Bioinformatics and Statistical Genetics	Relatedness analysis: allele sharing
Iván Galván-Femenía & Jan Graffelman	December 23, 2016

Name: .....

You can make use of the R-package **genetics** (and other packages) to compute your answers. Prepare a .pdf file with all your answers and figures. Send your work by email to the course instructor (`ivan.galvan@upc.edu`) before Friday the 6^{th} of January 2017.

1. Download the file `MEX_chr22.rda` from the website of the course. This file contains genotype information of 76 individuals from the Mexican population. The genotype information concerns 20,531 SNPs on chromosome 22. Load this data into the R environment. The data file contains a matrix `X.Geno` containing the allele counts (0, 1 or 2) for one of the alleles of each SNP. The data file also provides a matrix `X.Fam` with pedigree information.

2. (1p) What percentage of the data is missing? .....

   .....

3. (1p) Plot the percentage of missing values per SNP. How many SNPs have a percentage larger than 1%? .....

   .....

4. (1p) Compute the histogram of the minor allele frequency per SNP. How many markers have minor allele frequency larger than 0.40? How many smaller than 0.05? .....

   .....

5. (2p) Consider SNPs with minor allele frequency larger than 0.40. Compute the mean (m) and the standard deviation (s) of the shared IBS alleles for all pairs of individuals. Plot s against m. Use the pedigree information to identify parent-offspring (PO) pairs. Do you think that there is a clear separation between PO and the rest of pairs of individuals? Comment on your findings. .....

   .....

6. (2p) Repeat the question 5 considering SNPs with minor allele frequency smaller than 0.05. Do you think that there is a clear separation between PO and the rest of pairs of individuals? Comment on your findings. .....

   .....

7. (2p) Consider again SNPs with minor allele frequency larger than 0.40. Compute and plot the fraction of loci sharing 0 against the fraction of loci sharing 2 IBS alleles for all pairs of individuals.

Use the pedigree information to identify PO pairs. Do you think that there is a clear separation between PO and the rest of pairs of individuals? Comment on your findings. .....

.....

8. (2p) Repeat the question 7 considering SNPs with minor allele frequency smaller than 0.05. Do you think that there is a clear separation between PO and the rest of pairs of individuals? Comment on your findings. .....

.....