

BIOLOGY 664: Integrated Bioinformatics Using R for Both Wet and Dry Scientists

Problem Set 3: Hypothesis tests on data.frames

Due: At the beginning of class Tuesday, March 4th.

Part 1 (This section is the same as Part 1 of Problem Set 1)

Merge the most relevant data found in the 3 tables (golub.gnames, golub, and golub.cl) that make-up the golub data in the library(multtest) into one data.frame with the following properties:

Name: golub.df

Dimensions: patient rows and named gene columns, and an additional named column for the cancer classifications

Column Names: use the gene name (column 2) from golub.gnames and "classification"

Classification Column: use a factor column in golub.df that uses "ALL" and "AML" as the classifications

Part 2

Reformulate the answers to the Chapter 3 Exercise 9 and a new Exercise 8 below to use your new golub.df data.frame.

New Exercise 8 (Re-worded so as to use your golub.df dataframe)

- a) Perform a hypothesis test to see if the distribution of the expression values for the Zyxin gene for the ALL patients are normally distributed.
- b) Perform a hypothesis test to see if the distribution of the expression values for the Zyxin gene for the ALL patients are distributed according to $N(1.6, 0.4^2)$.