

Estimating heritability with simple experimental design:

- We have seen that additive values of individuals are primarily responsible for the resemblance between relatives and can be quantified by heritability $h^2 = V_A/V_P$
- Simple design: pairs of parents and offspring (estimates obtained by means of coefficient of regressing offspring values onto P)
- Or groups of full sibs or half sibs (estimates obtained by intraclass correlation coefficient)
- ICC: how strongly units in the same group relate to each other, commonly used in QG to quantify the degree to which individuals with fixed relatedness resemble each other
- Estimation based on degree of resemblance between parents and offspring
 - Mean of the trait of an offspring is by definition half of additive value A , so $cov(\bar{O}, P) = \frac{V_A}{2}$
 - From above it can be derived that the regression of the mean of phenotypic values of offspring on their parents is an estimate of half the heritability of the evaluated trait
 - with both sexes, we take \bar{P} as the average phenotypic value between parents, and the regression of mean of offspring on parental average provides a measure of heritability
 - Maternal effects: a cause of resemblance between parents and offspring that is influenced by maternal environment including body weight and milk production, and especially for behavioural traits in humans
 - Variance for heritability is minimized when only one offspring per family is evaluated and is approximated by $\sigma_b^2 \approx \frac{k}{n}$ where k is the number of parents and n is the number of data pairs
 - Standard error of the estimate can be reduced by taking parents with phenotypic values at the extremes of a distribution because this has a stronger effect on the regression
- Estimation based on resemblance between siblings
 - covariance in a trait between half sibs is a quarter, and the intraclass correlation is $t_{HS} = \frac{V_A/4}{V_P} = \frac{h^2}{4}$
 - Variance is increased when half or full sibs grow in the same environment
 - For full sibs, the intraclass correlation is $t_{HS} = \frac{h^2}{4} + 1/4d^2 + c^2$, where d^2 represents the proportion of phenotypic variance due to dominance variance, and c^2 is the proportion of variance due to the common environment
 - common environmental variance can imply an increase or decrease in resemblance between siblings
 - The error for heritability estimates by groups of full sibs can be approximated by $\sigma^2 \approx 2(1 + jt)^2/nj^2$
 - The standard error of the estimate of heritability is inversely proportional to the square root of the number of families analyzed
- Assortative mating:
 - Regression of average value of parents on their progeny is not affected by assortative mating
- Estimating based on degree of similarity between twins
 - Monozygotic twins have variance between pairs that includes genetic variance and common environmental variance
 - Within a pair, only environmental variance exists
 - In humans, things like cell activity, reproduction and social interactions have low heritability and high common environments, skeletal, dermatological and pthalmological characters have high average heritabilities
- Coefficients of additive and dominance relationships

- The correlation of additive values between two individuals $r = 2f_{xy}\sqrt{(1 + F_X)(1 + F_Y)}$
- The correlation of dominance values between two individuals with parents A,B and C,D is $u = f_{AC}f_{BD} + f_{AD}f_{BC}$
- For monozygotic, $r = 1$, $u = 1$
- for full sibs, $r = 0.5$ and $u = 0.25$
- for half sibs $r = 0.25$ and $u = 0$
- for parents and offspring, $r = 0.5$ and $u = 0$

Estimation of genetic correlation:

- correlations can be estimated for traits X and Y for phenotypic, additive genetic and environmental correlations
- All 3 correlations can be incorporated into the expression $r_p = r_A h_X h_Y + r_E \sqrt{(1 - h_x^2)(1 - h_y^2)}$
- Estimate of genetic correlation $r_A = \frac{cov_A(X,Y)}{\sqrt{V_{AX}V_{AY}}}$
- Experimental designs that minimize the sampling variance of the heritability will minimize error of correlation estimates

