

1. In a population of pea plants, the frequency of the red-flower color allele (R) is 0.7 and the frequency of the white flower-color allele (r) is 0.3. The red allele (R) is dominant to the white allele (r).

A pea plant with red flowers and known genotype Rr, is mated to a pea plant with red flowers and unknown genotype.

(a) In the population, what is the probability that a pea plant has red flowers?

(b) What is the probability that the second parent (a pea plant with red flowers and unknown genotype) has genotype RR?

2. For ABO blood types, the frequency of A allele is 0.3, of B allele is 0.1, and of O allele is 0.6.

Person of blood type AB, and person of blood type A but with unknown genotype, have a child.

(a) In the population, what is the probability that a person will be AA, AO?

(b) For the specified parents, what is the probability that their child is AA, AB, BB, AO, BO, OO?

3. Karl collected genetic array data from some samples and conducted a series of hypothesis tests of the type:

$$Y = a + b * X$$

where Y is a phenotype of interest, a is an intercept, b is a SNP-specific slope, and X is the genotype 0, 1, 2 for AA, AB, BB of a biallelic SNP.

- (a) Does the genotype coding 0, 1, 2 imply an additive, multiplicative, dominance, or recessive inheritance model?
- (b) Karl would like to control the rate at which he reports false positives to his supervisor? He wants to have a family-wise error rate in alignment with a 5% p-value. What is his significance threshold?
- (c) For the 10 SNPs their p-values are:
0.1,0.2,0.15,0.02,0.6,0.8,0.002,0.05,0.5

Which SNPs pass the significance threshold?

4. Joy wants to consider population stratification effects in her genetic analysis study. She performs a PCA with a fast software like `SNPRelate` or `EIGENSTRAT`. Her PCA plot looks like below. Provide an explanation for points in the middle.

