Assignment 4

STAT 32950

Ki Hyun

Due: 09:00 (CT) 2023-04-18

Problem 1.

(a)

- For x_1:

$$\begin{bmatrix} 6 & 5 & 8 & 4 & 7 \\ 3 & 1 & 2 & & \\ 2 & 5 & 3 & 2 & \end{bmatrix} = \begin{bmatrix} 4 & 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & & \\ 4 & 4 & 4 & 4 & \end{bmatrix} + \begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ -2 & -2 & -2 & & \\ -1 & -1 & -1 & -1 & \end{bmatrix} + \begin{bmatrix} 0 & -1 & 2 & -2 & 1 \\ 1 & -1 & 0 & & \\ -1 & 2 & 0 & -1 & \end{bmatrix}$$

- For x_2:

$$\begin{bmatrix} 7 & 9 & 6 & 9 & 9 \\ 3 & 6 & 3 & & \\ 3 & 1 & 1 & 3 & \end{bmatrix} = \begin{bmatrix} 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & & \\ 5 & 5 & 5 & 5 & \end{bmatrix} + \begin{bmatrix} 3 & 3 & 3 & 3 & 3 \\ -1 & -1 & -1 & & \\ -3 & -3 & -3 & -3 & \end{bmatrix} + \begin{bmatrix} -1 & 1 & -2 & 1 & 1 \\ -1 & 2 & -1 & & \\ 1 & -1 & -1 & 1 & \end{bmatrix}$$

(b)

Noting that:

$$\bar{\mathbf{x}} = \begin{pmatrix} 4 \\ 5 \end{pmatrix},\ \bar{\mathbf{x}}_{t=1} = \begin{pmatrix} 2 \\ 3 \end{pmatrix},\ \bar{\mathbf{x}}_{t=2} = \begin{pmatrix} -2 \\ -1 \end{pmatrix},\ \bar{\mathbf{x}}_{t=3} = \begin{pmatrix} -1 \\ -3 \end{pmatrix}$$

We can compute the `MANOVA` table from the data as below

	Matrix of sum of squares and cross-products	Degrees of freedom
Treatments	$\mathbf{B} = \begin{bmatrix} 36 & 48 \\ 48 & 84 \end{bmatrix}$	$g - 1 = 2$
Residuals	$\mathbf{W} = \begin{bmatrix} 18 & -13 \\ -13 & 18 \end{bmatrix}$	$\sum_{l=1}^{3} n_l - g = 9$
Total	$\mathbf{B} + \mathbf{W} = \begin{bmatrix} 54 & 35 \\ 35 & 102 \end{bmatrix}$	$\sum_{l=1}^{3} n_l - 1 = 11$

(c)

$$|\mathbf{B}| = \det\left(\begin{bmatrix} 36 & 48 \\ 48 & 84 \end{bmatrix}\right) = 36 \cdot 84 - 48^2 = 720$$

$$|\mathbf{W}| = \det\left(\begin{bmatrix} 18 & -13 \\ -13 & 18 \end{bmatrix}\right) = 18^2 - 13^2 = 155$$

$$|\mathbf{B} + \mathbf{W}| = \det\left(\begin{bmatrix} 54 & 35 \\ 35 & 102 \end{bmatrix}\right) = 54 \cdot 102 - 35^2 = 4283$$

$$\therefore \Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{155}{4283}$$

(d)

```r
lambda_star = 155/4283
n = 12
p = 2
g = 3

chisq_q = -log(lambda_star)*(n - 1 - (p + g)/2)
p_val = 1 - pchisq(chisq_q, df = p*(g - 1))

print(paste0("The p-value from the data is: ", p_val))
```

```
## [1] "The p-value from the data is: 1.13011001031671e-05"
```

(e)

$$\begin{cases} H_0: & \text{Treatment 1, 2, and 3 have effect of 0} \\ H_A: & \text{At least 1 treatment had non-zero effect} \end{cases}$$

Problem 2.

```
skull = read.table("T6-13.DAT")
colnames(skull) = c("x1", "X2", "x3", "x4", "period")
```

(a)

```
X <- cbind(skull$x1, skull$X2, skull$x3, skull$x4)
summary(manova(X ~ skull$period))
```

```
##              Df  Pillai approx F num Df den Df  Pr(>F)
## skull$period  1 0.13924   3.4374      4     85 0.01182 *
## Residuals    88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The `MANOVA` test above concludes that the null-hypothesis may be objected under significance level $\alpha = 0.05$. Therefore, difference in period may be associated with, on overage, a difference in skull based on the four variables.

(b)

