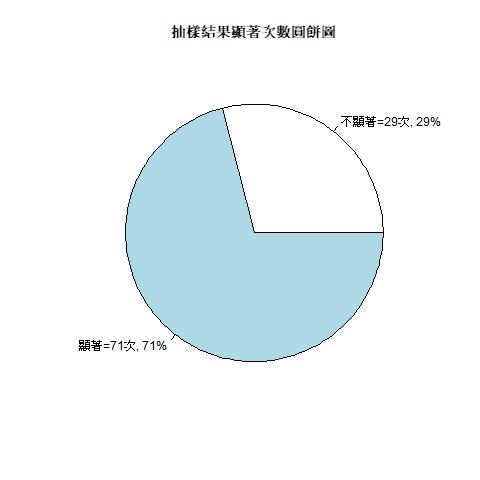
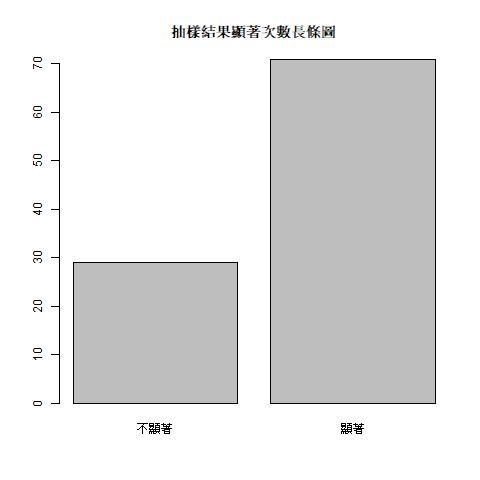
作業2

公衛三 梁嫚芳 B07801003

1. 請讀入partA\_III.csv，並將樣本依Gender分 類，針對G1-G5進行t-test，選擇其中具有 顯著差異(P < 0.01)的基因進行後續分析： 隨機挑選Gender為0及Gender為1的樣本各 250人，並再次進行t-test檢定後儲存其pvalue，重複上次讀隨機取樣及檢定過程 100次，測試當樣本數減半時仍能發現顯 著差異的次數有多少次，將檢定結果依P < 0.05分為顯著或不顯著後利用圓餅圖及 長條圖呈現(分別儲存為output1\_pie.JPG及 output1\_barplot.PNG)

顯著差異(p<0.01)的基因進行後續分析：(G4)



Code:

## 1

# 請讀入partA\_III.csv

setwd("D:/OneDrive - 國立台灣大學/109-2/計算生物學原理與應用/hw")

A3 <- read.csv(file.choose())

# 並將樣本依Gender分類，針對G1-G5進行t-test

ttest\_pvalue <- c()

for(i in 6:10){

ttest\_pvalue[i-5] <- t.test(A3[which(A3$Gender==0), i], A3[which(A3$Gender==1), i])$p.val

}

# 選擇其中具有顯著差異(P < 0.01)的基因進行後續分析

gene\_index <- which(ttest\_pvalue < 0.01)

# gender index

G0\_index <- which(A3$Gender==0)

G1\_index <- which(A3$Gender==1)

# 重複讀隨機取樣及檢定過程100次

ttest\_pvalue\_100 <- c()

for(k in 1:100){

# 隨機挑選Gender為0及Gender為1的樣本各250人

# 用GENDER當index去抽

G0\_sample <- A3[sample(G0\_index, 250), ]

G1\_sample <- A3[sample(G1\_index, 250), ]

# 並再次進行t-test檢定後儲存其pvalue

ttest\_pvalue <- c()

# 記得要只選前面有顯著的基因gene\_index

for(i in (5 + gene\_index)){

ttest\_pvalue[i-5] <- t.test(G0\_sample[, i], G1\_sample[, i])$p.val

}

ttest\_pvalue\_100 <- rbind(ttest\_pvalue\_100, ttest\_pvalue)

}

# 測試當樣本數減半時仍能發現顯著差異的次數有多少次

# 將檢定結果依P< 0.05分為顯著或不顯著後

sig <- table(ttest\_pvalue\_100 < 0.05)

