

HW #4

I. Data

- Trt 1: $[6 \ 7]$, $[5 \ 9]$, $[8 \ 6]$, $[4 \ 9]$, $[7 \ 9]$
- Trt 2: $[3 \ 3]$, $[1 \ 6]$, $[2 \ 3]$
- Trt 3: $[2 \ 3]$, $[5 \ 1]$, $[3 \ 1]$, $[2 \ 3]$

Ia. Break up the observations (indexed by j) into

mean, trt (indexed by t), + residual components

$$x_{tj} = \bar{x} + (\bar{x}_t - \bar{x}) + (x_{tj} - \bar{x}_t)$$

by constructing data arrays for each component variable.

- First Component:

$$\cdot \bar{x} = \frac{1}{12}(6+5+8+4+7+3+1+2+2+5+3+2) = \frac{1}{12}(48) = 4$$

$$\cdot \bar{x}_1 = \frac{1}{5}(6+5+8+4+7) = \frac{1}{5}(30) = 6$$

$$\cdot \bar{x}_2 = \frac{1}{3}(3+1+2) = \frac{1}{3}(6) = 2$$

$$\cdot \bar{x}_3 = \frac{1}{4}(2+5+3+2) = \frac{1}{4}(12) = 3$$

- Array:

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

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

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

- Second Component

$$\cdot \bar{x} = \frac{1}{12}(7+9+6+9+9+3+6+3+3+1+1+3) = \frac{1}{12}(60) = 5$$

$$\cdot \bar{x}_1 = \frac{1}{5}(7+9+6+9+9) = \frac{1}{5}(40) = 8$$

$$\cdot \bar{x}_2 = \frac{1}{3}(3+6+3) = \frac{1}{3}(12) = 4$$

$$\cdot \bar{x}_3 = \frac{1}{4}(3+1+1+3) = \frac{1}{4}(8) = 2$$

• Array:

$$\begin{bmatrix} 7 & 9 & 6 & 9 & 9 \end{bmatrix} = \begin{bmatrix} 5 & 5 & 5 & 5 & 5 \end{bmatrix} + \begin{bmatrix} 8-5 & 8-5 & 8-5 & 8-5 & 8-5 \end{bmatrix}$$
$$+ \begin{bmatrix} 5 & 5 & 5 & 5 & 5 \end{bmatrix} + \begin{bmatrix} 4-5 & 4-5 & 4-5 & 4-5 & 4-5 \end{bmatrix}$$
$$+ \begin{bmatrix} 5 & 5 & 5 & 5 & 5 \end{bmatrix} + \begin{bmatrix} 2-5 & 2-5 & 2-5 & 2-5 & 2-5 \end{bmatrix}$$
$$+ \begin{bmatrix} 7-8 & 9-8 & 6-8 & 9-8 & 9-8 \end{bmatrix}$$
$$+ \begin{bmatrix} 3-4 & 6-4 & 3-4 & 3-4 & 3-4 \end{bmatrix}$$
$$+ \begin{bmatrix} 3-2 & 1-2 & 1-2 & 1-2 & 1-2 \end{bmatrix}$$

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

1b. Use part (a) to construct the one-way MANOVA table.

• 3 levels of treatment $\Rightarrow g = 3 \Rightarrow g-1 = 2$

• 12 observations $\Rightarrow n = 12 \Rightarrow n-g = 9$

$$\mathbf{B} = \sum_{l=1}^3 n_l (\bar{x}_l - \bar{x})(\bar{x}_l - \bar{x})^T$$

$$= n_1 (\bar{x}_1 - \bar{x})(\bar{x}_1 - \bar{x})^T + n_2 (\bar{x}_2 - \bar{x})(\bar{x}_2 - \bar{x})^T + n_3 (\bar{x}_3 - \bar{x})(\bar{x}_3 - \bar{x})^T$$

$$= 5 \left(\begin{bmatrix} 6 \\ 8 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \end{bmatrix} \right) \left(\begin{bmatrix} 6 \\ 8 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \end{bmatrix} \right)^T + 3 \left(\begin{bmatrix} 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \end{bmatrix} \right) \left(\begin{bmatrix} 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \end{bmatrix} \right)^T$$

$$+ 4 \left(\begin{bmatrix} 3 \\ 2 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \end{bmatrix} \right) \left(\begin{bmatrix} 3 \\ 2 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \end{bmatrix} \right)^T$$

