Bioinformatics and Statistical Genetics	Association analysis (12p)
Jan Graffelman	December 17, 2018

Name: .....

Name: .....

In this practical we perform association tests for a binary disease indicator and a genetic polymorphism. Prepare a .pdf file with all your answers and figures. Send your work by email to the course instructor (`jan.graffelman@upc.edu`) no later than the 24^{th} of December 2018.

1. The file rs394221.dat contains genotype information, for cases and controls, of polymorphism rs394221, which is presumably related to Alzheimer's disease.Load the data file into the R environment.

2. (1p) What is the sample size? What is the number of cases and the number of controls? Construct the contingency table of genotype by case/control status. .....

   .....

3. (1p) Explore the data by plotting the percentage of cases as a function of the genotype, ordering the latter according to the number of M alleles. Which allele increases the risk of the disease? .....

   .....

4. (2p) Test for equality of allele frequencies in cases and controls by doing an alleles test. Report the test statistic, its reference distribution, and the p-value of the test. Is there evidence for different allele frequencies? .....

   .....

5. (2p) Which are the assumptions made by the alleles test? Perform and report any addtional tests you consider adequate to verify the assumptions. Do you think the assumptions of the alleles test are met? ...

   .....

6. (2p) Perform the Armitage trend test for association between disease and number of M alleles. Report the test statistic, its reference distribution and the p-value of the test. Do you find evidence for association? .

   .....

7. (4p) Test for association between genotype and disease status by a logistic regression of disease status on genotype, treating the latter as categorical. Do you find significant evidence for association? Which allele increase the risk for the disease? Give the odds ratios of the genotypes with respect to base line genotype *mm*. Provide 95% confidence intervals for these odds ratios. .....

   .....

   .....

   .....

   .....

   .....