```
g = length(unique(skull$period))
for(i in 1:(g - 1)){
  for(j in (i+1):g){
    group_x <- skull %>% filter(period == i) %>% .[-5]
    group_y <- skull %>% filter(period == j) %>% .[-5]
    diffmean = colMeans(group_x) - colMeans(group_y)
    n1 = nrow(group_x); n2 = nrow(group_y); p = ncol(skull) - 1
    Spool = (n1 - 1)/(n1 + n2 - 2) * cov(group_x) +
      (n2 - 1)/(n1 + n2 - 2) * cov(group_y)
    # Hotelling's T2
    T_sq = t(diffmean) %*% solve((1/n1 + 1/n2)*Spool) %*% diffmean
    # F test p-value
    p_val = 1 - pf((n1 + n2 - 1 - p) * T_sq/(p*(n1 + n2 - 2)),
                   df1 = p, df2 = n1 + n2 - 1 -p)

    print(paste0("The Hotelling's T-squared test on period = ", i,
                 " and period = ", j))
    print(paste0("has a p-value of: ", p_val))
  }
}
```

```
## [1] "The Hotelling's T-squared test on period = 1 and period = 2"
## [1] "has a p-value of: 0.813943332652345"
## [1] "The Hotelling's T-squared test on period = 1 and period = 3"
## [1] "has a p-value of: 0.0203528071261222"
## [1] "The Hotelling's T-squared test on period = 2 and period = 3"
## [1] "has a p-value of: 0.0190790259695431"
```

At a $\alpha = 0.05$ significance level, the time period pairs $(\text{period} = 1, \text{period} = 3)$ and $(\text{period} = 2, \text{period} = 3)$ differ across the 4 variables significantly, treating each pair of time periods as two independent samples of equal covariance structure.

(c)

i)

$$\binom{4}{1} \cdot \binom{3}{2} = 4 \times 3 = 12$$

ii)

First, we need to know the critical t-value for the 85% Bonferroni simultaneous confidence interval. Given that there are p components (i.e., $p = 4$),our desired Bonferroni significance level α (i.e., $\alpha = 1 - 0.85 = 0.15$), and with n_1, n_2 observations in periods 1 and 2 respectively, the formula for the critical value is:

$$t_{n_1+n_2-1,\alpha/2p}$$

The numerical critical t-value can be calculated as below:

```
n_1 <-  skull %>%
  filter(period == 1) %>%
  nrow(.)

n_2 <-  skull %>%
  filter(period == 2) %>%
  nrow(.)

n = n_1 + n_2

alpha <- 1 - 0.85

p <- ncol(skull) - 1

q <- 1 - (alpha / (2 * p))

cr = qt(q, df = n - 1)
```

The critical t-value from the 85% Bonferroni simultaneous confidence interval for comparing component i between periods 1 and 2 is 2.1282897.

The standard error for component i can be expressed as:

$$\sqrt{\frac{\frac{1}{n_1-1}\sum_{j=1}^{n_1}(x_{1i,j} - \bar{x}_{1i})^2 + \frac{1}{n_2-1}\sum_{j=1}^{n_2}(x_{2i,j} - \bar{x}_{2i})^2}{n_1 + n_2}}$$

Therefore, the confidence interval for the difference in mean becomes:

$$\left((\bar{x}_{1i} - \bar{x}_{2i}) - t_{n_1+n_2-1,\alpha/2p} \cdot \sqrt{\frac{\frac{1}{n_1-1}\sum_{j=1}^{n_1}(x_{1i,j} - \bar{x}_{1i})^2 + \frac{1}{n_2-1}\sum_{j=1}^{n_2}(x_{2i,j} - \bar{x}_{2i})^2}{n_1+n_2}},\right.$$
$$\left.(\bar{x}_{1i} - \bar{x}_{2i}) + t_{n_1+n_2-1,\alpha/2p} \cdot \sqrt{\frac{\frac{1}{n_1-1}\sum_{j=1}^{n_1}(x_{1i,j} - \bar{x}_{1i})^2 + \frac{1}{n_2-1}\sum_{j=1}^{n_2}(x_{2i,j} - \bar{x}_{2i})^2}{n_1+n_2}}\right)$$

Or, using some numerical values:

$$\left((\bar{x}_{1i} - \bar{x}_{2i}) - 2.1282897 \cdot \sqrt{\frac{\frac{1}{29}\sum_{j=1}^{30}(x_{1i,j} - \bar{x}_{1i})^2 + \frac{1}{29}\sum_{j=1}^{30}(x_{2i,j} - \bar{x}_{2i})^2}{60}},\right.$$
$$\left.(\bar{x}_{1i} - \bar{x}_{2i}) + 2.1282897 \cdot \sqrt{\frac{\frac{1}{29}\sum_{j=1}^{30}(x_{1i,j} - \bar{x}_{1i})^2 + \frac{1}{29}\sum_{j=1}^{30}(x_{2i,j} - \bar{x}_{2i})^2}{60}}\right)$$

Problem 3.

