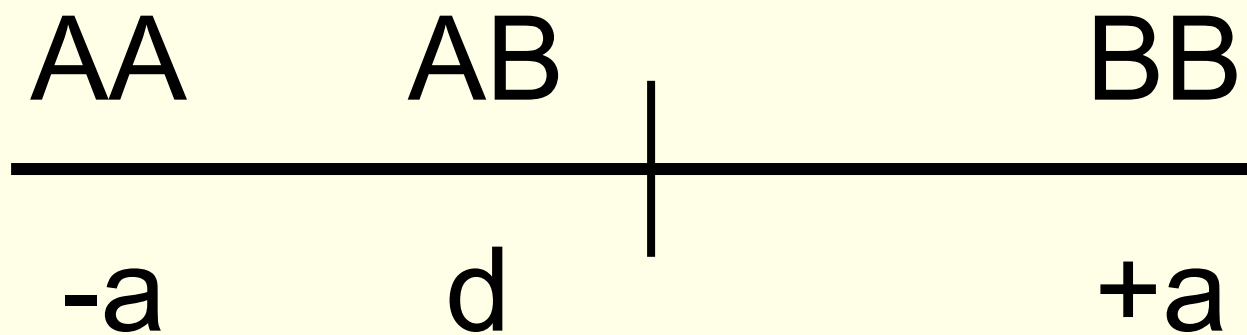
$$\sigma_p^2 = \text{Total phenotypic variance}$$

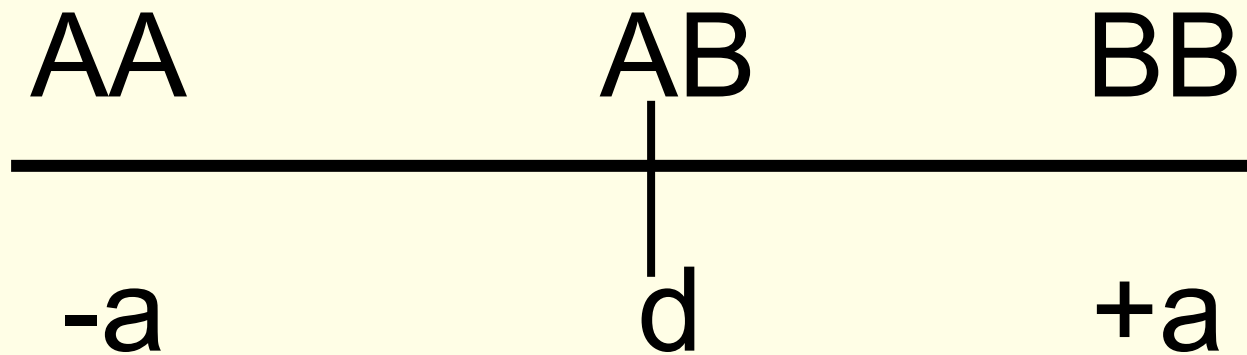


$$\sigma_g^2 = \sigma_a^2 + \sigma_d^2$$

σ_a^2 = Additive genetic variance

σ_d^2 = Dominance variance





If the heterozygote is half way between the two homozygotes, there's a "dose-response" effect, d is zero, and there is no dominance.

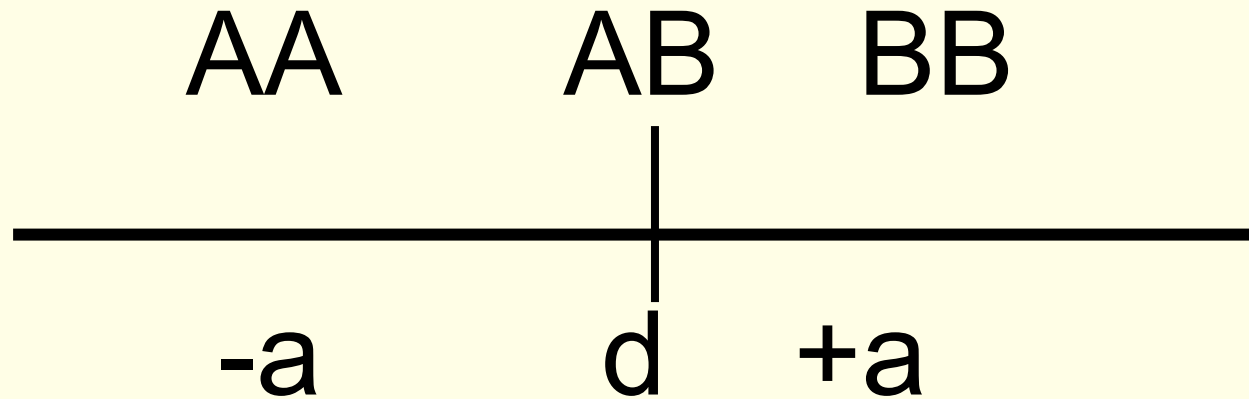


$$\sigma^2_a = 2pq[a + d(q-p)]^2$$

$$\sigma^2_d = (2pqd)^2$$

$$\sigma_g^2 = \sigma_a^2 + \sigma_d^2 + \sigma_i^2$$

σ_i^2 = Interaction variance (epistasis)



Interaction variance exists when a (or d) is a function of the genotypes at another locus.


$$\sigma_e^2 = \sigma_c^2 + \sigma_{ue}^2$$

σ_c^2 = Common or shared environment

σ_{ue}^2 = Unique environment



Heritability (h^2): the proportion of the phenotypic variance in a trait that is attributable to the additive effects of genes.

