

Exam Project

Technical background information on how the data were generated

Genetics

2500 families consisting of

- 2 parents
- 0, 1, 2, 3 or 4 children, with 20% probability each

1000 diallelic genetic markers with

- Minor allele frequencies of 0.5 for markers A, B and C; 0.1 for markers D and E; 0.1 for markers F and F+1; and randomly taken between 0.01 and 0.5 for all other markers
- Hardy-Weinberg equilibrium for all markers
- No Linkage Disequilibrium between any pair of markers except between markers F and F+1
- Linkage Disequilibrium between markers F and F+1 modeled by defining correlation r (taken to be $\sqrt{0.8}$)
- Mendelian transmission of haplotypes for marker pair F, F+1 (without recombination) and of independent alleles for all other markers

Continuous phenotype

Mean model with additive contributions of

- Overall mean (taken to be 0)
- Pure epistasis of markers A and B following model below but centered and scaled such that mean is 0 and a certain percentage (taken to be 5%) of the total phenotypic variance (taken to be 1) is explained by the pair (remark: means and variance are calculated weighted with genotype frequencies, and hence depend on minor allele frequencies)

Prototype			Result		
0	1	0	-0.2236	0.2236	-0.2236
1	0	1	0.2236	-0.2236	0.2236
0	1	0	-0.2236	0.2236	-0.2236

- Pure epistasis of markers B and C following model below but centered and scaled such that mean is 0 and a certain percentage (taken to be 5%) of the total phenotypic variance is explained by the pair

Prototype			Result		
0	0	1	-0.1581	-0.1581	0.4743
0	0.5	0	-0.1581	0.1581	-0.1581
1	0	0	0.4743	-0.1581	-0.1581

- Two-locus model of markers D and E following the same model as A and B but centered and scaled such that mean is 0 and a certain percentage (taken to be 5%) of the total phenotypic variance is explained by the pair (remark: of the total two-locus variance each mean effect amounts to 29% and the remaining 42% is epistatic)

Prototype			Result		
0	1	0	-0.1447	0.3455	-0.1447
1	0	1	0.3455	-0.1447	0.3455
0	1	0	-0.1447	0.3455	-0.1447

- Main effect of marker F following additive model but centered and scaled such that mean is 0 and a certain percentage (taken to be 5%) of the total phenotypic variance is explained by the marker

Prototype			Result		
0	1	2	-0.1826	0.1826	0.5477

- Main effect of marker F+1 following advantage heterozygous model but centered and scaled such that mean is 0 and a certain percentage (taken to be 5%) of the total phenotypic variance is explained by the marker

Prototype			Result		
0	1	0	-0.1732	0.2887	-0.1732

Total phenotypic variance fixed (taken to be 1) and consisting of

- Contributions due to the five non-constant mean model contributions (taken to be 5% each, hence 25% in total): this is not explicitly put in the variance model as it is already incorporated in the mean model (remark: this variance is in fact a population-level effect originating from the different means and the presence of different genotypes in the population, therefore it makes sense not to include it in an individual-level variance model)
- Polygenic variance (the total heritability of the trait is taken to be 60%, hence the polygenes explain 35% of the phenotypic variance): this is modeled in the variance model as $2K\sigma_{poly}^2$ where K is the kinship matrix and σ_{poly}^2 is the polygenic variance
- Environmental variance (explaining the remaining 40% of the phenotypic variance) this is modeled in the variance model as $I\sigma_{env}^2$ where I is the unit matrix and σ_{env}^2 is the environmental variance

	Effects present	Models	SNPS
(A,B)	Pure epistasis, no main effects		(107,304)
(B,C)	Pure epistasis, no main effects	Different epistasis model than for (A,B)	(304,564)
(D,E)	Both interaction and main effects present	Same model as for (A,B)	(285,890)
F	Main effect	Additive model	645
F+1	Main effect	Advantage heterozygous model, SNP in LD with F	646

Assignment

Write an individual report which summarizes all your findings (homeworks 1-5) and discusses the results in the light of the foregoing information about the truth.

Compile your written report around the following items (not exclusively):

- Data quality control: e.g., pedigree structure, loops in data, missing data, Hardy-Weinberg equilibrium, kinships
- Main effects analysis:
 - o Comparison population based and family based approach
 - o Why are there more /less / other main effects seen?
 - o Can the simulated models (additive, advantage heterozygous models) be retrieved? Why or why not?
 - o What is the effect of LD between SNPs on main effects results?
 - o What is the effect of the actual epistasis on main effects results?
- Interaction analysis:
 - o Comparison of methods
 - o Why are there more/less/ other pairs picked up?
 - o Can one see differences in epistasis effects between the pairs found? Can you explain?
 - o What is the effect of LD to discover interacting pairs?
 - o What is the effect of an interacting pair belonging to a higher-order complex interaction system (A,B,C)?

The written report is handed in electronically before the exam date (as discussed in class)

Use the mail subject title "genetic epi exam project".

During the oral part we further discuss this report.