Complete the following computations on gene means of the Golub data set.

a) Compute the mean expression values for every gene among “ALL” patients.

**a) Rscript:**

source("http://www.bioconductor.org/biocLite.R")

biocLite()

biocLite("multtest")

library(multtest)

data(golub)

gol.fac <- factor (golub.cl, levels=0:1, labels= c ("ALL","AML"))

meanALL<- apply(golub[,gol.fac=="ALL"],1,mean)

meanALL

**b) Compute the mean expression values for every gene among “AML” patients.**

**b). Rscript:**

library(multtest)

data(golub)

gol.fac <- factor (golub.cl, levels=0:1, labels= c ("ALL","AML"))

meanALL<- apply(golub[,gol.fac=="ALL"],1,mean)

meanALL

**c) Give the biological names of the three genes with the largest mean**

**expression value among “ALL” patients.**

**c)Rscript:**

library(multtest); data(golub)

gol.fac <- factor(golub.cl,levels=0:1, labels= c("ALL","AML"))

mall <- apply(golub[,gol.fac=="ALL"], 1, mean)

maml <- apply(golub[,gol.fac=="AML"], 1, mean)

o <- order(abs(mall), decreasing=TRUE)

print(golub.gnames[o[1:3],2])

**Answer**:

[1] "GB DEF = Chromosome 1q subtelomeric sequence D1S553"

[2] "37 kD laminin receptor precursor/p40 ribosome associated protein gene"

[3] "RPS14 gene (ribosomal protein S14) extracted from Human ribosomal protein S14 gene"

**d) Give the biological names of the three genes with the largest mean**

**expression value among “AML” patients.**

d). Rscript:

library(multtest); data(golub)

gol.fac <- factor(golub.cl,levels=0:1, labels= c("ALL","AML"))

mall <- apply(golub[,gol.fac=="ALL"], 1, mean)

maml <- apply(golub[,gol.fac=="AML"], 1, mean)

o <- order(abs(maml), decreasing=TRUE)

print(golub.gnames[o[1:3],2])

**Answer**:

[1] "GB DEF = mRNA fragment for elongation factor TU (N-terminus)"

[2] "GB DEF = HLA-B null allele mRNA"

[3] "Globin, Beta"

**Problem 2 (30 points)**

**Complete the following computations using the Golub data set.**

1. **Save the expression values of the first five genes (in the first five rows) for the AML patients in a csv file “AML5.csv.”**
2. **Rscript:**

library(multtest); data(golub)

expval<-rbind(golub[1:5,28:38])

write.table (expval, file= "AML5.csv")

**Answer**:

Attached file AML5.csv

1. **Save the expression values of the first five genes for the ALL patients in a plain text file “ALL5.txt.”**

**Rscript:**

library(multtest); data(golub)

eall<-rbind(golub[1:5,1:27])

write.table(eall, file= "ALL5.txt")

**Answer**:

Attached file ALL5.txt

**c) Compute the standard deviation of the expression values on the first patient, of the 100th to 200th genes (total 101 genes).**

**Rscript:**

library(multtest); data(golub)

stddev<-sd(golub[100:200,1])

stddev

**Answer**:

[1] 0.9174976

1. **Compute the standard deviation of the expression values of every gene, across all patients. Find the number of genes with standard deviations greater than 1.**

**Rscript:**

library(multtest); data(golub)

stdd<-sd(golub[,1:38])

stdd

i<-1

ctr<-0

for (i in golub[i,1:38])

{

i<-i+1

std<-sd(golub[i,1:38])

if(!is.na(std > 1))

{

ctr<-ctr+1

}

}

ctr

**Answer:**

Standard deviation of expression value of every gene - stdd

[1] 0.9998404

number of genes with standard deviations greater than 1 – ctr

[1] 3

**e) Do a scatter plot of the 101st gene expressions against the 102nd gene**

**expressions, labeling the x-axis and the y-axis with the genes’ biological**

**names. Do this using xlab= and ylab= control options.**

**Rscript:**

plot(golub[101,1:38],golub[102,1:38],xlab=golub.gnames[101,2],ylab=golub.gnames[102,2])

**Answer:**

Answer2e.png

**Problem 3 (20 points)**

**Complete a–c using the ALL data set.**

**Load the ALL data from the ALL library, and use str and openVignette() for**

**further orientation.**

**a) Use exprs(ALL[,ALL$BT=="B1"] to extract the gene expressions from the patients in disease stage B1. Produce a histogram of these gene expressions.**

**Rscript:**

library(ALL)

data(ALL)

vignette()

hist(exprs(ALL[,ALL$BT=="B1"]),xlab="Gene Expression",ylab="Frequency",main="Gene expression of patients in disease stage B1")

**Answer:**

Answer3a.png

**b) Compute the mean gene expressions over these B1 patients.**

**Rscript**:

library(ALL)

data(ALL)

vignette()

meangene<-apply(exprs(ALL[,ALL$BT=="B1"]),1,mean)

meangene

**Answer:**

In Rscript.

**c) Give the gene identifiers of the three genes with the largest mean.**

**Rscript:**

library(ALL)

data(ALL)

vignette()

B1<-exprs(ALL[,ALL$BT=="B1"])

meangene<-apply(B1,1,mean)

meangene

o <- order(abs(meangene), decreasing=TRUE)

print(B1[o[1:3],1])

**Answer:**

AFFX-hum\_alu\_at 31962\_at 31957\_r\_at

13.72101 13.36514 13.05370

**Problem 4 (20 points)**

**To complete a and b, work with the “trees” data set that comes with R.**

**a) Find the type of the trees data object.**

**Rscipt:**

class(trees)

**Answer:**

"data.frame"

**b) Produce a figure with two overlaid scatterplots: girth versus height and girth versus volume. Do the height plot with blue “+” symbols, and do the volume plot with red “o” symbols. You need to set the ylim= control option so that all points from the two plots can show up on the merged figure.**

**Rscript**:

plot(trees[,1],trees[,2],col="blue",pch="+", ylim=c(1,90),xlim=c(1,32),xlab="Girth",ylab="Height and Volume",main="Overlaid scatter plot")

points(trees[,3],col="red")

legend("bottomright",c("Height","Volume"),col=c("blue","red"),lty=c(1,1))

**Answer:**

Answer4b.png