```r
basket = read.table("basketball.csv", header = T, sep = ",")
```

(a)

```r
# without interaction
mfit_1 <-  lm(cbind(Field, Freethrow, Avgpt) ~ ., data = basket)
# with interaction
mfit_2 <-  lm(cbind(Field, Freethrow, Avgpt) ~ .*., data = basket)
```

```r
cor(mfit_1$residuals)
```

```
##               Field Freethrow     Avgpt
## Field     1.0000000 0.1661680 0.4231698
## Freethrow 0.1661680 1.0000000 0.2509584
## Avgpt     0.4231698 0.2509584 1.0000000
```

The residuals between the two response variables, average points scored per game and percent of successful field goals (out of 100 attempted), seem to have the highest correlation in residuals in the linear models without interaction.

```r
cor(mfit_2$residuals)
```

```
##               Field Freethrow     Avgpt
## Field     1.0000000 0.1491372 0.3989468
## Freethrow 0.1491372 1.0000000 0.2283571
## Avgpt     0.3989468 0.2283571 1.0000000
```

The linear models with interaction also had the same result. However, the correlation between the two response variables' residuals became lower.

(b)

```r
manova(cbind(Field, Freethrow, Avgpt) ~ Height + Weight, data = basket)
```

```
## Call:
##    manova(cbind(Field, Freethrow, Avgpt) ~ Height + Weight, data = basket)
##
## Terms:
##                     Height    Weight Residuals
## Field               0.0416    0.0059    0.1220
## Freethrow           0.0357    0.0095    0.4863
## Avgpt               8.7576   13.7782 1821.9296
## Deg. of Freedom          1         1        51
##
## Residual standard errors: 0.04891026 0.09764684 5.976965
## Estimated effects may be unbalanced
```

```r
summary(manova(cbind(Field, Freethrow, Avgpt) ~ Height + Weight,
               data = basket), test = "Wilks")
```

```
##           Df   Wilks approx F num Df den Df    Pr(>F)
## Height     1 0.64156   9.1256      3     49 6.646e-05 ***
## Weight     1 0.92500   1.3244      3     49    0.2771
## Residuals 51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
manova(cbind(Field, Freethrow, Avgpt) ~ Weight + Height, data = basket)
```

```
## Call:
##    manova(cbind(Field, Freethrow, Avgpt) ~ Weight + Height, data = basket)
##
## Terms:
##                    Weight    Height Residuals
## Field              0.0451    0.0024    0.1220
## Freethrow          0.0448    0.0005    0.4863
## Avgpt              0.1787   22.3571 1821.9296
## Deg. of Freedom         1         1        51
##
## Residual standard errors: 0.04891026 0.09764684 5.976965
## Estimated effects may be unbalanced
```

```r
summary(manova(cbind(Field, Freethrow, Avgpt) ~ Weight + Height,
               data = basket), test = "Wilks")
```

```
##           Df   Wilks approx F num Df den Df    Pr(>F)
## Weight     1 0.63088   9.5566      3     49 4.464e-05 ***
## Height     1 0.94814   0.8935      3     49    0.4512
## Residuals 51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The `Weight` variable seems to be more important based on the MANOVA tables.

Problem 4.

```
mat = matrix(0, 9, 9)
mat[row(mat) <= col(mat)] = scan("T12-13.DAT")
X = t(mat)
```

(a)

```
cmdscale(as.dist(X), k = 3)
```

```
##               [,1]        [,2]        [,3]
##  [1,]  0.5119010 -0.27797661  0.24210462
##  [2,] -1.3184960  0.69177869  0.62299269
##  [3,]  0.4696574 -0.07075632  0.18553022
##  [4,]  0.3874028  0.08774518  0.04893247
##  [5,]  0.2336943  0.29550962 -0.32518484
##  [6,]  0.4688497  0.13734912 -0.21876261
##  [7,]  0.5814134 -0.34919001  0.45732159
##  [8,] -1.1180751 -1.12218941 -0.31595964
##  [9,] -0.2163475  0.60772973 -0.69697450
```

```
cmdscale(as.dist(X), k = 4)
```

