

Assessing the Genetic Component of a Phenotype

A phenotype is the appearance of an individual, which results from the interaction of the person's genetic makeup and their environment. Phenotypes can be categorical or numerical. If we are interested in the genetic component of a trait there are different methods we can use for analysis.

We define a trait for analysis, determine what we are interested in studying, and then the methods that need to be used. For example, if we want to know how BMI is linked to gene effects we would need to consider other factors such as age, sex, smoking, etc. Thus, a Multiple Linear Regression might be appropriate, were we can measure the variability of each factor.

Variance of Phenotypic Traits

σ^2_T = The observed (phenotypic) variability of a trait

$\sigma^2_T = \sigma^2_G + \sigma^2_E$ = The phenotypic variability can be partitioned in to variability due to **genetic** and **environmental** effects

$\sigma^2_G = \sigma^2_A + \sigma^2_D$ = The genetic component can be further partitioned into **additive** and **dominance** genetic variance

We can write a model for the trait as:

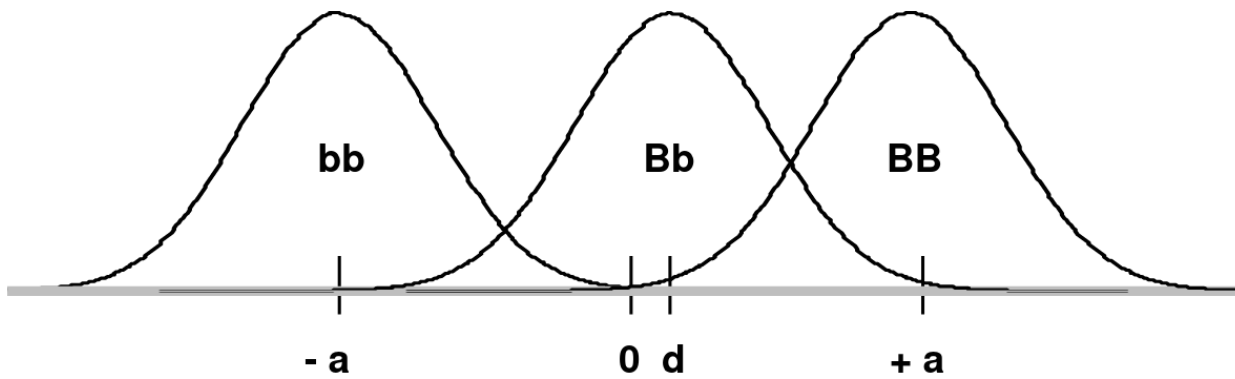
$$T = (A + D) + E = G + E$$

Assumptions of this model:

- Genetic and environmental factors are uncorrelated
- Standard deviation of trait is the same for all individuals

Additive and Dominance Components

Consider a frequency distribution of trait values for two alleles B and b, where B creates a high trait value and b creates a low trait value on a continuous scale which is shifted so that the midpoint between the mean of BB (+a) and bb (-a) is 0:



d is the mean of the Bb group

In an **additive** model $d = 0$ (no dominance; dominance variance = 0)

In a **recessive model**: $d = -a$ (Bb would overlap the bb distribution)

In a **dominant model** $d = +a$ (Bb would overlap the BB distribution)

The **degree of dominance** can be defined as d/a

Heritability

The **heritability** of a trait is the proportion of total phenotypic variance that is due to genetic effects.

Heritability can be defined as:

$$h^2 = \sigma_G^2 / \sigma_T^2 = (\sigma_A^2 + \sigma_D^2) / (\sigma_A^2 + \sigma_D^2 + \sigma_E^2)$$

The above formula is also called Broad sense heritability. Narrow sense heritability (just the additive component):

$$h_n^2 = \sigma_A^2 / \sigma_T^2$$

We use the expected resemblance among relatives to estimate h^2 . It is a function of covariance between relatives and coefficient of relationship (AKA "Additive coefficient")

The additive coefficient of a relationship C is the expected proportion of alleles shared IBD by a relative pair, defined as:

$$C = 2^{-R}, \text{ where } R \text{ is the degree of relationship}$$

- $R = 0$: MZ twins $\rightarrow C = 1$

- R = 1: 1st degree relationship; sib, parent-offspring -> C = 1/2
- R = 2: 2nd degree relatives: half-sibs, grandparent-grandchild, avuncular
- R = 3: 1st cousins

Relationship Class	Relationship	Coefficient of relationship ("Additive coefficient")	Dominance coefficient
	MZ twins	1	1
First degree	Parent-Offspring	$\frac{1}{2}$	0
	Full sibling	$\frac{1}{2}$	$\frac{1}{4}$
	DZ twins	$\frac{1}{2}$	$\frac{1}{4}$
Second degree	Half-sibling	$\frac{1}{4}$	0
	Grandchild-Grandparent	$\frac{1}{4}$	0
	Uncle(Aunt)-Niece(Nephew) Avuncular	$\frac{1}{4}$	0
	Double first cousins	$\frac{1}{4}$	$\frac{1}{16}$
Third degree	Great grandchild-Great grandparent	$\frac{1}{8}$	0
	First Cousins	$\frac{1}{8}$	0

Recall sharing a allele **Identical-By-Descent** means relatives who share the exact same copy of an allele by inheritance.