Broad sense heritability

$$h^2 = \frac{\sigma_g^2}{\sigma_p^2}$$

Additive genetic (narrow sense) heritability

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2}$$

Modeling the Phenotype

$$p = \mu + \sum \beta_i x_i + a + e$$

μ Baseline mean

β Regression coefficients

x Scaled covariates

a Additive genetic effects

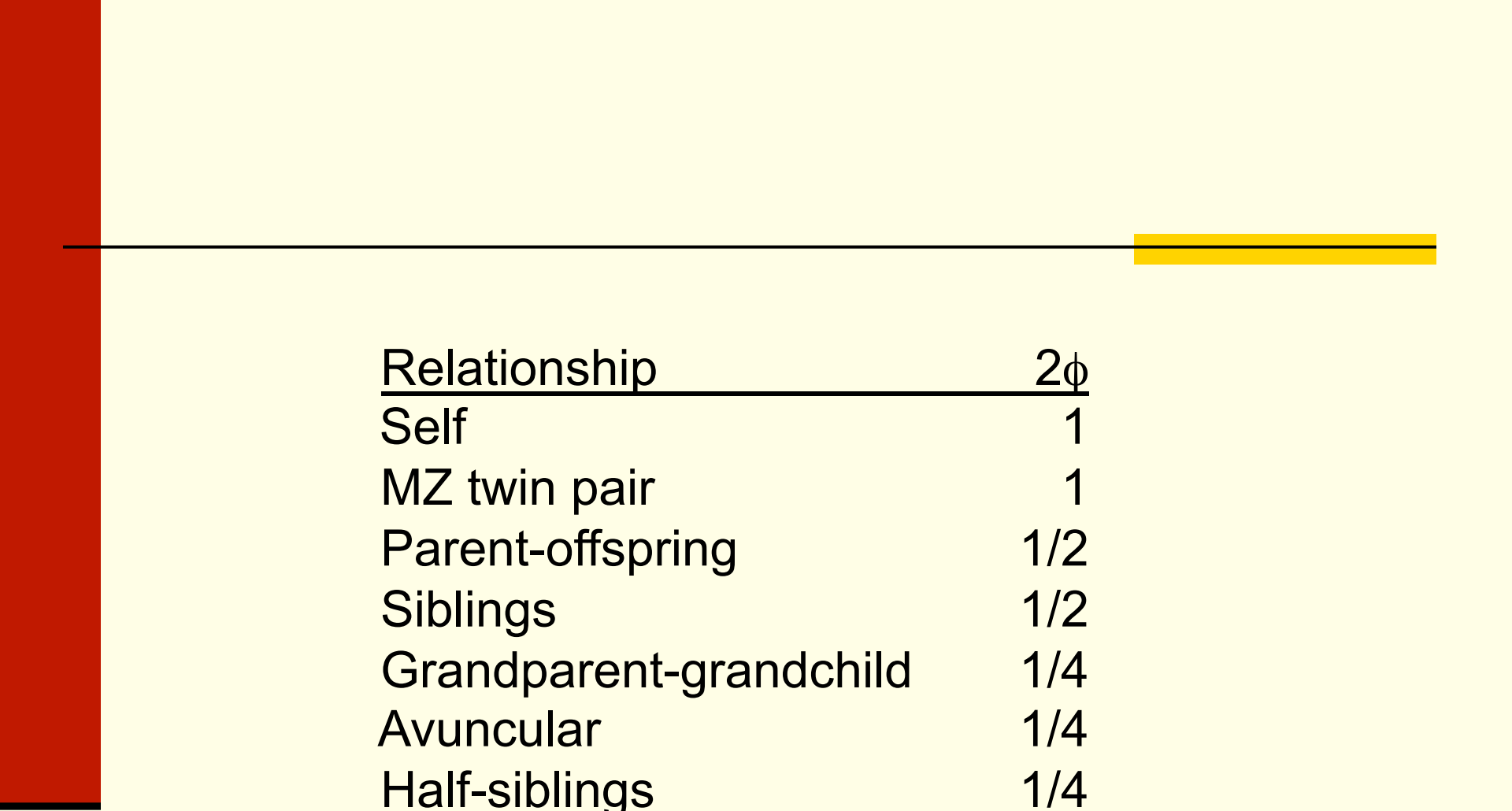
e Random environmental effects

Modeling phenotypic covariance

$$\Omega = 2\Phi\sigma_a^2 + I\sigma_e^2$$

σ_a^2 = additive genetic variance

σ_e^2 = variance due to unique environmental influences



Relationship	2ϕ
Self	1
MZ twin pair	1
Parent-offspring	1/2
Siblings	1/2
Grandparent-grandchild	1/4
Avuncular	1/4
Half-siblings	1/4
1st cousins	1/8
2nd cousins	1/32