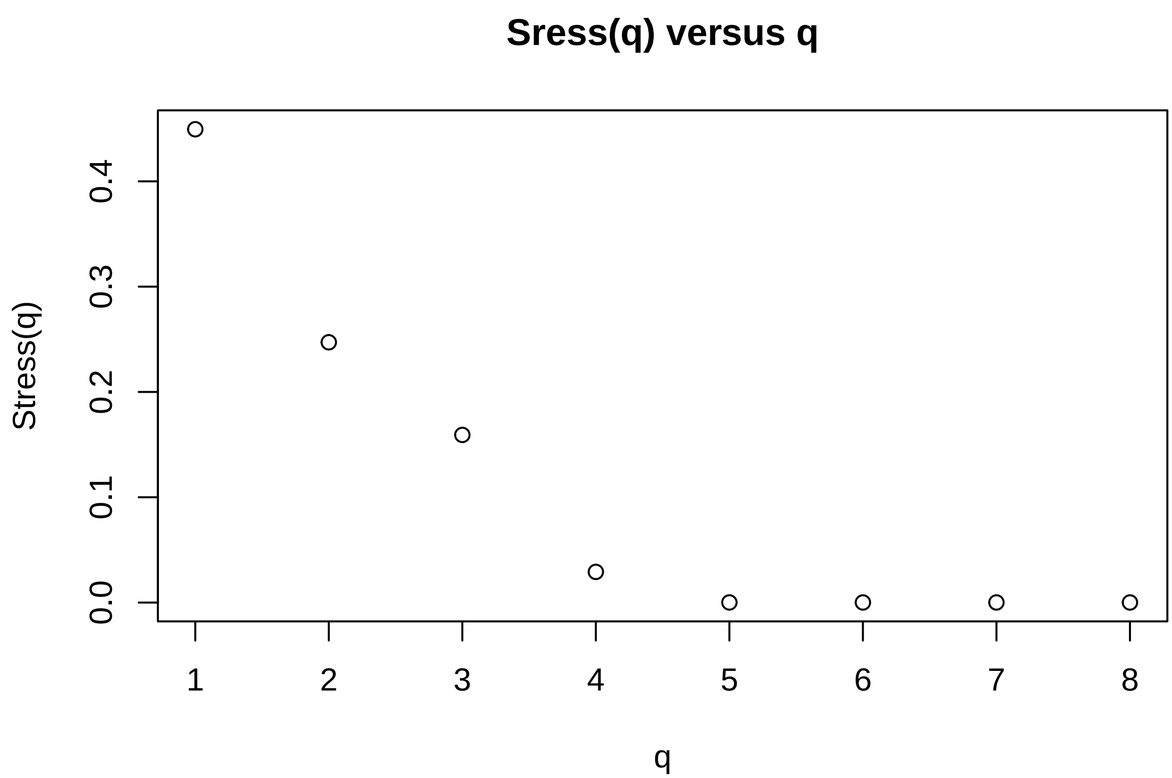
```
##               [,1]        [,2]        [,3]        [,4]
##  [1,]  0.5119010 -0.27797661  0.24210462  0.67644341
##  [2,] -1.3184960  0.69177869  0.62299269  0.04985327
##  [3,]  0.4696574 -0.07075632  0.18553022 -0.30157380
##  [4,]  0.3874028  0.08774518  0.04893247 -0.34374144
##  [5,]  0.2336943  0.29550962 -0.32518484 -0.05196534
##  [6,]  0.4688497  0.13734912 -0.21876261  0.13932144
##  [7,]  0.5814134 -0.34919001  0.45732159 -0.17841618
##  [8,] -1.1180751 -1.12218941 -0.31595964 -0.05212137
##  [9,] -0.2163475  0.60772973 -0.69697450  0.06220001
```

```
cmdscale(as.dist(X), k = 5)
```

```
##               [,1]        [,2]        [,3]        [,4]         [,5]
##  [1,]  0.5119010 -0.27797661  0.24210462  0.67644341  0.118956421
##  [2,] -1.3184960  0.69177869  0.62299269  0.04985327 -0.023615842
##  [3,]  0.4696574 -0.07075632  0.18553022 -0.30157380  0.058979347
##  [4,]  0.3874028  0.08774518  0.04893247 -0.34374144  0.101917878
##  [5,]  0.2336943  0.29550962 -0.32518484 -0.05196534  0.121452747
##  [6,]  0.4688497  0.13734912 -0.21876261  0.13932144 -0.281320799
##  [7,]  0.5814134 -0.34919001  0.45732159 -0.17841618 -0.101635526
##  [8,] -1.1180751 -1.12218941 -0.31595964 -0.05212137 -0.005024992
##  [9,] -0.2163475  0.60772973 -0.69697450  0.06220001  0.010290767
```

(b)