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

$$= 5 \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} + 3 \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} + 4 \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 20 & 30 \\ 30 & 45 \end{bmatrix} + \begin{bmatrix} 12 & 6 \\ 6 & 3 \end{bmatrix} + \begin{bmatrix} 4 & 12 \\ 12 & 36 \end{bmatrix}$$

$$= \begin{bmatrix} 36 & 48 \\ 48 & 84 \end{bmatrix}$$

$$\begin{aligned}
W &= \sum_{\ell=1}^3 \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)(x_{\ell j} - \bar{x}_\ell)^T \\
&= \sum_{j=1}^5 (x_{1j} - \bar{x}_1)(x_{1j} - \bar{x}_1)^T + \sum_{j=1}^3 (x_{2j} - \bar{x}_2)(x_{2j} - \bar{x}_2)^T + \sum_{j=1}^4 (x_{3j} - \bar{x}_3)(x_{3j} - \bar{x}_3)^T \\
&= \left(\begin{bmatrix} 6 \\ 7 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right) \left(\begin{bmatrix} 6 \\ 7 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right)^T + \left(\begin{bmatrix} 5 \\ 9 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right) \left(\begin{bmatrix} 5 \\ 9 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right)^T + \left(\begin{bmatrix} 8 \\ 6 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right) \left(\begin{bmatrix} 8 \\ 6 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right)^T \\
&\quad + \left(\begin{bmatrix} 4 \\ 9 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right) \left(\begin{bmatrix} 4 \\ 9 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right)^T + \left(\begin{bmatrix} 7 \\ 9 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right) \left(\begin{bmatrix} 7 \\ 9 \end{bmatrix} - \begin{bmatrix} 6 \\ 8 \end{bmatrix}\right)^T + \left(\begin{bmatrix} 3 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \end{bmatrix}\right) \left(\begin{bmatrix} 3 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \end{bmatrix}\right)^T \\
&\quad + \left(\begin{bmatrix} 1 \\ 6 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \end{bmatrix}\right) \left(\begin{bmatrix} 1 \\ 6 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \end{bmatrix}\right)^T + \left(\begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \end{bmatrix}\right) \left(\begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \end{bmatrix}\right)^T + \left(\begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) \left(\begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}\right)^T \\
&\quad + \left(\begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) \left(\begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}\right)^T + \left(\begin{bmatrix} 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) \left(\begin{bmatrix} 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}\right)^T + \left(\begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) \left(\begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}\right)^T \\
&= \begin{bmatrix} 0 \\ -1 \end{bmatrix} [0 \ -1] + \begin{bmatrix} -1 \\ 1 \end{bmatrix} [-1 \ 1] + \begin{bmatrix} 2 \\ -2 \end{bmatrix} [2 \ -2] + \begin{bmatrix} -2 \\ 1 \end{bmatrix} [-2 \ 1] + \begin{bmatrix} 1 \\ 1 \end{bmatrix} [1 \ 1] \\
&\quad + \begin{bmatrix} 1 \\ -1 \end{bmatrix} [1 \ -1] + \begin{bmatrix} -1 \\ 2 \end{bmatrix} [-1 \ 2] + \begin{bmatrix} 0 \\ -1 \end{bmatrix} [0 \ -1] \\
&\quad + \begin{bmatrix} -1 \\ 1 \end{bmatrix} [-1 \ 1] + \begin{bmatrix} 2 \\ -1 \end{bmatrix} [2 \ -1] + \begin{bmatrix} 0 \\ -1 \end{bmatrix} [0 \ -1] + \begin{bmatrix} -1 \\ 1 \end{bmatrix} [-1 \ 1] \\
&= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} + \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
&\quad + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\
&\quad + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 10 & -6 \\ -6 & 8 \end{bmatrix} + \begin{bmatrix} 2 & -3 \\ -3 & 6 \end{bmatrix} + \begin{bmatrix} 6 & -4 \\ -4 & 4 \end{bmatrix} \\
&= \begin{bmatrix} 18 & -13 \\ -13 & 18 \end{bmatrix}
\end{aligned}$$

