Marriage and Friendship

In general, the correlation between the husband and the wife is

in which and are indicator for the husband and the wife, and is the genetic effect. Depending on the context, can be summary statistics or true effects .

For convenience, the phenotype and genotype have both been standardized.

**Scenario 1: predicted values vs predicted values**

If is estimated as provided in summary statistics, . is the number of QTLs, is the additive genetic variance, is averaged LD square between a QTL and its surrounding markers, is the phenotypic variance, is the variance of the marker, and is the sample size of the discovery set.

Let’s call the marriage cohort as the testing set.

Assume for each locus, the frequency between husbands and wives are same

in which is the genetic relatedness between husbands and wives, is the genetic correlation between husbands and wives, and is quasi-Fst because it may be positive or negative.

1) We can assume no kinship between husbands and wives, and then .

2) can be positive or negative, depending on the context. For example, in the sample if blue eyes like to marry blue eyes, and brown eyes to brown eyes, indicating positive genetic correlation; if blue eyes likes to marry brow eyes, indicating negative genetic correlation.

3) is quasi-Fst. For example the sample are Swedish and Greek subgroups. If Swedish marry Swedish, Greek marry Greek, ; if Swedish always marry Greek in the sample .

These three setting have been used in my commentary to “[Friendship and nature selection](http://journal.frontiersin.org/Journal/10.3389/fgene.2014.00400/full)”, but the argument for is put on to the context Swedish find Swedish / Greek as friends.

plug it into ,

(Equ 1)

always, but given increased p-value cut-off, this term is getting smaller.

Inference:

1) Assume the marriage is random across the nations, .

If marries completely depends on the match up of the height, will increase if include more SNPs. However, it will reach plateau somewhere (say p-value cut-off 0.1) because the sample size of the testing set will blur , the magnitude of which depends on how big the effect is, and eventually gain nothing.

2) If no associative mating , but always see within-nation marriage so that . . It means that if is from a trait not related mating at all, if C increases when includes more markers, it indicated the existence of population stratification.

Testing for population stratification

Step 1: Only select loci which have their summary statistic p-value < 0.5 (can be other values), so as to knock out .

Step 2: pruning, only use markers in linkage equilibrium. markers are remained.

**Scenario 2: predicted values vs observed values**

For example, now it is to calculate the correlation between the predicted values of husbands to the observed values of wives. If is true genetic effects, .

in which is the genetic correlation between summary stats.

Genetic effect

1. Genetic effects can be estimated using population data, which may be subject to population stratification; or, using family-based method to estimate effects. It will help get much better in Equation 1, but it only adjusts the blue part rather than the purple part.
2. If there is associative mating for height, the estimated heritability from their children should be inflated as predicted by the theory (Lynch & Walsh, Equ 4.23c; Lynch &Walsh, Fig 7.7). However, the question is different methods pick up different things. Some estimates , and some methods which will scale the inbreeding part and estimates .
3. Adjustements. After adjustment for PCs, , in which is a value between 0 and 1, and is between 0 and 1. The best scenario will be that after adjustment and . However, PC may reduce both.