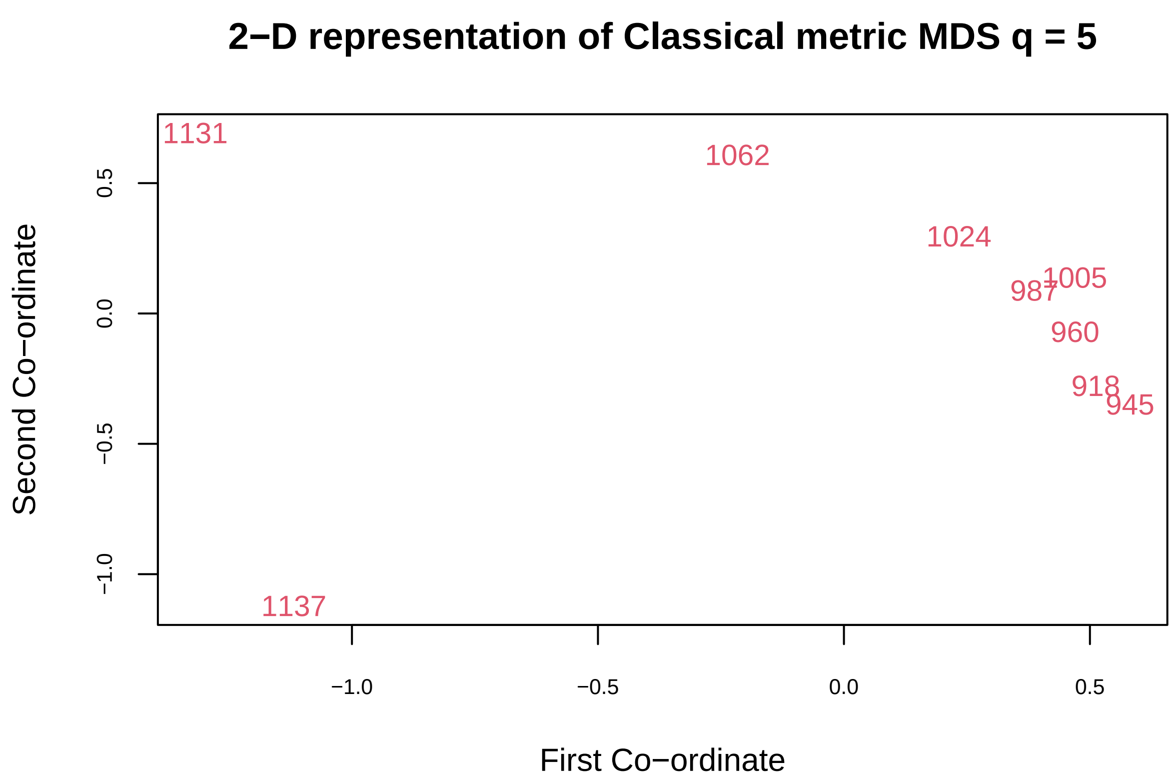
```
Stress_q <- function(q){
  x <- as.dist(X)
  mds <- dist(cmdscale(x, k = q), method = "euclidean")
  numerator = sum((x - mds)^2)
  denominator = sum(x^2)
  return(sqrt(numerator/denominator))
}

q <- 1:8

stresses <- map_dbl(q, Stress_q)

plot(q, stresses, xlab = "", ylab = "")
title(main = "Sress(q) versus q",
      xlab = "q", ylab = "Stress(q)")
```



As the dimension increases, the difference in original distance and the distance from classical metric multidimensional scaling decreases. Therefore, the stress decreases and approaches 0 as q increases to 8.