Thus, the MANOVA table is:

<u>Source of Variation</u>	<u>Matrix of Sum of Squares & Cross-Products</u>	<u>df</u>
Treatments	$B = \begin{bmatrix} 36 & 48 \\ 48 & 84 \end{bmatrix}$	2
Residuals	$W = \begin{bmatrix} 18 & -13 \\ -13 & 18 \end{bmatrix}$	9
Total	$B+W = \begin{bmatrix} 54 & 35 \\ 35 & 102 \end{bmatrix}$	11

1a. Evaluate $|W|$, $|B|$, $\Lambda^* = \frac{|W|}{|B+W|}$

$$\cdot |W| = \det \left(\begin{bmatrix} 18 & -13 \\ -13 & 18 \end{bmatrix} \right)$$

$$= (18)(18) - (-13)(-13)$$

$$= 324 - 169$$

$$\Rightarrow |W| = 155$$

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

$$= (36)(84) - (48)(48)$$

$$= 3024 - 2304$$

$$\Rightarrow |B| = 720$$

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

$$= (54)(102) - (35)(35)$$

$$= 5508 - 1225$$

$$= 4283$$

• So, $\Lambda^* = \frac{|w|}{|B+w|}$

$$\Rightarrow \Lambda^* = \frac{155}{4283} \approx 0.036$$

Id. Use Bartlett's approximation $-\log(\Lambda^*)(n-1 - \frac{p+1}{2}) \approx_{H_0} \chi^2_{p(g-1)}$ to find the observed p-value from the data.

$$-\log(\Lambda^*)(n-1 - \frac{p+1}{2}) \approx -\log(0.036)(12-1 - \frac{2+3}{2}) = -\frac{17}{2} \log(0.036) \approx 28.211$$

$$\chi^2_{p(g-1)} = \chi^2_{2(3-1)} = \chi^2_4$$

Let $X \sim \chi^2_4$. Using R,

$$p = P(X > 28.211)$$

$$\Rightarrow p \approx 1.130 \times 10^{-5}$$

Ie. Write the hypothesis test (H_0, H_1) that the p-value in Part (d) refers to. Describe in words.

$$\begin{cases} H_0: \tau_1 = \tau_2 = \tau_3 = 0 \\ H_1: \tau_t \neq 0 \text{ for some } t \in \{1, 2, 3\} \end{cases}$$

where τ_t is the treatment effect for treatment t ,

$0 = [0]$, & we have imposed the constraint

$$\sum_{t=1}^3 n_t \tau_t = 0 \text{ for identification purposes.}$$

• Intuitively, the null hypothesis is that all three treatments have no effect on individuals' outcomes $x = (x_1, x_2)$ (so their effects—or lack thereof—are no different from one another's).

• The alternative hypothesis is that at least one treatment has a nonzero effect on individuals' outcomes $x = (x_1, x_2)$ (so that at least one treatment's effect is different than the others').