Relative Pair Type	Probability X Alleles shared IBD		
	X=0	X=1	X=2
Monozygotic twins	0	0	1
Siblings	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Parent-child	0	1	0
Grandparent-Grandchild	$\frac{1}{2}$	$\frac{1}{2}$	0
Half-siblings	$\frac{1}{2}$	$\frac{1}{2}$	0
Avuncular	$\frac{1}{2}$	$\frac{1}{2}$	0
Cousins	$\frac{3}{4}$	$\frac{1}{4}$	0
Double first cousins	$\frac{9}{16}$	$\frac{6}{16}$	$\frac{1}{16}$

The additive coefficient is expected proportion of alleles shared IBD by the pair, so we can also define it as

$$\sum_{i=0}^2 x_i * p(x_i)$$

$p(x) = x/2$ = the proportion of shared alleles

For example, the parent child relationship would have a additive coefficient of $0*(0/2) + 1*(1/2) + 0*(2/2) = 1/2$

The **kinship coefficient** is the probability that a randomly selected pair of alleles from a individual is IBD. It is always half of the coefficient of relationship.

Estimating Heritability

Using Co-variance

Recall the properties of covariance:

$$\text{Estimate of cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, X) = \text{var}(X)$
- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
- $\text{Cov}(cX, Y) = c \cdot \text{Cov}(X, Y)$; where c is some constant
- The unit of covariance is xy
- Positive covariance: Value of x tends to be high when value of y is high
- Negative covariance: Value of x tends to be high when value of y is low

A standardized measure of covariance is correlation which is a scale of -1 to 1

$$\text{cor}(x, y) = \frac{\text{cov}(x, y)}{\text{sd}(x)\text{sd}(y)}$$

We can use the properties of covariance to determine the covariance between quantitative measure on a parent and an offspring. This follows the same $T = (A + D) + E = G + E$ concept as above:

$$\text{Cov}(\text{Parents}, \text{Offspring}) = \text{Cov}(A_p + D_p + E_p, A_o + D_o + E_o)$$

In this example, we can break this down into:

- $\text{Cov}(A_p, A_o) =$ Additive covariance, offspring always inherits exactly 1 allele from parent so $= (1/2)\sigma_A^2$
- $\text{Cov}(E_p, E_o)$, covariance of environmental component between parent and offspring is assumed to be 0
- $\text{Cov}(D_p, D_o) =$ Since offspring do not inherit pairs of alleles the dominance environmental component is also 0
- The cross terms, $\text{Cov}(A, D)$ and etc. which we assume to be 0

Note: This only works if the assumptions of independence is true and standard deviation is the same for all individuals.

Since we are assuming standard deviation is the same across generations we can write heritability between parent and offspring as a product of correlation:

$$h_n^2 = \sigma_A^2 / \sigma_T^2 = 2 * \text{Cov}(\text{parent}, \text{offspring}) / \text{SD}(\text{parent}) * \text{SD}(\text{child}) = 2 * \text{Cor}(\text{parent}, \text{offspring})$$

Thus, we can estimate narrow-sense heritability using 2 times the observed correlation between parents and offspring.

Using Linear Regression

Alternatively, we can use the linear regression coefficient to estimate heritability. Where Y is offspring and X is parent:

$$Y = \mu + \alpha + \epsilon$$

$$\hat{\alpha} \sim \text{cor}(X, Y) \times \text{SD}(Y) / \text{SD}(X)$$

$$\hat{\beta}_{PO} = \text{cor}(\text{parent}, \text{offspring}) \times \frac{\text{SD}(\text{offspring})}{\text{SD}(\text{parent})} \stackrel{\text{(assumed equal)}}{=} \text{cor}(\text{parent}, \text{offspring}) = r_{PO}$$

$$h_n^2 = 2 * r_{PO} = 2 * \hat{\alpha}_{PO}$$

Note that the estimate obtained from beta may differ from the estimate obtained from r if the standard deviations of parents and offspring are not equal.