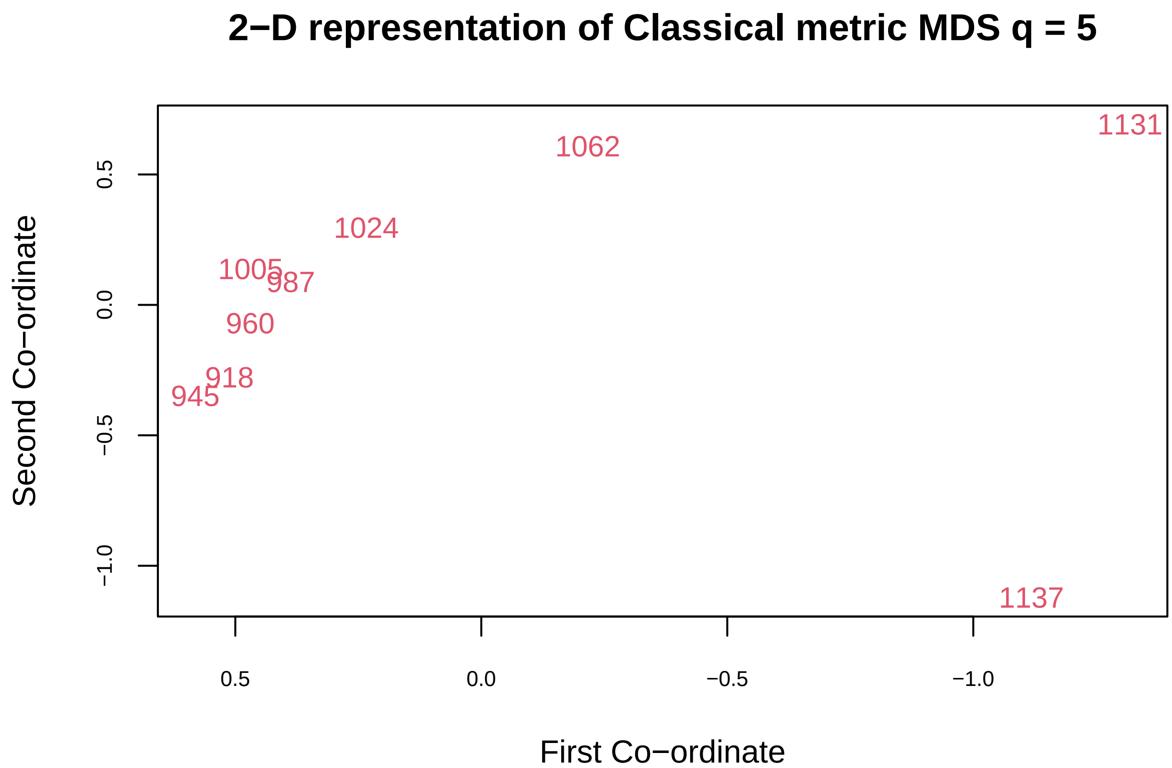
(c)

```
X5 = cmdscale(as.dist(X), k = 5)
plot(X5[,1], X5[,2],type="n",
     xlab = "", ylab = "", cex.axis=.7)
text(X5[,1], X5[,2],c("918", "1131", "960", "987", "1024", "1005", "945",
                      "1137", "1062"), cex=.9,lwd=2,col=2)
title(main="2-D representation of Classical metric MDS q = 5",
      xlab = "First Co-ordinate", ylab = "Second Co-ordinate")
```



With the exception of A.D. 1137 co-ordinate, there seems to be a slight trend that would be more apparent in the reversed scale for the first co-ordinate

```
X5 = cmdscale(as.dist(X), k = 5)
plot(X5[,1], X5[,2],type="n",
     xlab = "", ylab = "", cex.axis=.7,
     xlim = rev(range(X5[,1])))
text(X5[,1], X5[,2],c("918", "1131", "960", "987", "1024", "1005", "945",
                      "1137", "1062"), cex=.9,lwd=2,col=2)
title(main="2-D representation of Classical metric MDS q = 5",
      xlab = "First Co-ordinate", ylab = "Second Co-ordinate")
```


2−D representation of Classical metric MDS q = 5
Second Co−ordinate
First Co−ordinate
0.5
0.0
−0.5
−1.0
1131
1062
1024
1005
987
960
918
945
1137

Problem 5.

```r
data = read.table("HairEyeAll.txt")
rownames(data) = c("Black","Brown", "Red", "Blond") # for variable Hair
colnames(data)= c("Brown","Blue", "Hazel", "Green") # for variable Eye
```

(a)

```r
Q5 = as.matrix(data)
P = Q5/sum(Q5)
round(P, 2)
```

```
##       Brown Blue Hazel Green
## Black  0.11 0.03  0.03  0.01
## Brown  0.20 0.14  0.09  0.05
## Red    0.04 0.03  0.02  0.02
## Blond  0.01 0.16  0.02  0.03
```

(b)

```r
hair = Q5 %*% c(1,1,1,1)
eye = c(1,1,1,1) %*% Q5
E = hair %*% eye / sum(Q5)
round(E, 1)
```

```
##       Brown  Blue Hazel Green
## Black  40.1  39.2  17.0  11.7
## Brown 106.3 103.9  44.9  30.9
## Red    26.4  25.8  11.2   7.7
## Blond  47.2  46.1  20.0  13.7
```

(c)

```r
round((data - E)^2/E, 2) %>%
  kable()
```

	Brown	Blue	Hazel	Green
Black	19.35	9.42	0.23	3.82
Brown	1.52	3.80	1.83	0.12
Red	0.01	2.99	0.73	5.21
Blond	34.23	49.70	4.96	0.38

The null and alternative hypotheses are:

$$\begin{cases} H_0: & \text{Hair variable and Eye variable are independent} \\ H_A: & \text{Hair variable and Eye variable are not independent} \end{cases}$$

The degree of freedom of the χ^2 is $df = (I-1)(J-1) = 3 \cdot 3 = 9$.