STAT 32950: Homework 4

Robert Winter

Table of Contents

1	Exercise 2: MANOVA and Confidence Region	1
1.1	Part (a)	2
1.1.1	Part (i)	2
1.1.2	Part (ii)	3
1.1.3	Part (iii)	3
1.1.4	Part (iv)	5
1.2	Part (b)	5
1.3	Part (c)	7
1.3.1	Part (i)	7
1.3.2	Part (ii)	7
1.3.3	Part (iii)	9
2	Exercise 3: Multidimensional Scaling	10
2.1	Part (a)	10
2.2	Part (b)	11
2.3	Part (c)	13
3	Exercise 4: Correspondence Analysis	14
3.1	Part (a)	14
3.2	Part (b)	16
3.3	Part (c)	16
3.4	Part (d)	17

1 Exercise 2: MANOVA and Confidence Region

Researchers have suggested that a change in skull size over time is evidence of the interbreeding of a resident population with immigrant populations. Four measurements were made of male Egyptian skulls for three different time periods: Period 1 is 4000 B.C., Period 2 is 3300 B.C., and Period 3 is 1850 B.C. The measured variables are

- X_1 = maximum breadth of skull (mm)
- X_2 = basibregmatic height of skull (mm)
- X_3 = basialveolar length of skull (mm)
- X_4 = nasal height of skull (mm)

The data are in T6-13.DAT.

```
skull =
  ↵  read.table("C:/Users/rewin/OneDrive/Documents/STAT_32950/Homework/HW4/T6-13.DAT")
  colnames(skull) = c("x1", "x2", "x3", "x4", "period")
  skull = skull %>%
    mutate(period = as.factor(period))
```

1.1 Part (a)

Conduct a one-way MANOVA (periods as “treatments”) of the Egyptian skull data.

1.1.1 Part (i)

State the numerical values of p (the number of component variables), g (the number of samples or treatments), and sample sizes n_i for $i = 1, \dots, g$.

```
skull %>% group_by(period) %>% summarize(n = n())
```

```
# A tibble: 3 x 2
  period      n
  <fct>   <int>
1 1          30
2 2          30
3 3          30
```

In this case, we have:

- $p = 4$ (corresponding to the four different measurements for each skull),
- $g = 3$ (corresponding to the three different time periods), and
- $n_1 = n_2 = n_3 = 30$ (since we have data for 30 skulls from each of the three time periods, making this a balanced dataset).

1.1.2 Part (ii)

Write down the null hypothesis H_0 that the MANOVA table is used for (define your notations if you use any).

For each time period $t \in \{1, 2, 3\}$ and for each skull $j \in \{1, \dots, 30\}$ within each of those time periods, let

$$X_{tj} = \mu + \tau_t + \varepsilon_{tj},$$

where X_{tj} is a vector consisting of the four measurements for the j^{th} skull in time period t , μ is a vector consisting of the population mean skull measurements across all three time periods, τ_t is the average deviation from the total population mean skull measurements among the sub-population of skulls from time period t , and ε_{tj} is the j^{th} skull from time period t 's personal deviation from the mean skull measurements associated with its time period (i.e., $\mu + \tau_t$).

The MANOVA table will test the null hypothesis

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \mathbf{0},$$

where we have imposed the usual constraint that $\tau_1 + \tau_2 + \tau_3 = 0$ for identification purposes. Intuitively, the MANOVA table tests the null hypothesis that, on average, there are no differences between male Egyptians' skull measurements across the three time periods. Equivalently (since we have imposed the sum-zero constraint), the MANOVA table tests the null hypothesis that, on average, each time period has no effect on male Egyptians' skull measurements.

1.1.3 Part (iii)

Compute Wilks' Λ^* and the related F statistic under H_0 (ref. p. 303 Table 6.3 in Johnson & Wichern, 6th ed.). Use $\alpha = 0.05$ to conduct the hypothesis test and state your conclusion.

