

Assignment2

2023-01-23

Problem 1 Compute the gene means of golub data set

Installed package multtest from BiocManager

Problem 1 a) mean expression values for every gene among ALL patients

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

b) mean expression values for every gene among AML patients

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

c) biological names of three genes with largest mean expression among AML patients

```
meanALL <- apply(golub[,gol.fac=="ALL"], 1, mean)
o <- order(abs(meanALL), decreasing=TRUE)
print(golub.gnames[o[1:3],2])
```

```
## [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) biological names of three genes with largest mean expression value among AML patients

```
meanAML <- apply(golub[,gol.fac == "AML"], 1,mean)
o <- order(abs(meanAML),decreasing = TRUE)
print(golub.gnames[o[1:3],2])
```

```
## [1] "GB DEF = mRNA fragment for elongation factor TU (N-terminus)"
## [2] "GB DEF = HLA-B null allele mRNA"
## [3] "Globin, Beta"
```

Problem 2

a) To save the expression values of first five genes for AML patients in a csv

```
firstFive <- golub[1:5,gol.fac == "AML"]
write.csv(firstFive,"AML5.csv")
```

b) To save the expression values of first five genes for ALL patients in a text

```
first_five_ALL <- golub[1:5,gol.fac == "ALL"]  
write.table(first_five_ALL, "ALL5.txt")
```

c) Compute the standard deviation of expression values on 1st patient of 100th to 200th genes

```
patient1 <- (golub[100:200,1])  
standardDeviation <- sd(patient1)  
print(standardDeviation)
```

```
## [1] 0.9174976
```

d) To compute the standard deviation of every gene across all the patients

```
standard_deviation <- apply(golub,1,sd)  
print(length(which(standard_deviation>1)))
```

```
## [1] 123
```

e) A scatter plot of the 101th gene expressions against the 102th gene expressions

```
gene_exp = golub[101,]  
  
gene_exp2 = golub[102,]  
  
name1 = golub.gnames[101,2]  
name1
```

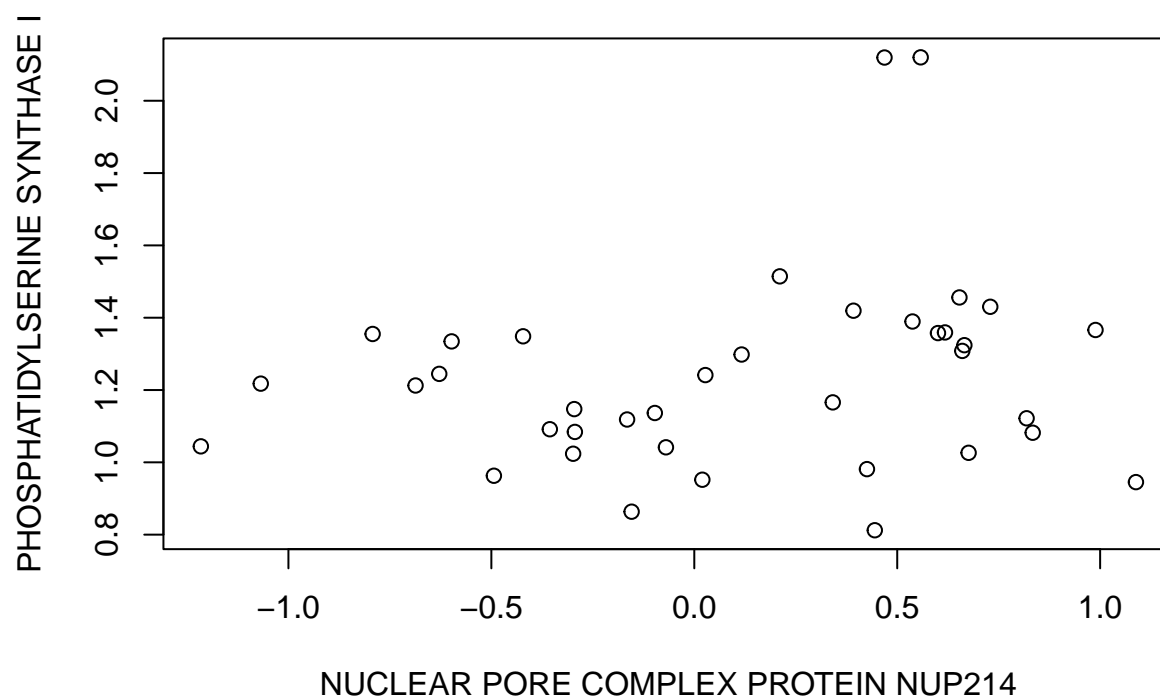
```
## [1] "NUCLEAR PORE COMPLEX PROTEIN NUP214"
```

```
name2 = golub.gnames[102,2]  
name2
```

```
## [1] "PHOSPHATIDYLSERINE SYNTHASE I"
```

```
plot(gene_exp,gene_exp2,xlab =name1,ylab = name2, main = "scatterplot of gene expression 101th against 102th")
```

scatterplot of gene expression 101th against 102th



Problem 3

Work with the ALL data set, Load the ALL data from the ALL library and use str and openVignette() for a further orientation.

```
library("ALL")
str("ALL")
```

```
## chr "ALL"
```

```
# openVignette("ALL")
```

- Produce one histogram of gene expression using `exprs(ALL[,ALL$BT=="B1"])` to extract the gene expression from patients in disease stage B1

```
gene_expression_B1 <- exprs(ALL[, ALL$BT == "B1"])
hist(gene_expression_B1, xlab = "Gene Expression", ylab = "Frequency", main = "Histogram plot of Gene Expression for B1 patients")
```

- mean gene expressions for every gene over these B1 patients.

```
meanB1 <- apply(exprs(ALL[,ALL$BT=="B1"]),1, mean)
head(meanB1)
```

- gene identifiers of the three genes with the largest mean

```
meanB1 <- apply(exprs(ALL[,ALL$BT=="B1"]),1, mean)
sorted <- sort(meanB1, decreasing=TRUE)
print(sorted[1:3])
```

Problem 4

To produce 2 overlaid scatterplots Height vs Girth, Volume vs Girth.

```
plot(trees$Girth, trees$Height ,pch = "+", xlab = "Girth", main = "Overlaid scatter plot of Girth vs h
points(trees$Girth, trees$Height, pch = "+", col = "blue")
points(trees$Girth, trees$Volume, pch = "o", col = "red")
legend("topright", c("Height", "Volume"), col = c("blue", "red"), pch = c("+", "o"))
```

