QUANTITATIVE TRAITS

Genetic Model for Trait

Suppose gene **T** affects a trait: its genotype may affect the probability an individual has a disease or its genotype may affect the value of some measurable quantity. There may be other genes also affecting the trait, and there may be non-genetic effects.

No suppose G is the genotypic effect of **T** on the trait and E is the environmental effect (or all other effects). An individual is observed to have phenotypic (trait) value (disease status or measured value) Y and a simple linear model is

$$Y = G + E$$

The mean environmental effect is taken to be zero, so the mean phenotypic effect is equal to the mean genotypic value.

Genetic Model for Trait

"If we could replicate a particular genotype in a number of individuals and measure them under environmental conditions normal for the population, their mean environmental deviations would be zero, and their mean phenotypic value would consequently be equal to the genotypic value of that particular genotype. This is the meaning of the genotypic value of an individual."

Falconer DS. 1960. Introduction to Quantitative Genetics. Ronald Press, New York. p. 113

Genetic Model for Trait

Extensions to this model can include $G \times E$ interaction, but at present G and E will be considered independent and their variances sum to the variance of Y:

$$\text{Var}(Y) = \text{Var}(G) + \text{Var}(E) \quad \text{or} \quad \sigma_Y^2 = \sigma_G^2 + \sigma_E^2$$

In general, the number of alleles for gene **T** is not known, but a convenient start is to suppose there are two: an ancestral form and a more recent form that may increase the chance of being affected or lead to detrimental values of a measured trait. Write the two alleles as T, t and the three genotypes as TT, Tt, tt. The three genotypic values are G_{TT}, G_{Tt}, G_{tt}.

Additive and Dominance Variance

In a population there is a mean genotypic effect, μ_G, and a variance of genotypic effects, σ^2_G:

$$\begin{aligned}\mu_G &= \pi_T^2 G_{TT} + 2\pi_T\pi_t G_{Tt} + \pi_t^2 G_{tt} \\ \sigma^2_G &= \pi_T^2(G_{TT} - \mu_G)^2 + 2\pi_T\pi_t(G_{Tt} - \mu_G)^2 + \pi_t^2(G_{tt} - \mu_G)^2\end{aligned}$$

and the variance can be partitioned into additive and dominance components:

$$\sigma^2_G = \sigma^2_{A_T} + \sigma^2_{D_T}$$

$$\sigma^2_{A_T} = 2\pi_T\pi_t[\pi_T(G_{TT} - G_{Tt}) + \pi_t(G_{Tt} - G_{tt})]^2$$

$$\sigma^2_{D_T} = \pi_T^2\pi_t^2(G_{TT} - 2G_{Tt} + G_{tt})^2$$

Additive Traits

If the genetic value of trait heterozygotes is the average of the values of the two trait homozygotes, $G_{Tt} = (G_{TT} + G_{tt})/2$ then

$$\sigma^2_{A_T} = 2\pi_T\pi_t(G_{Tt} - G_{tt})^2$$

$$\sigma^2_{D_T} = 0$$

and the genetic variance is entirely additive.

If the population has only one of the two trait alleles, $\pi_T\pi_t = 0$ and there is no genetic variance. Otherwise, additive genetic variance is maximized when the two trait alleles are equally frequent.

Heritability

A convenient single parameter to describe the trait genetic variance *in a particular population* is the heritability h^2 defined as

$$h^2 = \frac{\sigma^2_{A_T}}{\sigma^2_Y}$$

or the proportion of phenotypic (trait) variance due to additive allelic effects.

The phenotypic variance, the genetic variance and the additive and dominance variance components all depend on trait genotypic (or allele) frequencies and so are different in different populations.

The genotypic effects G are not known but the variance components and heritability can be estimated.

Trait Mean in Inbred Populations

Finding the mean and variance for quantitative traits in populations where there is inbreeding and/or relatedness and Hardy-Weinberg equilibrium does not hold, requires modification of genotype probabilities. For inbred populations, the mean trait value requires the inbreeding coefficient. For a random member of a population inbred to an extent F (relative to a reference population), the genotype probabilities are

$$\begin{aligned} P_{TT} &= \pi_T^2 + F\pi_T\pi_t \\ P_{Tt} &= 2\pi_T\pi_t(1-F) \\ P_{tt} &= \pi_t^2 + F\pi_T\pi_t \end{aligned}$$

Trait Mean in Inbred Populations

The expected trait value μ_F in an inbred population is

$$\mu_F \;=\; \mu_0 + FH$$

where μ_0 is the value in a HWE population ($F = 0$) and $H = \pi_T\pi_t(G_{TT} - 2G_{Tt} + G_{tt})$ is a measure of dominance.

Inbreeding Depression

Recent paper by Clark DW, et al. 2019. Nature Communications. Published October 31, 2019:

From data on 1.4 million individuals, e.g. "F_{ROH} equivalent to the offspring of first cousins is associated with a 55% decrease in the odds of having children."

Note the need for estimation of individual inbreeding coefficients.

Clark et al., 2019

Nearly one billion people live in populations where consanguineous marriages are common.

Burden of disease thought to be disproportionately due to increased homozygosity of rare, recessive variants.

The fraction of each autosomal genome in ROH $>$ 1.5 Mb correlates well with pedigree-based estimates of inbreeding.

Genetic Variance and Covariance in Inbred Populations

For individual U, the genetic variance for an additive trait is

$$\sigma^2_{G_U} = (1 + F_U)\sigma^2_A$$

For individuals U, V, the genetic covariance for an additive trait is

$$\text{Cov}_{G_{UV}} = 2\theta_{UV}\sigma^2_A$$

regardless of inbreeding.

Total Variances and Covariances for Additive Trait

Trait values have both genetic and environmental components. The simplest model of $Y = G + E$ leads to the variance of trait values Y among individuals U in a non-inbred population of unrelated individuals:

$$\text{Var}(Y_U) = \sigma^2_A + \sigma^2_E$$

This is also referred to as the phenotypic variance σ^2_P.

For an additive trait and for individuals that have no shared environment, the variance-covariance matrix for a sample of related and inbred individuals has elements

$$\begin{aligned} \text{Var}(Y_U) &= (1 + F_U)\sigma^2_A + \sigma^2_E \\ \text{Cov}(Y_U, Y_V) &= 2\theta_{UV}\sigma^2_A \end{aligned}$$

Genetic Relationship Matrix

Vector $\boldsymbol{Y}$ of trait values for individuals $i = 1, 2, \ldots n$ has GRM

$$\boldsymbol{G} = \begin{bmatrix} (1+F_1) & 2\theta_{12} & \ldots & 2\theta_{1n} \\ 2\theta_{21} & (1+F_2) & \ldots & 2\theta_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ 2\theta_{n1} & 2\theta_{2n} & \ldots & (1+F_n) \end{bmatrix}$$

The trace of this matrix is

$$\text{tr}(\boldsymbol{G}) = \sum_{i=1}^{n} G_{ii} = n(1+F_W)$$

and the sum of the off-diagonal elements is

$$\Sigma_{\boldsymbol{G}} = \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} G_{ij} = n(n-1)\theta_S$$

to define the average inbreeding and kinship values F_W, θ_S for the sample.

GRM

Historically, the GRM was obtained from known pedigrees, particularly for experimental populations of plants and animals.

Beginning with Yu J, et al. 2006. Nature Genetics 38:203 it has been recognized that pedigree information may not be available, it may not be accurate, and it can be different from the "gold standard" GRM.

Instead, the GRM may be constructed with estimated inbreeding and kinship coefficients.