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 vales 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 $lnL(b, P|y, Xb) = c - 1/2log|P| - 1/2(y-Xb)'P^{-1}(y-Xb)$, 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 effeects 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?

$$N_e = \frac{4N_fN_m}{N_m + N_f} = \frac{4(4)16}{20} = 12.8$$

(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 tha tcan 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?

$$1 - F_{IT} = (1 - F_{IS})(1 - F_{ST})$$

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

$$N_{autosomal} = \frac{4N_mN_f}{N_m + N_f} = \frac{4(16)32}{16 + 32} = 42.67$$

(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.

$$C^2 = Us = 0.2(0.1) = 0.02$$

(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?

$$S^2_K = Var(1, 3, 4, 0, 2, 0, 5, 1) = 3.429$$

Self Assessment

(1) The different types of effective size (variance, inbreeding, eigenvalue, coalescence, etc. usually coincide exactly or approximately in their asymptotic value.
True: the text states that except in very specific situations, the predictions of effective population size coincide, or only differ in second order terms.

(2) To average different population sizes, the harmonic mean is usually used, because the effective size usually affects the denominator of the expressions in which it is found.
True, the text states that when we want to average (effective) population sizes for predicting their impact on drift or inbreeding, we use the harmonic mean, because small values have the most relevance and are in the denominator in these expressions.

(3) In populations with a certain percentage of matings between relatives, the variance of the contributions of parents to progeny decreases with respect to that corresponding to a panmictic population.
False, Hardy-Weinberg disequilibrium implies a decrease of heterozygotes and increase of homozygoytes. The increase of homozygotes corresponds to an increase in the variation between parental contributions by a factor of $1 + \alpha$. This results in a greater variance in allele frequencies, because homozygotes only contribute one allele.

(4) The effective size decreases as the generation interval increases in populations with overlapping generations.
False, effective size in overlapping generations is linearly proportional to generational interval $N_e \propto I_g$, so an increase in generational interval would increase N_e.

(5) The magnitude of the genetic drift that affects the genes of the X chromosome in XX-XY species or the Z chromosome in ZZ-ZW species is 25 percent less than that of autosomal genes.
Assuming an equal number of males and females, then true. $N_e = 3N/4$. The expression of sex-linked genes is given greater weight in the heterogametic sex. This could be false? My answer deals with N_e, not drift.

(6) The effective size referring to neutral genes is drastically reduced when there is linkage between these and other loci subjected to selection.
True, the cumulative effect of selection produced by mutations on neutral genes linked to genes that are selected can be approximated by $Q_c \approx \frac{1}{s+c}$. As linkage becomes greater, the percent recombination c increases which increases the cumulative effect.

(7) The effective size of a subdivided population always increases with the differentiation in allele frequencies between sub-populations.
True, the effective size of a subdivided population $N_e = \frac{Nn}{1-F_{ST}}$. As differentiation increases (approaches 1), N_e approaches infinity. Reasoning: there is little to no genetic drift because if sub-populations remain isolated, different allelic variants could become fixed in the different subpopulations without being lost, resulting in a high variance effective size. This scenario is only true if sub-populations contribute identically to the offspring in each generation.

(8) With equal contributions from parents to offspring, if mating between relatives is forced, the long-term effective size increases in comparison with the panmictic scenario.
True, the thinking is along the same lines as question 7.

(9) The demographic methods of estimating the effective size tend to produce underestimates, by not taking into account all possible sources of genetic drift in the population.

False, demographic methods have the advantage of incorporating sources like genetic drift and inbreeding.

(10) The larger the effective population size, the larger is the expected linkage disequilibrium between two closely linked loci.
False, linkage and N_e are inversely proportional as shown by: $N_e \approx Nexp\frac{-U}{s+(L/2)}$, where L is the linkage in the chromosomal segment. So a larger N_e must arise from lower L.