If both parents are available we can regress the offspring value of the mean parental phenotype value:

$$\text{offspring} = \alpha + \beta_{MO} (\text{average parent}) + E$$

$$h^2 = \hat{\alpha}_{MO}$$

Beta estimates heritability directly in the average parent version. I'm not going to show the math here, but know this is in part due to $\text{SD}(\text{average parent}) = \text{SD}(\text{parent}) / \sqrt{2}$

Example with Sibling Pairs

Looking at the relationship table above in the siblings row: additive coefficient is 1/2 and dominance coefficient 1/4. If data is available on N sibling pairs only:

$$\text{Cov}(\text{Sibling 1}, \text{Sibling 2}) = (1/2) \sigma_A^2 + (1/4) \sigma_D^2$$

If $\sigma_D^2 = 0$, 2 times the intraclass correlation (ICC) is an estimate of heritability:

$$h^2 = 2 \times ICC = 2 \times \frac{\sum_{i=1}^n (x_{1i} - \bar{x})(x_{2i} - \bar{x})}{(N-1)SD(x)^2}$$

Where \bar{x} and $SD(x)$ are computed combining data on all siblings.

If σ_D^2 does not equal 0, this estimate is between the narrow and broad heritability because:

$$2 \text{ Cor (Sibling 1, Sibling 2)} = \frac{2 \times \left(\frac{1}{2} \sigma_A^2 + \frac{1}{4} \sigma_D^2 \right)}{\sigma_T^2} = \frac{\sigma_A^2 + \frac{1}{2} \sigma_D^2}{\sigma_T^2}$$

Intraclass Correlation Vs. Pearson (Product-Moment) Correlation

When pairs consist of individuals of two different classes (grandparent-grandchild, parent-offspring) we call this **pairwise correlation** and we can use a simple Pearson correlation coefficient:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

But when the pairs have no obvious order (siblings or cousins), the intraclass correlation is used:

$$ICC = \frac{\sum_{i=1}^n (x_{1i} - \bar{x})(x_{2i} - \bar{x})}{\frac{n-1}{2n-1} \sum_{i=1}^n \sum_{j=1}^2 (x_{ji} - \bar{x})^2}$$

where n is the number of pairs and \bar{x} is the mean value across all individuals

The main difference between the product-moment correlation and ICC:

- ICC: SD in the denominator is pooled across individuals
- Pearson: SD for the two types of relatives are computed separately

ICC can be extended to sets of 3 or more. The formula for ICC is:

$$ICC = \frac{MSB - MSW}{MSB + (k - 1)MSW}$$

k = number of siblings per sibship

MSB = Mean Square Between (model)

MSW = Mean Square Within (error)

If the sibships are of unequal size using the average sibship size will provide an approximate h^2 estimate

$$h^2 = 2 * ICC$$

Twin Studies

Monozygotic twins are genetically identical and dizygotic twins are genetic the same as a full sibling, sharing half their genes on average. A common estimate of heritability in studies with both mono and dizygotic twins can estimate it as:

$$h^2 = 2 * (ICC_{MZ} - ICC_{DZ})$$

If σ_D^2 does not equal 0, this estimate is greater than both the narrow and broad heritability because:

$$2 [\text{Cor}(MZ_1, MZ_2) - \text{Cor}(DZ_1, DZ_2)] =$$

$$2 \times \left(\frac{(\sigma_A^2 + \sigma_D^2)}{\sigma_T^2} - \frac{\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2}{\sigma_T^2} \right) = 2 \times \left(\frac{\frac{1}{2}\sigma_A^2 + \frac{3}{4}\sigma_D^2}{\sigma_T^2} \right) = \frac{\sigma_A^2 + \frac{3}{2}\sigma_D^2}{\sigma_T^2}$$

This is in part because we cannot assume there is no shared environmental variance in siblings and especially twins.

Notes on Estimating Heritability

- We can obtain different estimates using different subsets of data
- Differences in estimates may be due to "sampling variation"
- If the dominance variance is not 0, some differences may be expected
 - Siblings and twins incorporate dominance variance; parent-offspring do not
- Siblings and twins tend to share environment/exposure factors
- We may get smaller estimates with pairs that are more similar (Age, weight, etc)

- Precision of heritability estimates depends on the SE of the correlation coefficient or regression coefficient used for the estimation
- Large sample sizes are needed to get precise measures of heritability
- Heritability is a ratio of variances, not an average
- Law of large numbers does not necessarily apply here, heritability does not necessarily follow a normal distribution
- Maximum likelihood approach can be used to estimate σ^2_T , σ^2_A , and σ^2_D -- and can also estimate shared environments when we are not assuming it to be 0

We can estimate the linear mixed effects (variance component model):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + G + e_i \quad i = 1, \dots, n$$

$$\text{cov}(Y_i, Y_j) = \begin{cases} \sigma_G^2 + \sigma_e^2 & \text{if } i = j \\ \phi_{ij} \sigma_G^2 & \text{if } i \neq j \end{cases}$$