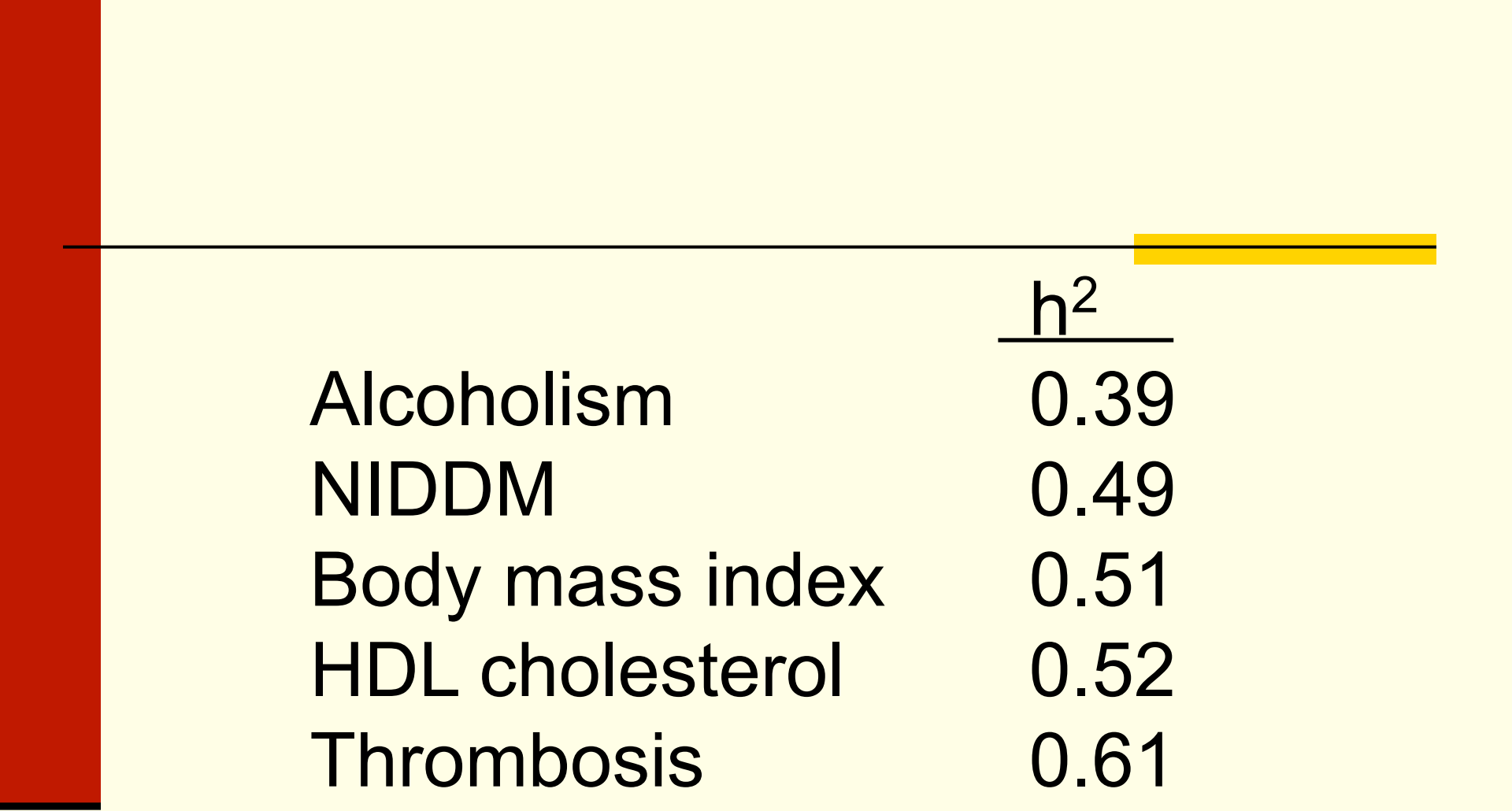
Hypothesis Testing:

Null hypothesis $h^2 = 0$

Parameters Estimated

Model	σ^2_a	σ^2_e
Sporadic	0	+
Additive	+	+

Twice the difference in ln likelihoods between the two models is distributed as a mixture of chi-square distributions.



	<u>h²</u>
Alcoholism	0.39
NIDDM	0.49
Body mass index	0.51
HDL cholesterol	0.52
Thrombosis	0.61
Height	0.81

Limitation

This model assumes that the only source of correlation among family members is genetic.

Solution

Add more components to the model.