# 利用圓餅圖及長條圖呈現(分別儲存為output1\_pie.JPG及output1\_barplot.PNG)

jpeg("output1\_pie.JPG")

pie(sig, main="抽樣結果顯著次數圓餅圖", labels=c(paste0("不顯著=", sig[1], "次, ", sig[1] / sum(sig) \* 100, "%"), paste0("顯著=", sig[2], "次, ", sig[2] / sum(sig) \* 100, "%")))

dev.off()

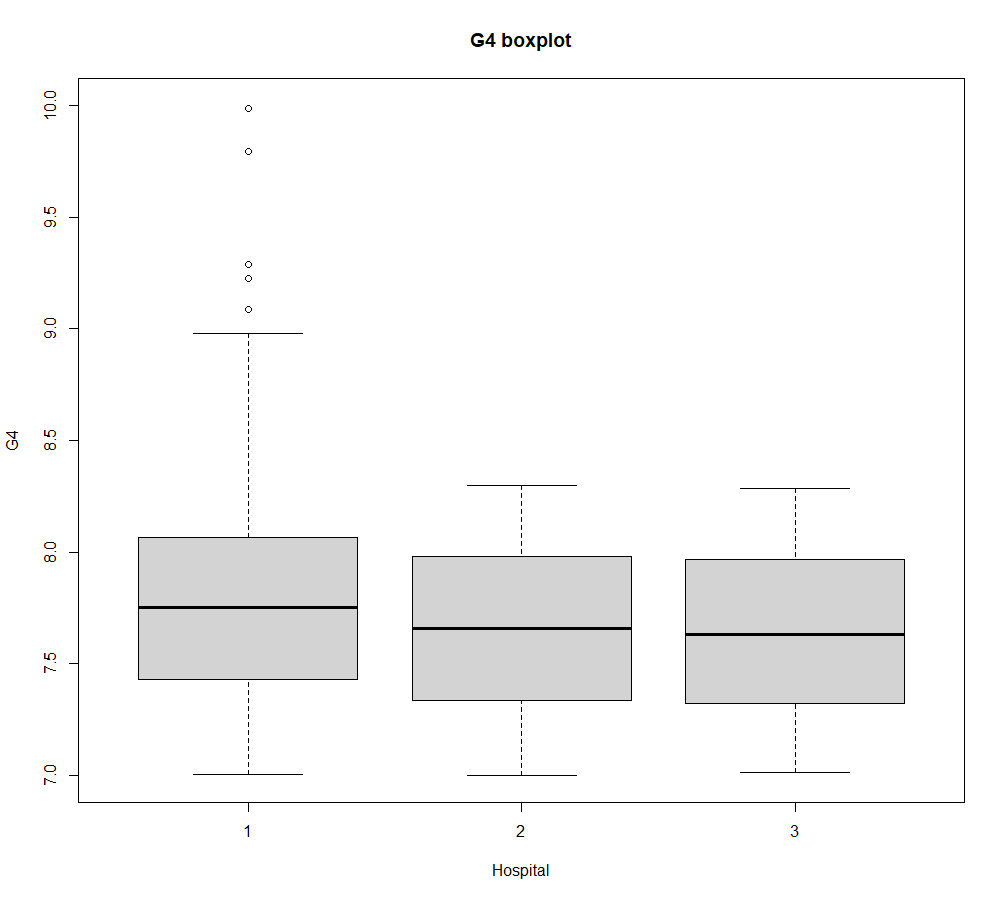
png("output1\_barplot.PNG")

barplot(sig, main="抽樣結果顯著次數長條圖", names.arg=c("不顯著","顯著"))

dev.off()

1. 請讀入partA\_III.csv，並將樣本依 Hospital分類，針對G1-G5進行anova檢 定，選擇其中具有顯著差異(P < 0.01)的 基因進行後續分析：將基因數值利用盒 鬚圖呈現，請做出一張圖包含三張子圖 形，x軸請顯示Hospital 1 or 2 or 3，y軸 請固定最小值為7，最大值為10，將圖 形儲存為output2.PDF

