Howework 5
Genome 560
Due May 6, 2008

Question 1 [5 pts] Consider the following ten p-values :

1	0.0020
2	0.0045
3	0.0060
4	0.0080
5	0.0085
6	0.0090
7	0.0175
8	0.0250
9	0.1055
10	0.5350

Use the Benjamini and Hochberg FDR definition to decide which p-values would be rejected at a FDR = 0.05. If you'd like to demonstrate your R prowess, write a function that accepts an arbitrarily long vector of p-values and returns both the FDR and a 0 or 1 flag of whether to accept or reject the null hypothesis for the kth hypothesis.

Question 2 [5 pts] Let's say we have performed a microarray study to compare gene expression levels between a set of tumor and non-tumor cells. After testing for differential expression and correcting for multiple hypothesis tests, we look up our favorite gene and it has p-value of 5.5×10^5 and a q-value of 0.013. Interpret the p-value and q-value. What is the distinction between the two?

Question 3 [15 pts] Download the data file HW5.data from the course web site. The data consists of a quantitative phenotype measured in 225 individuals that were genotyped for two SNPs. The sex of each individual is also provided. Perform a multiple linear regression modeling the phenotype as a function of the three predictors (SNP1, SNP2, and Sex). Find the best model to account for the observed data, interpret the regression parameters, and perform the appropriate diagnostics to ensure the assumptions of the regression model are valid. If they are not, identify what assumptions have been violated and discuss what could be done to fix it. Finally, submit the R commands you used to perform the ananlyses.