Shared environment

Modeling the Phenotype:

$$p = \mu + \sum \beta_i x_i + a + c + e$$

μ Baseline mean

β Regression coefficients

x Scaled covariates

a Additive genetic effects

c Shared environmental effects

e Random environmental effects

Variance Decomposition

$$\sigma_p^2 = \sigma_a^2 + \sigma_c^2 + \sigma_e^2$$

Shared environmental (household) effects

$$C^2 = \frac{\sigma_c^2}{\sigma_p^2}$$

Modeling phenotypic covariance

$$\Omega = 2\Phi\sigma_a^2 + H\sigma_c^2 + I\sigma_e^2$$

σ_a^2 = additive genetic variance

σ_c^2 = variance due to shared environmental influences

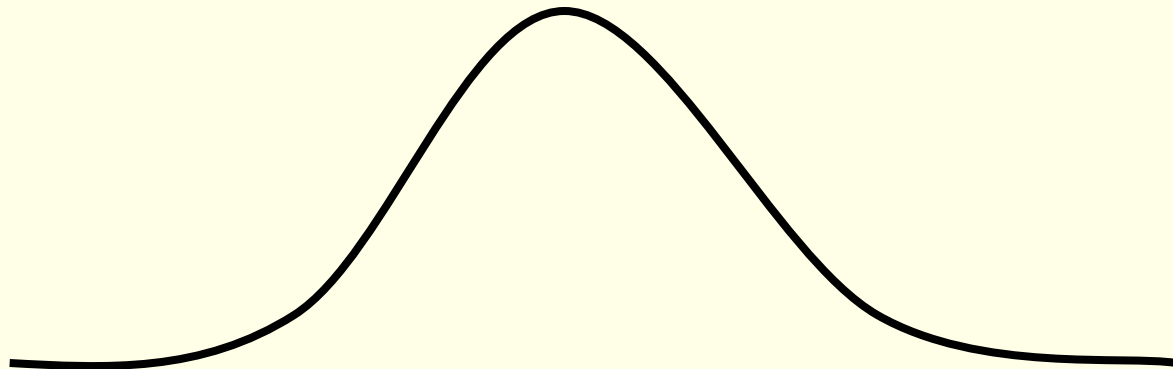
σ_e^2 = variance due to unique environmental influences

Assumption of VC analysis:

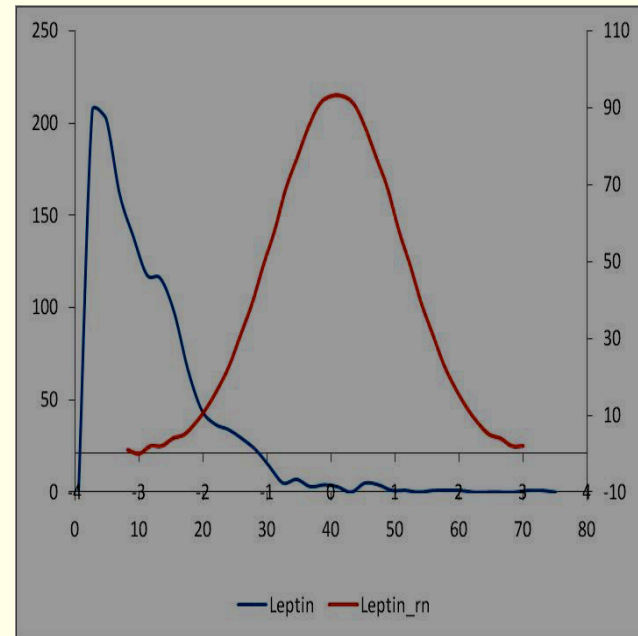
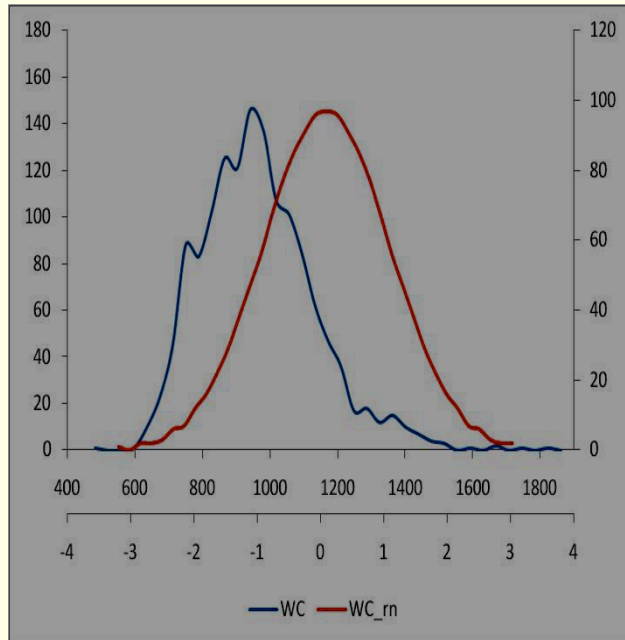
Trait is normally distributed.

What happens if this assumption is violated?

4th Central Moment:
Kurtosis



Data Transformation: Waist Circumference, Serum Leptin



What about those covariates?