Estimating variance and predicting additive values with complex structure

- the ML estimate for μ is \bar{y}
- the ML estimate for $\sigma^2 = V + (\bar{y} - \mu)^2$
- ML estimates for variance do not account for loss in degrees of freedom when estimating fixed effects, so a correct is applied to produce restricted ML (REML)
- REML estimation and the animal model: $y = Xb + Za + e$, where b is the vector of fixed effects, X is the incidence matrix, a is the vector of additive effects of individuals, Z is the incidence matrix, and e is the vector of residual errors
- The additive genetic covariance between any two individuals is $r\sigma_A^2$, where r is the numerator of the coefficient of additive relationships between individuals (twice their coancestry, $2f$)
- The variance-covariance matrix G includes all additive covariances between individuals and is given by $A\sigma_A^2$ where A is the matrix with the numerator of the coefficient of additive relationships between individuals
- Matrix A is obtained by multiplying the coancestry values between all individuals by 2 (which was obtained using the tabular method)
- Matrix R is the residual variance-covariance matrix which reduces to $I\sigma_A^2$
- The matrix of phenotypic variances-covariances becomes $P = ZGZ' + I\sigma_R^2$, and the vector y in the animal model contains both vectors of means (Xb) and the matrix of phenotypic variances-covariances P
- The ML estimator can be derived as $\ln L(b, P|y, Xb) = c - 1/2 \log|P| - 1/2(y - Xb)'P^{-1}(y - Xb)$, where $|P|$ denotes the determinant
- The basic animal model can be extended to other random effects (environmental, indirect genetic effects such as social interactions), maternal effects. This would entail adding new terms of random effects to the expression.
- Animal model can also be extended to dominance effects, and also to detect G by E interactions
- Variance estimates are not affected by biases from finite census size, assortative mating, selection or inbreeding because the matrix of additive relationship factors account for these
- Predicting additive values using BLUP
 - BLUP allows for estimating the additive values correcting for any type of fixed effects and uses all information
 - The BLUP estimate for the prediction of additive values is: $\hat{a} = cov(y', a)P^{-1}(y - Xb) = AZ'\sigma_A^2[ZGZ' + R]^{-1}(y - Xb)$

- Estimating fixed effects by using generalized least squares is $\hat{b} = (X'P^{-1}X)^{-1}X'P^{-1}y$
- Using molecular coancestries (with genetic markers) can estimate genetic variance components with the expression $\hat{h}^2 = \frac{cov(Z, f_m)}{2var(f_M)}$
- Can add term Zq to the animal model expression where q is a vector of additive effects of individuals associated with the marker, then use likelihood ratio tests to determine if the chosen marker is significant
- comparing estimates of heritability using genealogical and molecular data: reflect with greater reality the degree of coancestry but are subject to lower precision

Problems

- (1) The aim is to carry out an experiment to estimate the heritability of a trait using pairs of parents and offspring. It is expected that the estimate of heritability will be around 0.6. It is intended to evaluate only one offspring per couple and a single parent or both. How many pairs of data would have to be evaluated in each case to obtain an estimate of heritability with a standard error equal to or less than 0.05?

$$SE(h^2) = \frac{2}{\sqrt{n}}$$

$$n = \left(\frac{2}{SE(h^2)}\right)^2 = \left(\frac{2}{0.05}\right)^2 = 1600$$

$$SE(h^2) = \sqrt{\frac{2}{n}}$$

$$n = \left(\frac{\sqrt{2}}{SE(h^2)}\right)^2 = \left(\frac{\sqrt{2}}{0.05}\right)^2 = 800$$

- (2) An analysis of families has provided a value of the phenotypic correlation between full sibs of $t_{FS} = 0.12$. (a) What is the estimate of heritability that can be obtained with this data? (b) It is later discovered that the parents of the families did not mate randomly but with positive assortative mating for the character under study, having estimated that the phenotypic correlation between the individuals of pairs is $\rho = 0.5$. How is the estimate of heritability modified? (c) If the phenotypic correlation between pairs were maximal, what would be the value of the heritability?

$$t_{FS} = 0.5h^2 + 0.25d^2 = c^2$$

$$t_{FS} = 0.5h^2$$

$$2t_{FS} = h^2 = 2(0.12) = 0.24$$

$$t_{FS} = 0.5h^2(1 + \rho h^2)$$

$$t_{FS} = 0.5h^2 + 0.25(h^2)^2$$

$$0 = 0.5h^2 + 0.25(h^2)^2 - t_{FS}$$

$$h^2 = x$$

$$0 = 0.5x + 0.25x^2 - t_{FS} = 0.5x + 0.25x^2 - 0.12$$

$$x = \frac{1}{5}(\sqrt{37} - 5) \approx 0.217$$

$$\rho = 1$$

$$0 = 0.5x + 0.25x^2 - t_{FS} = 0.5x + 0.5x^2 - 0.12$$

$$x = \frac{1}{5}$$

- (3) In an analysis of mono (MZ) and dizygotic (DZ) twins, the following components of the variance between pairs (B), within pairs (W) and total (T) have been obtained: $\sigma_B^2 = 0.41, \sigma_W^2 = 0.18, \sigma_T^2 = 0.59$ for monozygotic twins and $\sigma_B^2 = 0.25, \sigma_W^2 = 0.32, \sigma_T^2 = 0.57$ for dizygotic twins. What estimates of genetic components can these data provide?

