

Homework 2

Sayan Mukherjee

- (1) Gene expression data. Download the matrix of gene expression data (simulated) from the course website. You will find a matrix with $n = 500$ individuals and $p = 10$ genes. Read in these data to a program like R. For this question, include the code that you used to answer these questions, and your answers, but not the plots themselves.
 - (a) Plot a histogram of each gene. Which ones look the farthest from normally distributed?
 - (b) Plot a quantile-quantile plot for each gene, comparing samples drawn from a normal distribution with the same mean and standard deviation as the gene in question with the actual samples. Which genes look the farthest from normally distributed? How do these plots compare with examining the histograms?
 - (c) Standardize one gene, and plot the same quantile-quantile plot with the standardized data. Does standardizing make the data look more or less similar to what is expected from a normal distribution? Are outliers controlled?
 - (d) Project one gene to the quantiles of a standard normal, and plot the same quantile-quantile plot with the quantile normalized data. How does quantile normalization impact the quantile-quantile plots? Are outliers controlled?
 - (e) Plot the covariance of the genes. Which genes appear to covary?
- (2) Finding *eQTLs*. Download a corresponding set of (simulated) genotypes from the course website.
 - (a) Which gene-SNP pairs appear to be associated ?
 - (b) Choose one associated gene-SNP pair, and one pair that does not appear associated. For both pairs, plot the SNP (x-axis) versus gene (y-axis). Overlay on top of this the best fit linear regression model parameters β . Include this figure in your solutions.
 - (c) What are the RSS for both of the pairs you have chosen? What is the r^2 value?
 - (d) Write out the likelihood of the gene expression data given the genotype data. Compute the likelihood of the gene expression data given the genotypes for your two chosen gene-SNP pairs (assume the residual variance $\sigma^2 = 1$).