Genetic association studies

EXERCISE 1: Researchers are interested in assessing possible association between candidate SNPs and the response to treatment in patients diagnosed with major depression (file "DM.txt NOTE: check how alleles are separated). The file also includes clinical information about other covariates of interest such as:

- HDRS: Hamilton Depression Rating Scale (continuous variable)

- PSICOT: Was the patient psychotic? (No and Yes)

- MELANCOL: Was the patient melancholic? (No and Yes)

- EPD_PREV: Number of previous episodes of depression (continuous variable)

- RESP: Response to treatment (outcome)

Answer the following questions:

1. Is it necessary to check HWE hypothesis for this example?

2. Is there any SNP associated with the response to the treatment? If so, write a sentence interpreting the results of that association (only for one SNP)

3. Does the result change after adjusting for other clinical covariates?

4. Create a plot with the p-values only for dominant, recessive and additive models

5. Compute the p-values using MAX-statistic for all SNPs

EXERCISE 2: Researchers are now interested in assessing the effect of haplotypes on the response to treatment. Answer the following questions:

1. Create a LD heatmap plot (NOTE: retrieve genomic positions using `biomaRt` - if you are having problems with that, just use consecutive positions (e.g., 1,2,3, ...))

2. Which is the combination of SNPs (e.g. number of consecutive haplotypes) that is more associated with the response to treatment? (NOTE1: Try only haplotypes of length 2 up to 4 - NOTE2: Remember that number of simulations have to be increased)

3. Which is the most frequent haplotype?

4. Compute the OR of association and its 95 % confidence intervals between the estimate haplotypes and the response to treatment