Modeling the Phenotype:

$$p = \mu + \sum \beta_i x_i + a + c + e$$

μ Baseline mean

β Regression coefficients

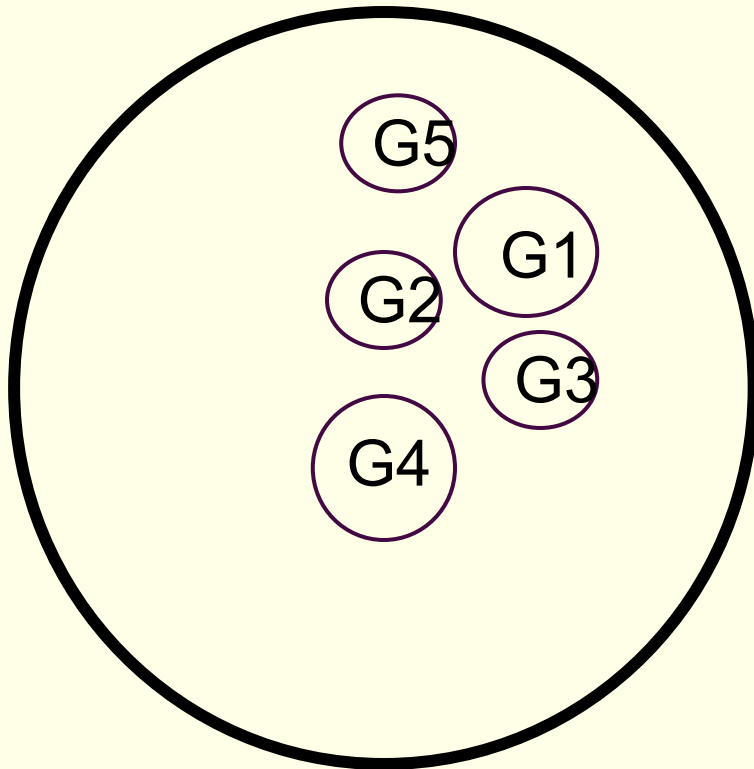
x Scaled covariates

g Additive genetic effects

c Shared environmental effects

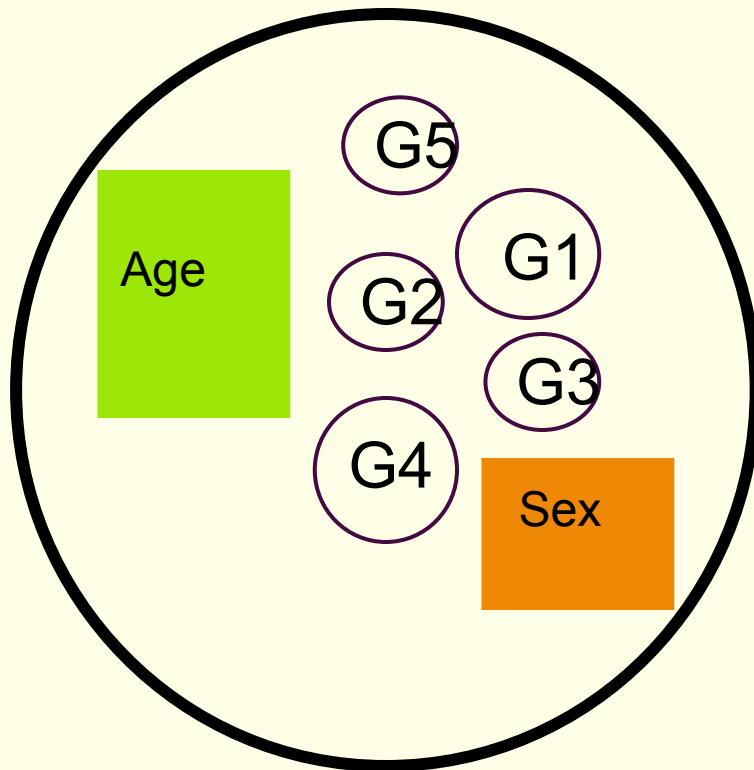
e Random environmental effects

What good are covariates?



$$h^2 = \frac{G1 + G2 + G3 + G4 + G5}{\text{Total}}$$

Covariates absorb variance



$$h^2 = \frac{G1 + G2 + G3 + G4 + G5}{\text{Total} - \text{Age} - \text{Sex}}$$

Hypothesis Testing:

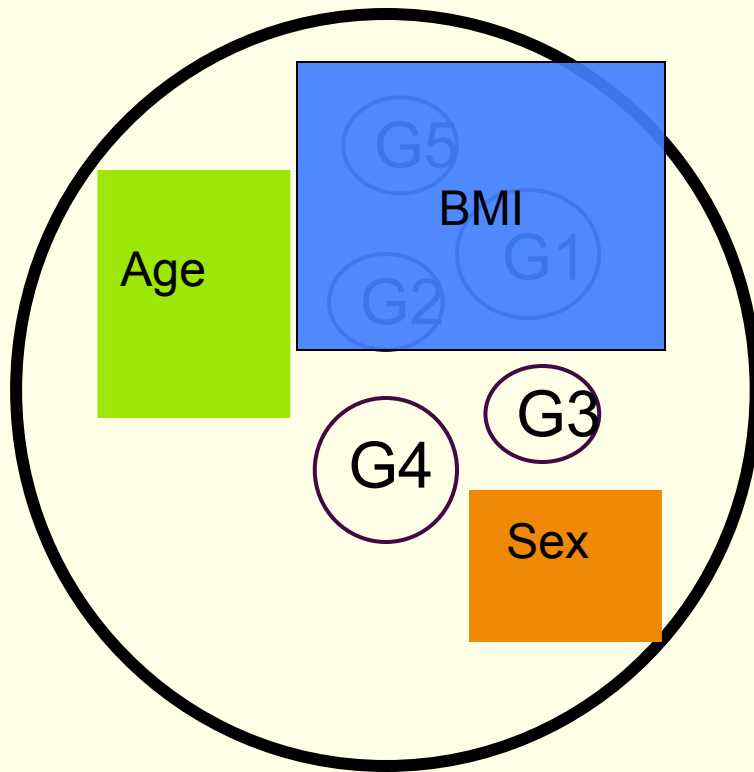
**Null hypothesis: regression coefficient
for covariate = 0**

Parameters Estimated

<u>Model</u>	<u>β</u>
Null	0
Alternate	+

Twice the difference in ln likelihoods between the two models is distributed as a chi-square with 1 df.

