Applied Multivariate analysis-HW7

ID : 111024517　　　　Name：鄭家豪

Pre-processing

Check missing value

From the data description, we know that the variable gender = 9 means Missing value, so calculating how many Missing values:

```
The sum of gender is 9(meaning missing): 37
```

```
The total observation of womenraw dataset: 33590
```

Here we know that there are 37 Missing values, which is not a large number compared to the total number of samples, so we choose to remove these 37 samples here.
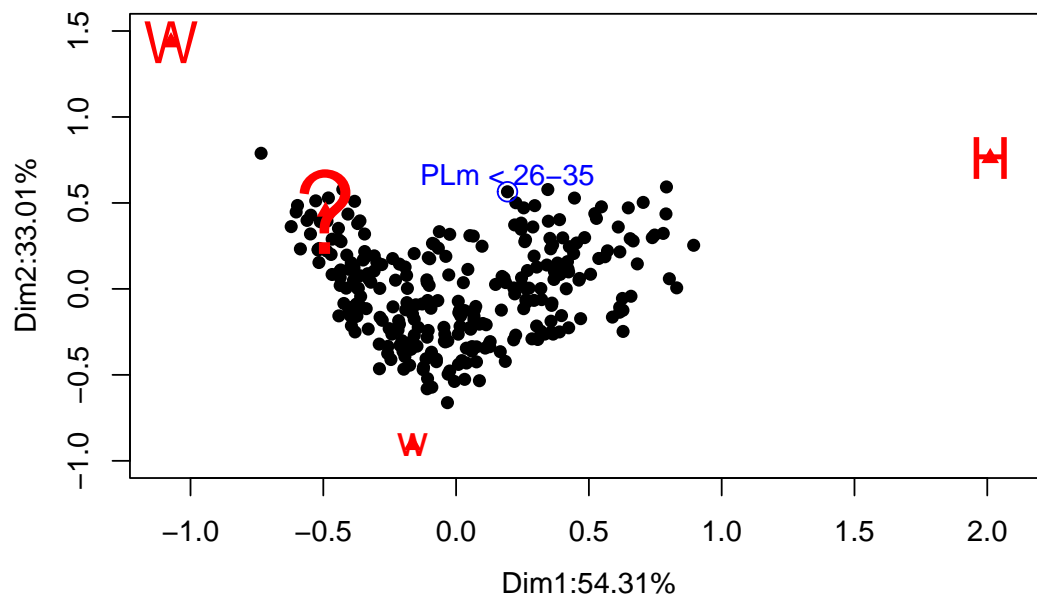
Construct contingency table

Using the data that has removed the Missing value to integrate into contingency table:(code in Appendix)

C	G	A	Q3			
			W	w	H	?
1	1	1	14	22	2	8
2	1	1	9	101	28	20
...	...	...	...	...	...	...
...	...	...	...	...	...	...
16	1	2	45	31	45	20
...	...	...	...	...	...	...
...	...	...	...	...	...	...
23	2	6	19	48	25	8
24	2	6	28	94	71	30

(Note:The PLm<26-35 for coding is (C,G,A)= (16,1,2)

Corresponding Analysis

CA plot: The coding is attached in Appendix

Appendix

```r
# Read data
library("readxl")
data <- read_excel("womenraw.xls")
cat("The sum of gender is 9(meaning missing):",sum(data$G==9),"\n")
cat("The total observation of womenraw dataset:",dim(data)[1])

# Contingency table
library(kableExtra)
library(tidyverse)
rm.data <- data[-which(data$G == 9),]
mat <- xtabs(~rm.data$C+rm.data$G+rm.data$A+rm.data$Q3,data=rm.data)
mat <- data.frame(mat)
colnames(mat) <- c("C","G","A","Q3","Freq")
contingency.mat <- matrix(0,nrow = 24*2*6,ncol = 4)
for (i in 1:4){
  index <- sort(unique(mat$Q3))[i]
  contingency.mat[,i] <- mat$Freq[which(mat$Q3==i)]
```

```r
}
contingency.mat <- data.frame(contingency.mat)
colnames(contingency.mat) =c("W","w","H","?")
con.mat <- cbind(mat[1:288,1:3],contingency.mat)
omit = matrix(rep("...", ncol(con.mat)*2), nrow = 2)
colnames(omit) <- colnames(con.mat)
a <- which(mat$C == 16 & mat$G == 1 & mat$A == 2 & mat$Q3==1)
print <- rbind(con.mat[1:2,],omit,con.mat[a,],omit,con.mat[287:288,])
kable(print,row.names = FALSE) %>%
add_header_above(c(" ", " ", " ", "Q3" = 4))


# CA plot
library(ca)
ca = ca(contingency.mat)
principal.coord <- cacoord(ca,type="rowprincipal")
row.coord <-  principal.coord$rows
plot(x = row.coord[,1],y=-row.coord[,2],pch=16,col="black",
     xlab="",ylab="",xlim=c(-1.1,2.1),ylim=c(-1,1.5))
mtext(paste0("Dim1:",100*round((ca$sv^2)[1]/sum(ca$rowinertia),4),"%"),
      side=1, line=2.5)
mtext(paste0("Dim2:",100*round((ca$sv^2)[2]/sum(ca$rowinertia),4),"%"),
      side=2, line=2.5)
points(x = ca$colcoord[,1],y = -ca$colcoord[,2],col="red",pch=17)
text(x=ca$colcoord[,1],y=-ca$colcoord[,2],cex = c(2,1.5,2,4),
     labels = c("W","w","H","?"),col = c("red","red","red","red"))
points(x = row.coord[a,1],y = -row.coord[a,2],
       col="blue",pch = 1,cex=1.5)
text(x= row.coord[a,1],y =-row.coord[a,2] + 0.1 ,cex=1,
     labels = "PLm < 26-35",col="blue")
```