



Northeastern University

College of Science

Module 7 Homework

Problem 1 (30 points)

For the Golub et al. (1999) data set, use appropriate Wilcoxon two-sample tests to find the genes whose mean expression values are higher in the ALL group than in the AML group.

- a) Use FDR adjustments at the 0.05 level. How many genes are expressed higher in the ALL group?
- b) Find the gene names for the top three genes with smallest p-values. Are they the same three genes with largest difference between the means in the ALL group and the AML group?

Please submit your R commands together with your answers to each part of the question.



Northeastern University

College of Science

Problem 2 (15 points)

For the Golub et al. (1999) data set, apply the Shapiro-Wilk test of normality to every gene's expression values in the AML group. How many genes do not pass the test at 0.05 level with FDR adjustment? Please submit your R script with the answer.



Northeastern University

College of Science

Problem 3 (15 points)

Gene "HOXA9 Homeo box A9" can cause leukemia (Golub et al., 1999). Use appropriate Wilcoxon two-sample tests to test if, for the ALL patients, the gene "HOXA9 Homeo box A9" expresses at the same level as the "CD33" gene. Please submit your R script with the answer.



Northeastern University

College of Science

Problem 4 (20 points)

The data set “UCBAdmissions” in R contains admission decisions by gender at six departments of UC Berkeley. For this data set, carry out appropriate test for independence between the admission decision and gender for each of the departments.

What are your conclusions? Please submit your R script with the answer.



Northeastern University

College of Science

Problem 5 (20 points)

There are two random samples $X_1 \dots X_n$ and $Y_1 \dots Y_m$ with population means μ_X and μ_Y and population variances σ_X^2 and σ_Y^2 . For testing $H_0: \sigma_X^2 = \sigma_Y^2$ versus

$H_A: \sigma_X^2 < \sigma_Y^2$, we can use a permutation test for the statistic $S = \frac{s_X^2}{s_Y^2}$.

Please program this permutation test in R. Use this nonparametric test on the “CD33” gene of the Golub et al. (1999) data set. Test whether the variance in the ALL group is smaller than the variance in the AML group. Please submit your R code with the answer.