**Module 6 Homework**

**1. (60 points)** On the Golub et al. (1999) data, consider the “H4/j gene” gene (row 2972) and the “APS Prostate specific antigen” gene (row 2989). Setup the appropriate hypothesis for proving the following claims. Chose and carry out the appropriate tests.

**(a)** The mean “H4/j gene” gene expression value in the ALL group is greater than

-0.9 (note that this is negative 0.9).

**Answer)**

**library("multtest")**

**data(golub)**

**gol.fac<-golub[2972,gol.fac=="ALL"]**

**t.test(gol.fac, mu=-0.9, alternative = "greater")**

**Output)**

**(b)** The mean “H4/j gene” gene expression value in ALL group differs from the mean “H4/j gene” gene expression value in the AML group.

**(c)** In the ALL group, the mean expression value for the “H4/j gene” gene is lower than the mean expression value for the “APS Prostate specific antigen” gene.

**(d)** Let pH4j denotes the proportion of patients for whom the “H4/j gene” expression values is greater than -0.6. We wish to show that pH4j in the ALL group is less than 0.5.

**(e)** The proportion pH4j in the ALL group differs from the proportion pH4j in the AML group.

**You should state the hypothesis, show the R commands for the tests, show the output of these tests, and state your conclusion based on these outputs.**

**2.** **(10 points)** Suppose that the probability to reject a biological hypothesis by the results of a certain experiment is 0.03. This experiment is repeated 3000 times.

**(a)** How many rejections do you expect?

**Answer)**

**n<-3000**

**p<-0.03**

**expreject <- (n\*p)**

**expreject**

**Output)**

**[1] 90**

**(b)** What is the probability of less than 75 rejections?

**Answer)**

**pbinom(74,3000,0.03)**

**Output)**

[1] 0.04537989

**3. (10 points)**

For testing H0: μ=5 versus HA: μ>5, we considers a new α=0.1 level test which rejects when falls between and .

Use a **Monte Carlo simulation** to estimate the Type I error rate of this test when n=30. Do 10,000 simulation runs of data sets from the . Please show the R script for the simulation, and the R outputs for running the script. Provide your numerical estimate for the Type I error rate. Is this test valid (that is, is its Type I error rate same as the nominal α=0.1 level)?

**Answer)**

**x.sim<-matrix(rnorm(10000\*30, mean=5, sd=4), ncol=30)**

**tstat<-function(x)(mean(x)-5)/sd(x)\*sqrt(length(x))**

**tstat.sim<-apply(x.sim,1,tstat)**

**power.sim<-mean(tstat.sim>qt(0.90,df=29))**

**power.sim+c(-1,0,1)\*qnorm(0.975)\*sqrt(power.sim\*(1-power.sim)/10000)**

**Output)**

**[1] 0.09509407 0.10100000 0.10690593**

* **So the Monte Carlo estimate of the Type I error rate is 0.101 with its 95% CI as (0.095, 0.1069). This does agree with the nominal level of α = 0.1.**

**4. (20 points)**

On the Golub et al. (1999) data set, do **Welch two-sample t-tests** to compare every gene’s expression values in ALL group versus in AML group.

1. Use Bonferroni and FDR adjustments both at 0.05 level. How many genes are differentially expressed according to these two criteria?

**Answer)**

data(golub, package = "multtest")

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

ALL <- golub[2972, gol.fac == "ALL"]

AML <- golub[2972, gol.fac == "AML"]

t\_test <- apply(golub, 1, function(x) t.test(x~gol.fac, var.equal = F))

p.values <- sapply(t\_test, function(x) x$p.value)

p.bon <-p.adjust(p=p.values, method="bonferroni")

p.fdr <-p.adjust(p=p.values, method="fdr")

sum(p.bon<0.05)

sum(p.fdr<0.05)

**Output)**

**> sum(p.bon<0.05)**

**[1] 103**

**> sum(p.fdr<0.05)**

**[1] 695**

1. Find the gene names for the top three strongest differentially expressed genes (i.e., minimum p-values). Hint: the gene names are stored in ***golub.gnames***.

**Answer)**

**data(golub, package = "multtest");**

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

**test <- apply(golub, 1, function(x) t.test(x ~ gol.fac)$p.value)**

**o <- order(test,decreasing=FALSE)**

**golub.gnames[o[1:3],2]**

**Output)**

[1] "Zyxin"

[2] "FAH Fumarylacetoacetate"

[3] "APLP2 Amyloid beta (A4) precursor-like protein 2"

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