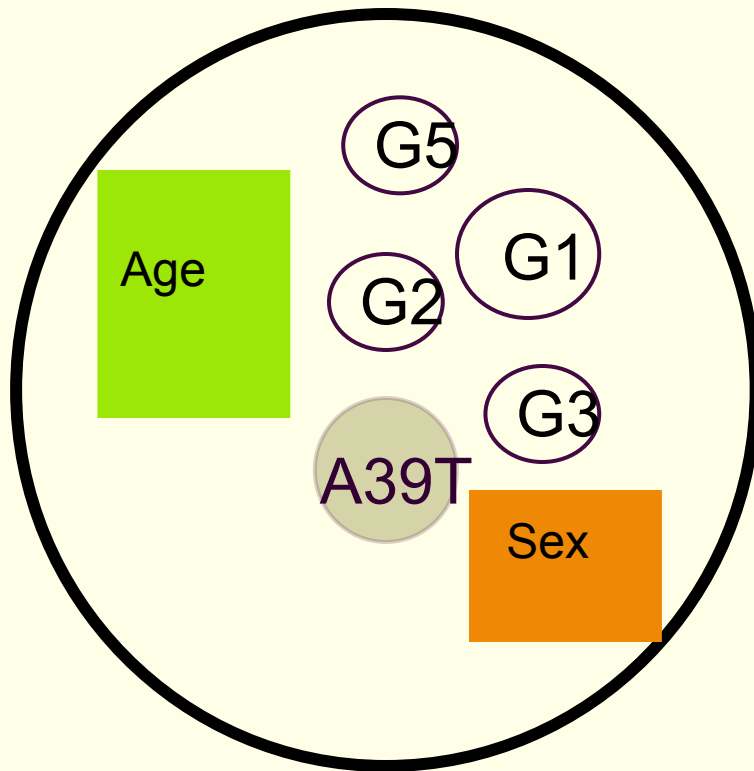
Caution! In general, you only want to use covariates that are demographic or environmental.



$$h^2 = \frac{G3 + G4}{\text{Total} - \text{Age} - \text{Sex} - \text{BMI}}$$

If we include BMI as a covariate, we reduce our power to detect G1, G2, or G5, genes that influence both BMI and the trait of interest.

Standard association analysis uses genotype as a covariate



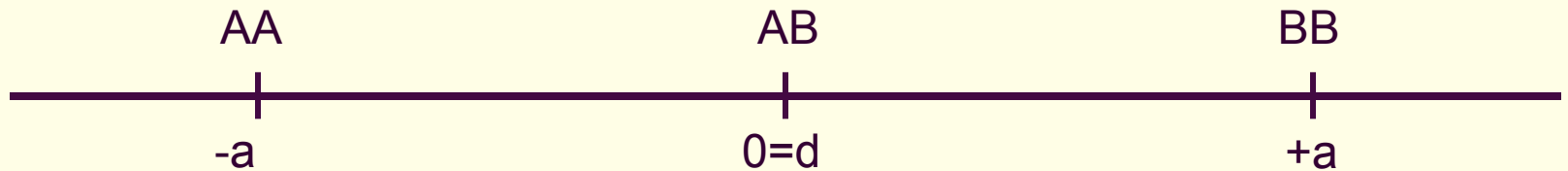
Measured genotype association tests whether the trait mean differs by genotype, taking into account the non-independence between family members.

QTN-Specific Heritability

$$h_{qtn}^2 = \frac{\sigma_{qtn}^2}{\sigma_p^2}$$

Additive genetic variance due to a QTN

assuming additivity:



$$\sigma_{QTN}^2 = H a^2, \text{ where}$$

$a = 1 / 2$ of phenotypic difference between homozygotes

$H = 2p(1 - p)$, i.e. heterozygosity

Genotypes as covariates

If effect of QTL is modeled as additive:

<u>Genotype</u>	<u>Cov</u>
AA	0
Aa	1
aa	2

Power for Association Studies

Power for association studies is a function of the sample size, the family configuration, the QTN variance, and the LD between a QTN and a genotyped marker.