The p-value of the test is:

```
chi_sq = sum((data - E)^2/E) # getting the quantile
1 - pchisq(chi_sq, df = 9)
```

```
## [1] 0
```

or similarly,

```
chisq.test(data)
```

```
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 138.29, df = 9, p-value < 2.2e-16
```

If we calculate the total inertia

```
Inertia = 0
hair_p = hair/sum(Q5)
eye_p = eye/sum(Q5)

for(i in 1:4){
  for(j in 1:4){
   Inertia = Inertia + (P[i, j] - hair_p[i]*eye_p[j])^2/(hair_p[i] * eye_p[j])
  }
}
```

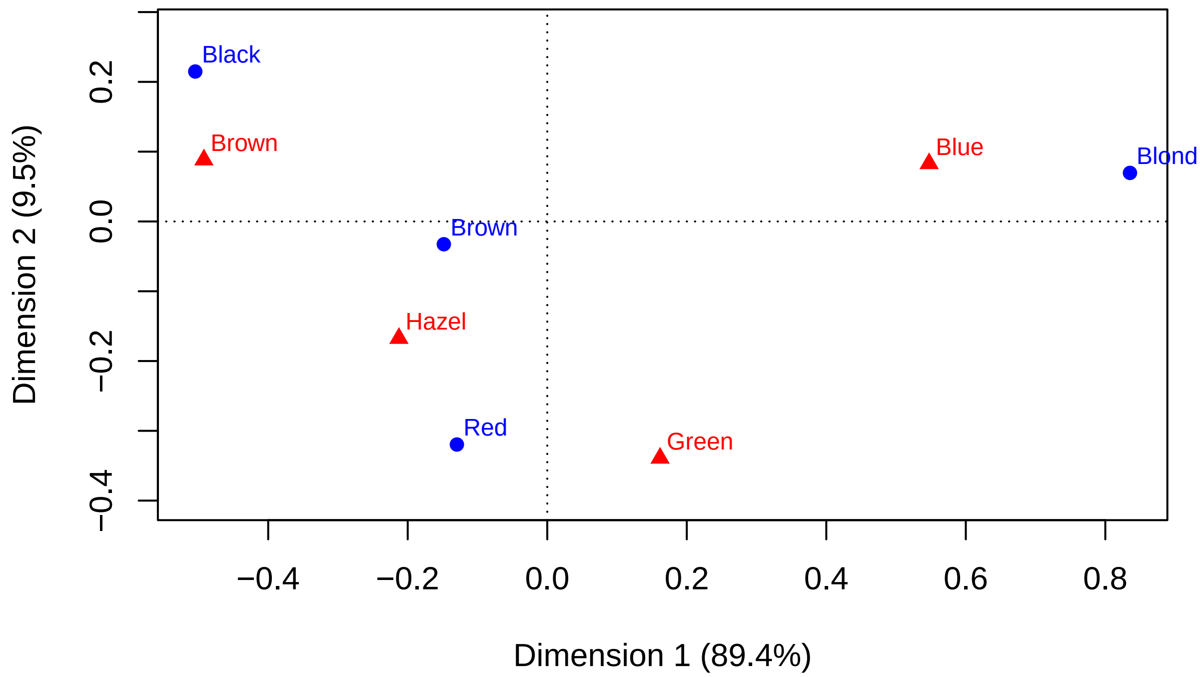
$$\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(p_{i,j} - r_i c_j)^2}{r_i c_j} \approx 0.23$$

From the test-statistics

$$\frac{\chi^2}{n} = \frac{138.29}{592} \approx 0.23$$

(d)

```
plot(ca(data), map = "symmetric")
```



About 98.9% of variation in the data is captured in the 2-dimensional CA plot.

Based on the quadrants of the two dimension plane, there seems to be some association between Eye category and Hair category. The Brown eye color is in the only eye color in the same quadrant as the Black hair color. The Blue eye color is the only eye color in the same quadrant as the Blue eye color. The Green eye color is in the same quadrant as no eye color. Though Hazel is in the same quadrant with two hair colors (Brown and Red) and Brown hair color is close to the origin, the remaining combinations seem to show clear association in the plot.

Problem 6.

(a)

i) and ii)

We know from the corollary on conditional distribution of multivariate normal:

$$(\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2) \sim N(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}) \ \cdots \ (6-1)$$

Similarly,

$$(\mathbf{X}_1 \mid \mathbf{X}_3 = \mathbf{x}_3) \sim N(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}(\mathbf{x}_3 - \boldsymbol{\mu}_3), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{31}) \ \cdots \ (6-2)$$

If we expand this to $\mathbf{X}_2$ and $\mathbf{X}_3$ as well:

$$\begin{aligned}(\mathbf{X}_2 \mid \mathbf{X}_3 = \mathbf{x}_3) \sim& N(\boldsymbol{\mu}_2 + \mathbf{0}\boldsymbol{\Sigma}_{33}^{-1}(\mathbf{x}_3 - \boldsymbol{\mu}_3), \boldsymbol{\Sigma}_{22} - \mathbf{0}\boldsymbol{\Sigma}_{33}^{-1}\mathbf{0})\\ =& N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})\end{aligned}$$

However, we also know from the corollary that

$$\mathbf{X}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$$

Therefore, we may conclude that $\mathbf{X}_2 \perp\!\!\!\perp \mathbf{X}_3$.