β : fixed effects, X : covariates, G : random effect, e : individual error

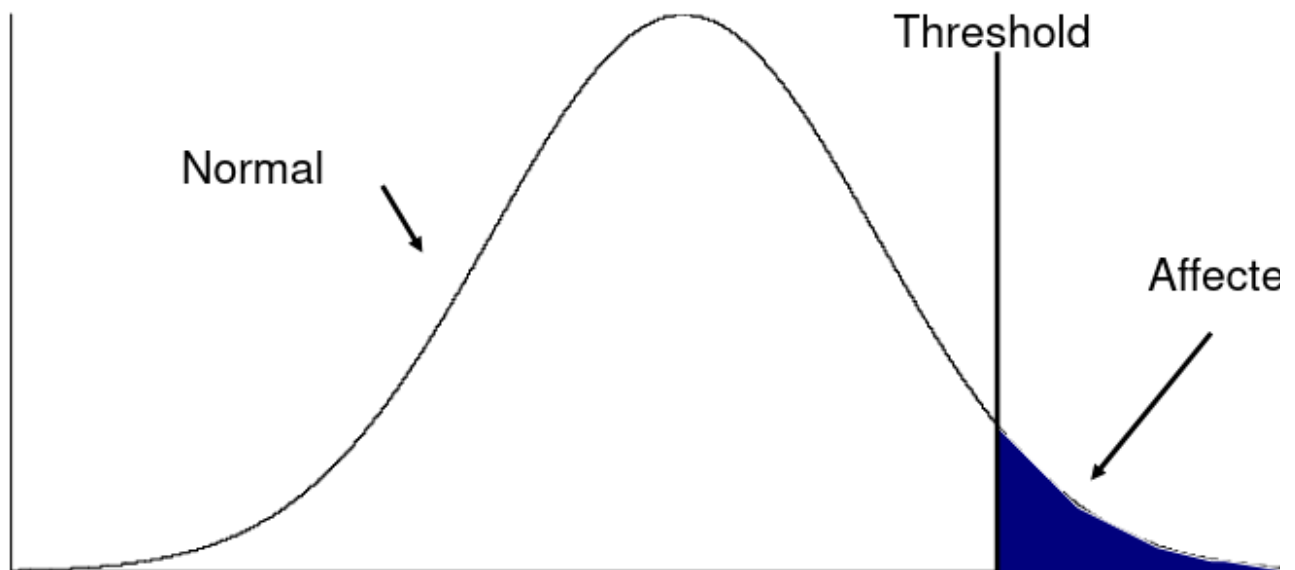
ϕ_{ij} : coefficient of relationship between individuals i and j (=0 if unrelated)

In which case we would estimate heritability as σ^2_G and test $H_0: \sigma^2_G = 0$

Heritability Summary

- Heritability is a population parameter, defined as the proportion of trait variance explained by additive genetic factors
 - Estimates of h^2 are usually based on resemblance between relatives
 - Highly heritable \rightarrow close relatives should be more correlated
- Heritability may differ by population
 - h^2 depends on both σ^2_T and σ^2_A
- h^2 estimates using close and distant relatives should be similar
 - Estimates based on twins, sibs and cousin pairs estimate the same population parameter (unless there is dominance variance)
- If a trait is not heritable ($h^2 = 0$), we do not expect close relatives to be more correlated than distant relatives
- Heritability is generally computed for quantitative traits with analysis of variance, but can also be estimated for discrete traits
 - Liability, or predisposition to disease, is measured on a quantitative scale
 - Threshold model: trait present if liability to disease exceeds a certain threshold
 - Heritability measures the liability between relatives

- There is specialized software for calculating this



- Referred to as the "latent variable" model, we can observe whether an individual has cross the threshold

Recurrence Risk Ratio

Recurrence risk ratio assess familial aggregation of dichotomous traits (diseases). λ_R denotes relative risk with a subscript indicating the specific type of relative examined.

A **proband** is a subject selected for a study due to disease status. In familial disease studies, it is the affected subject that brought the family into the study

$$\lambda_R = \lambda_{R|K} / \lambda_K$$

$\lambda_{R|K}$ = P(disease | relative of type R is affected) = the proportion (prevalence) of relatives of affected probands that are affect with disease

K = the proportion of affected individuals in the general population

General rule:

- If a trait is generic λ_R should decrease as degree of relationship increases
- Differences in shared environment may lead to differences in λ_R for different relative types of the same degree
- λ_R may be greater than 1 when the trait does not have a genetic component due to a shared enviornment

Revision #13

Created 27 September 2022 00:54:53 by Elkip

Updated 29 September 2022 17:33:37 by Elkip