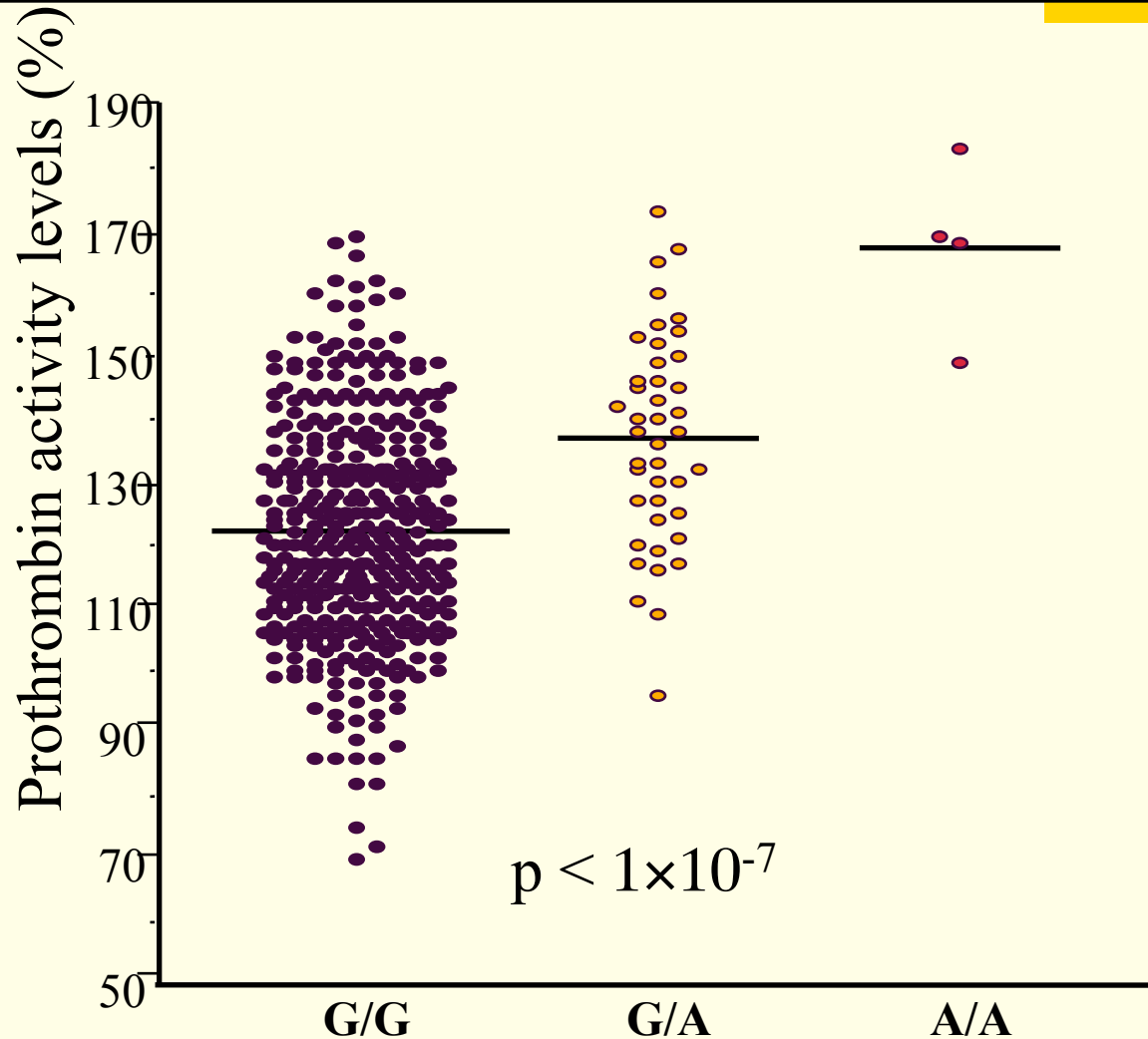
$$\text{effective } h^2_{\text{snp}} = \text{actual } h^2_{\text{qtn}} * r^2$$

Example: FXII levels by *F12* 46C/T genotype

	<u>FXII levels</u>
CC	128.88
CT	92.23
TT	55.58

$$p < 1 \times 10^{-7}$$

Prothrombin levels by G20210A genotype



Linkage analysis

$$\Omega = \hat{\Pi}\sigma_{qtl}^2 + 2\Phi\sigma_a^2 + I\sigma_e^2$$

Identity By Descent

Want to know the proportion of alleles shared by a relative pair that derive from a common ancestral source (IBD). This is in contrast to alleles shared identical by state (IBS) in which the ancestral source of the alleles is not considered.