顯著差異(P < 0.01)的基因： (G4)



Code:

## 2

# 請讀入partA\_III.csv

# 上題已讀入

# 將樣本依Hospital分類，針對G1-G5進行anova檢定

anova\_pvalue <- c()

for(i in 6:10){

anova\_pvalue[i-5] <- summary(aov(A3[, i]~factor(A3$Hospital), data=A3))[[1]][["Pr(>F)"]][1]

}

# 選擇其中具有顯著差異(P < 0.01)的基因進行後續分析

gene\_index <- which(anova\_pvalue < 0.01)

# 將基因數值利用盒鬚圖呈現

# 請做出一張圖包含三張子圖形

# x軸請顯示Hospital 1 or 2 or 3

# y軸請固定最小值為7，最大值為10

# 將圖形儲存為output2.PDF

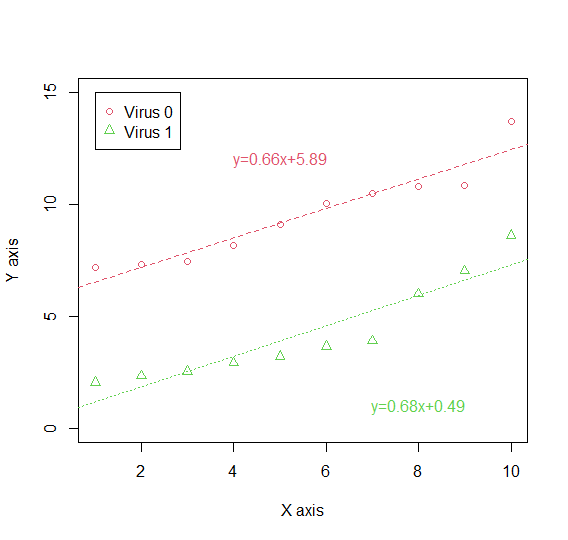
pdf("output2.PDF")

boxplot(G4 ~ Hospital, main="G4 boxplot", xlab="Hospital", ylim=c(7, 10), data=A3)

dev.off()

1. 請讀入partB\_III.csv，並將樣本依Virus分類， 針對G1-G5進行wilcoxon rank sum檢定，選擇 其中具有顯著差異(P < 0.01)的基因進行後續 分析：請在同一張圖形上將該基因對Virus變 數分類樣本後以不同形狀及不同顏色的資料 點繪出，請固定y軸最小值為0，最大值為15， 請針對這兩群數值加上迴歸線，並放上文字 說明其迴歸線公式及圖說說明兩群樣本為 Virus 0或Virus 1，將圖形儲存為output3.bmp (x軸請用Fake\_index->1:10)

顯著差異(P < 0.01)的基因： (G3)



Code:

## 3

# 請讀入partB\_III.csv

B3 <- read.csv(file.choose())

# 並將樣本依Virus分類， 針對G1-G5進行wilcoxon rank sum檢定

wil\_pvalue <- c()

for(i in 6:10){

wil\_pvalue[i-5] <- wilcox.test(B3[, i]~factor(B3$Virus))$p.value

}

gene\_index <- which(wil\_pvalue < 0.01)

# 請在同一張圖形上將該基因對Virus變數分類樣本後 以不同形狀及不同顏色的資料點繪出

# 請固定y軸最小值為0，最大值為15

# 將圖形儲存為output3.bmp (x軸請用Fake\_index->1:10)

bmp("output3.bmp")

ff <- 1:10

plot(ff, ff, type="n", ylim=c(0, 15), xlab="X axis", ylab="Y axis")

V\_0 <- sort(B3$G3[B3$Virus==0])

V\_1 <- sort(B3$G3[B3$Virus==1])

points(V\_0, pch=1, col=2)

points(V\_1, pch=2, col=3)

# 請針對這兩群數值加上迴歸線

abline(lm(V\_0~ff), lty=2, col=2)

abline(lm(V\_1~ff), lty=3, col=3)