$$t_{MZ} = \frac{\sigma_B^2}{\sigma_T^2} = \frac{0.41}{0.59} = 0.695$$

$$t_{DZ} = \frac{\sigma_B^2}{\sigma_T^2} = \frac{0.25}{0.57} = 0.439$$

$$2t_{DZ} - t_{MZ} = 0.182 = c^2 - 0.5d^2 \approx c^2$$

$$2(t_{MZ} - t_{DZ}) = 0.513 = h^2 + (3/2)d^2 \approx H^2$$

- (4) The weights P_i of for mice are available: A (16.6g) B (22.1g) C(18g) D(12.4g). Using genetic markers, we have molecular coancestries $f_{M,ij}$ with coancestry table provided in the question. Estimate the heritability of weight.

See the associated R file for this chapter for the answer to this question.

- (5) The aim is to estimate the heritability of a human trait using a design of full-sib pairs, estimating the genomic relationships with molecular markers. How many pairs should be evaluated to obtain a standard error of the heritability estimate equal to or less than 0.1?

$$SE[h^2] = \sqrt{\frac{2}{n * var(r_M)}}$$

$$L = 35$$

$$var(r_M) = 1/16L - \frac{1}{3L^2} = 0.0015$$

$$n = \frac{2}{(SE[h^2])^2 * var(r_M)} = \frac{2}{0.1^2 * 0.0015} = 133,333$$

Self Assessment

- (1) The estimate of heritability obtained from the regression of offspring values on that of their parents will necessarily be biased if the latter are not a random sample of the population.
False, if the group of parents were selected, the regression estimates would not be affected if the bias would occur to the same extent in the numerator and the denominator of the equation $b_{OP} = \frac{cov(O,P)}{\sigma_P^2}$
- (2) The variance of the means of full-sib families is equal to the covariance of the values of the members of the families.
- (3) The standard errors of heritability estimates obtained from full-sib families are always lower, the lower the value of the heritability. False, the standard error of heritability estimates is the same for full or half sibs and is approximated by $h^2 \sqrt{\frac{2}{n}}$, and is only inversely proportional to the square root of the number of families analyzed.
- (4) For a given total number of evaluated individuals, the estimates of heritability by the regression have less error if only one offspring per family is evaluated.
True, it is stated that of the total number of offspring is fixed, then σ_b^2 is minimal if only one offspring per family is evaluated, which then is approximated by k/n where k is the

number of parents and n is the number of data pairs.

- (5) The coefficient of dominance relationship between two cousins is $1/8$.
False, it should be 0 because this coefficient only exists when offspring X and Y share both parents. Cousins only share 1 parent.
- (6) The phenotypic correlation is the sum of the genetic and environmental correlations.
False, the phenotypic correlation is the sum of the covariance of genetic correlation and environmental correlation all divided by the product of standard errors of each correlation.
- (7) The estimates of variance obtained by ML are biased because they do not take into account the degrees of freedom lost when the fixed effects are estimated.
True, this is the reason why REML is made.
- (8) The prediction of additive values by BLUP is not biased by the fact that there is selection in the evaluated population.
True, in BLUP the variance of the additive value of an individual is equal to the additive variance of the base population modulated by inbreeding coefficient, so it accounts for selection, non-random mating or changes in population census size.
- (9) Given that the estimates of heritability obtained by twin studies include dominance and epistatic components of variance, this may contribute to the fact that estimates provided by this method are higher than those obtained by means of molecular markers.
True, this could be a reason why. The reason in the text was that the loci that control a trait do not have a large enough effect, or LD between causal and SNPs is not strong enough (due to being found at low frequencies in the population).
- (10) To obtain an estimate of heritability with a standard error of approximately 0.05 using molecular markers in pairs of full sibs in humans, of the order of 500,000 pairs would be needed.
True, use the equation on page 145 to get an answer of 490k.