IBS = association

IBD = linkage

π_{ij} = IBD probability for individuals i and j

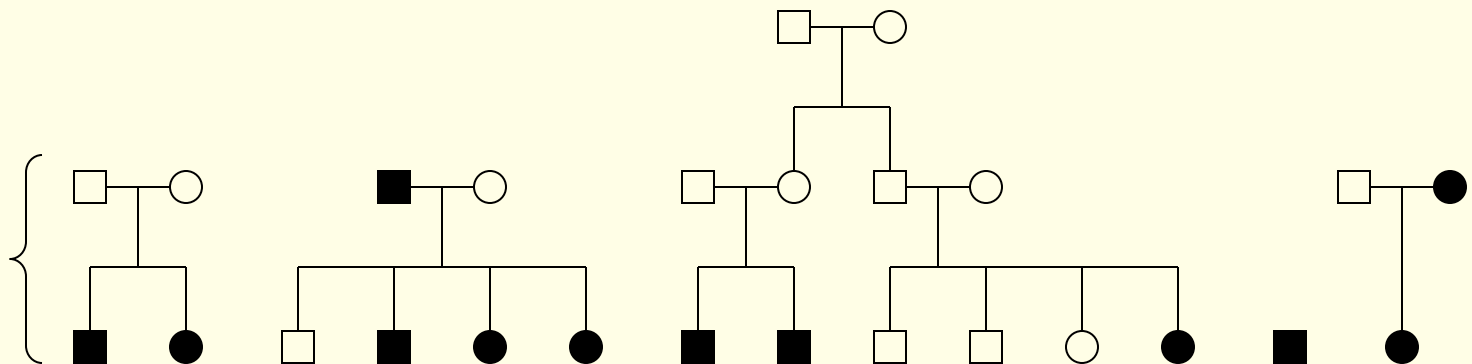
= $\frac{1}{2}$ Pr (1 allele shared IBD)
+ Pr (2 alleles shared IBD)

Variance component-based linkage analysis

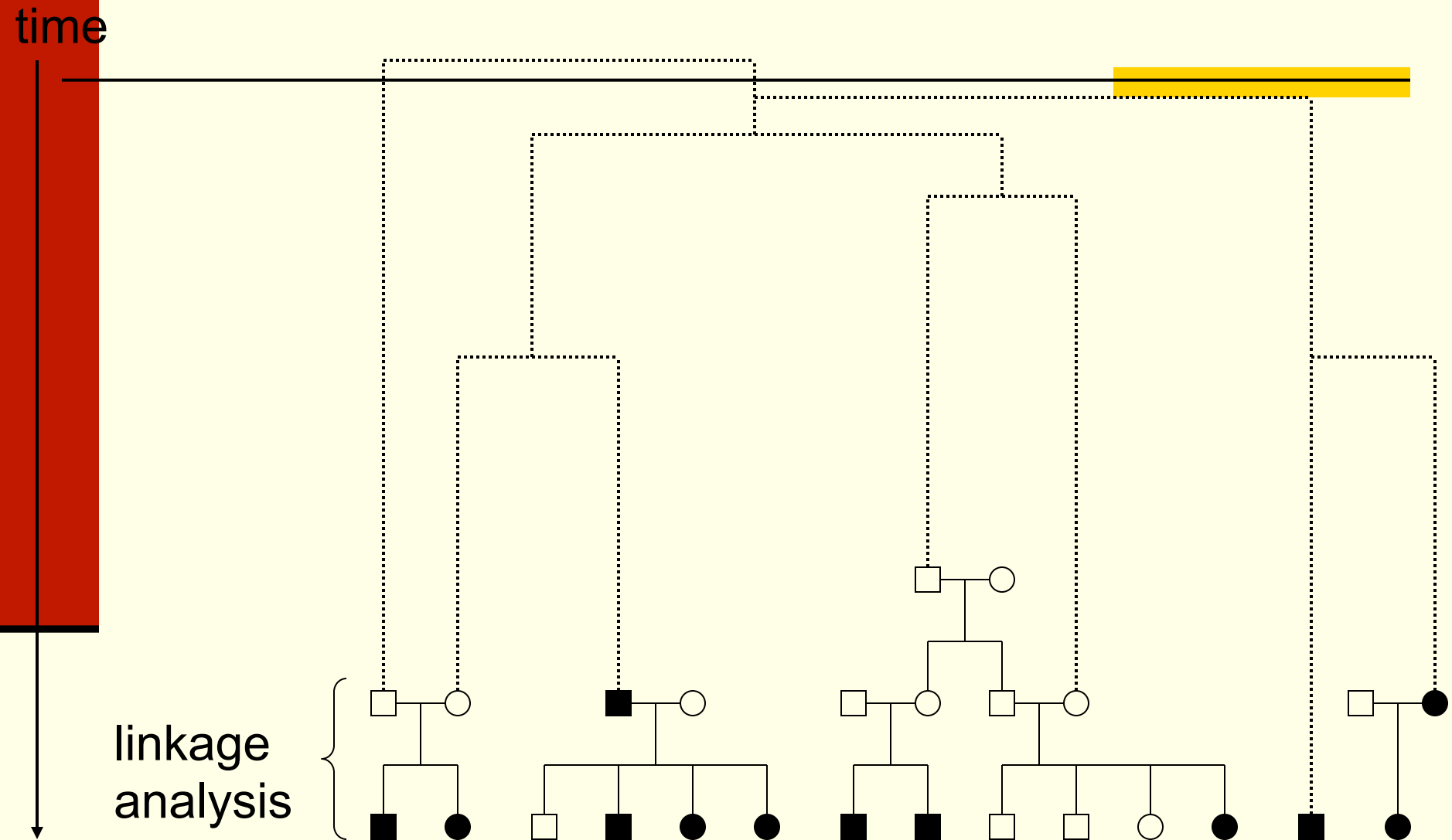
In a region containing a QTL influencing the trait, relatives who are phenotypically similar will share more alleles IBD than relatives who are phenotypically dissimilar

Linkage and association analyses

linkage
analysis



Linkage and association analyses



Now lets do it in software