Given this relationship:

$$\begin{aligned}&(\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2, \mathbf{X}_3 = \mathbf{x}_3)\\
=&\frac{f_{\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)}{f_{\mathbf{X}_2,\mathbf{X}_3}(\mathbf{x}_2,\mathbf{x}_3)}\\
&= \frac{f_{\mathbf{X}_1}(\mathbf{x}_1)f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2)f_{\mathbf{X}_3|\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}_3)}{f_{\mathbf{X}_2,\mathbf{X}_3}(\mathbf{x}_2,\mathbf{x}_3)}\\
&= \frac{f_{\mathbf{X}_1}(\mathbf{x}_1)f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2)f_{\mathbf{X}_3|\mathbf{X}_1}(\mathbf{x}_3)}{f_{\mathbf{X}_2}(\mathbf{x}_2)f_{\mathbf{X}_3}(\mathbf{x}_3)}\\
&(\because \mathbf{X}_2 \perp\!\!\!\perp \mathbf{X}_3)\\
&= \frac{f_{\mathbf{X}_1}(\mathbf{x}_1)f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2)}{f_{\mathbf{X}_2}(\mathbf{x}_2)} \cdot \frac{f_{\mathbf{X}_1}(\mathbf{x}_1)f_{\mathbf{X}_3|\mathbf{X}_1}(\mathbf{x}_3)}{f_{\mathbf{X}_3}(\mathbf{x}_3)} \cdot \frac{1}{f_{\mathbf{X}_1}(\mathbf{x}_1)}\\
&= \frac{f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1)f_{\mathbf{X}_1|\mathbf{X}_3}(\mathbf{x}_1)}{f_{\mathbf{X}_1}(\mathbf{x}_1)}\\
&(\because \text{Baye's Rule})\end{aligned}$$

We also know from the corollary that:

$$\mathbf{X}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \ \cdots \ (6-3)$$

Therefore, from (6 - 1), (6 - 2), and (6 - 3) and the conditional probability form above:

$$
\begin{aligned}
&(\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2, \mathbf{X}_3 = \mathbf{x}_3) \\
\sim & N(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) + \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}(\mathbf{x}_3 - \boldsymbol{\mu}_3), \\
&\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{31})
\end{aligned}
$$

Therefore,

$$\mathbf{E}(\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2, \mathbf{X}_3 = \mathbf{x}_3) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) + \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}(\mathbf{x}_3 - \boldsymbol{\mu}_3)$$

and

$$\mathbf{Var}(\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2, \mathbf{X}_3 = \mathbf{x}_3) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{31}$$

(b)

We know from the independence shown in `(a)` that:

$$\mathbf{X}_2 + \mathbf{X}_3 \sim N(\boldsymbol{\mu}_2 + \boldsymbol{\mu}_3, \boldsymbol{\Sigma}_{22} + \boldsymbol{\Sigma}_{33})$$

Since $\mathbf{X}_2 \perp\!\!\!\perp \mathbf{X}_3$,

$$\mathbf{Cov}(\mathbf{X}_1, \mathbf{X}_2 + \mathbf{X}_3) = \mathbf{Cov}(\mathbf{X}_1, \mathbf{X}_2) + \mathbf{Cov}(\mathbf{X}_1, \mathbf{X}_3)$$

Therefore, $\mathbf{X}_1$ and $\mathbf{X}_2 + \mathbf{X}_3$ have the multivariate normal distribution:

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 + \mathbf{X}_3 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 + \boldsymbol{\mu}_3 \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} + \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{21} + \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{22} + \boldsymbol{\Sigma}_{33} \end{bmatrix} \right)$$

We can directly infer to the corollary for the conditional distribution as:

$$
\begin{aligned}
&(\mathbf{X}_1 \mid \mathbf{X}_2 + \mathbf{X}_3 = \mathbf{x}_0) \\
\sim & N(\boldsymbol{\mu}_1 + (\boldsymbol{\Sigma}_{12} + \boldsymbol{\Sigma}_{13})(\boldsymbol{\Sigma}_{22} + \boldsymbol{\Sigma}_{33})^{-1}(\mathbf{x}_0 - (\boldsymbol{\mu}_2 + \boldsymbol{\mu}_3)), \\
&\boldsymbol{\Sigma}_{11} - (\boldsymbol{\Sigma}_{12} + \boldsymbol{\Sigma}_{13})(\boldsymbol{\Sigma}_{22} + \boldsymbol{\Sigma}_{33})^{-1}(\boldsymbol{\Sigma}_{21} + \boldsymbol{\Sigma}_{31}))
\end{aligned}
$$