# 並放上文字 說明其迴歸線公式

text(5, 12, col=2, paste0("y=",round(lm(V\_0~ff)$coefficients[2], 2), "x+", round(lm(V\_0~ff)$coefficients[1], 2)))

text(8, 1, col=3, paste0("y=",round(lm(V\_1~ff)$coefficients[2], 2), "x+", round(lm(V\_1~ff)$coefficients[1], 2)))

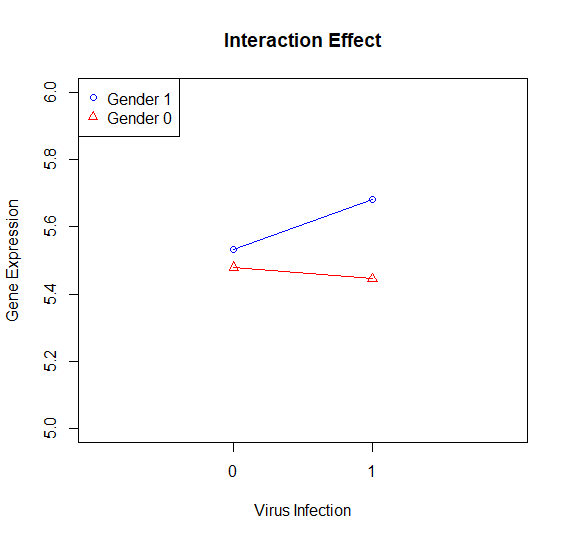
# 及圖說說明兩群樣本為 Virus 0或Virus 1

legend(1, 15, legend=c("Virus 0", "Virus 1"), pch=c(1, 2), col=c(2, 3))

dev.off()

1. 請讀入partC\_III.csv，並將樣本依Gender與Virus分類， 針對G1-G5進行線性迴歸，並加入Gender與Virus的 交互作用項，選擇其中具有顯著差異(P < 0.01)的基因畫出下圖：將圖形儲存為output4.pdf

顯著差異(P < 0.01)的基因： (G1)



Code:

## 4

# 請讀入partC\_III.csv

C3 <- read.csv(file.choose())

reg\_pvalue <- c()

# 並將樣本依Gender與Virus分類，針對G1-G5進行線性迴歸，並加入Gender與Virus的 交互作用項

for (i in 6:10){

reg\_pvalue[i-5] <- summary(lm(C3[, i] ~ factor(C3$Gender) + factor(C3$Virus) + factor(C3$Gender \* C3$Virus)))$coefficients[4, 4]

}

# 選擇其中具有顯著差異(P < 0.01)的基 因畫出下圖

gene\_index <- which(reg\_pvalue < 0.01)

# 交互作用效果

G0\_V0 <- mean(C3[which(C3$Gender==0 & C3$Virus==0), 6])

G0\_V1 <- mean(C3[which(C3$Gender==0 & C3$Virus==1), 6])

G1\_V0 <- mean(C3[which(C3$Gender==1 & C3$Virus==0), 6])

G1\_V1 <- mean(C3[which(C3$Gender==1 & C3$Virus==1), 6])

xx <- c(0, 1, 0, 1) # virus

yy <- c(G0\_V0, G0\_V1, G1\_V0, G1\_V1)

# 將圖形儲存為output4.pdf

pdf("output4.pdf")

plot(xx, yy, xlim=c(-1, 2), ylim=c(5, 6), xaxt="n", main="Interaction Effect", type="n", xlab="Virus Infection", ylab="Gene Expression")

axis(1, at=c(0,1), labels=c("0", "1"))

points(xx[1], yy[1], pch=2, col="Red")

points(xx[2], yy[2], pch=2, col="Red")

points(xx[3], yy[3], pch=1, col="Blue")

points(xx[4], yy[4], pch=1, col="Blue")

lines(xx[1:2], yy[1:2], col="Red")

lines(xx[3:4], yy[3:4], col="Blue")

legend("topleft", legend=c("Gender 1", "Gender 0"), pch=c(1,2), col=c("Blue", "Red"))

dev.off()