Using the `manova` function in R, we find that the between- and within-group matrices are (approximately) as follows:

```
q2manova = manova(cbind(skull$x1, skull$x2, skull$x3, skull$x4) ~ period,
                    data = skull)

B = summary(q2manova, test = "Wilks")$SS[1]$period %>% round(2)
W = summary(q2manova, test = "Wilks")$SS[2]$Residuals %>% round(2)

print("B = "); print(B)
```

```
[1] "B = "
[,1] [,2] [,3] [,4]
[1,] 150.20 20.30 -161.83 5.03
[2,] 20.30 20.60 -38.73 6.43
[3,] -161.83 -38.73 190.29 -10.86
[4,] 5.03 6.43 -10.86 2.02
```

```
print("W = "); print(W)
```

```
[1] "W = "
[,1] [,2] [,3] [,4]
[1,] 1785.40 172.5 128.97 289.63
[2,] 172.50 1924.3 178.80 171.90
[3,] 128.97 178.8 2153.00 -1.70
[4,] 289.63 171.9 -1.70 840.20
```

Thus, $|W| \approx 5.669 \times 10^{12}$, $|B + W| \approx 6.830 \times 10^{12}$, and so $\Lambda^* \approx \frac{5.669 \times 10^{12}}{6.830 \times 10^{12}} \approx 0.830$.

```
det(W); det(W+B); det(W)/det(W+B)
```

```
[1] 5.669254e+12
```

```
[1] 6.829577e+12
```

```
[1] 0.8301031
```

As shown in Table 6.3 of Johnson & Wichern, 6th ed., since $p = 4$ and $g = 3$, we have

$$\begin{aligned} & \left(\frac{\sum_{t=1}^3 n_t - p - 2}{p} \right) \left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) \sim F_{2p, 2(\sum_{t=1}^3 n_t - p - 2)} \\ & \Rightarrow \left(\frac{(30 + 30 + 30) - 4 - 2}{4} \right) \left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) \sim F_{2(4), 2(30+30+30-4-2)} \\ & \Rightarrow \frac{21(1 - \sqrt{\Lambda^*})}{\sqrt{\Lambda^*}} \sim F_{8, 168} \end{aligned}$$

Our F statistic is therefore $\frac{21(1 - \sqrt{\Lambda^*})}{\sqrt{\Lambda^*}} \approx \frac{21(1 - \sqrt{0.830})}{\sqrt{0.830}} \approx 2.049$, and the associated p -value is $\mathbb{P}(F_{8,168} > 2.049) \approx 0.044 < 0.05$. Thus, at the $\alpha = 0.05$ level, we reject the null hypothesis

that, on average, there were no differences between male Egyptians' skull measurements across the three time periods in favor of the alternative hypothesis that male Egyptians' skull measurements were significantly different during at least one of the three time periods as compared to the remaining periods.

```
L = det(W)/det(W+B) # Wilks' Lambda
21*(1-sqrt(L))/sqrt(L) # F statistic
```

```
[1] 2.049063
```

```
1-pf(21*(1-sqrt(L))/sqrt(L), 8, 168) # p-value
```

```
[1] 0.04358321
```

1.1.4 Part (iv)

What are the assumptions on the data to make the hypothesis test valid?

For the hypothesis test to be valid, we must assume that $\varepsilon_{tj} \sim \text{iid } N_4(\mathbf{0}, \Sigma) \forall t \in \{1, 2, 3\}, j \in \{1, \dots, 30\}$, where ε_{tj} is defined as in Part (ii) above. That is, within each time period, we assume that each skull's 4-vector of residuals are independently and identically multivariate normally distributed. Moreover, we must also assume that each time periods' skulls have equal population covariance matrices.

1.2 Part (b)

Apply Hotelling's T^2 to determine which pair of time periods differ, treating each pair of time periods as two independent samples of equal covariance structure.

We perform three Hotelling's T^2 tests to evaluate differences in skull measurements between pairs of time periods.

- In our pairwise comparison of Periods 1 and 2, we recover a T^2 statistic of approximately 0.028 and a corresponding p -value of 0.814; there is *not* statistical evidence that the four skull measurements were significantly different among Egyptian males in Period 1 and those in Period 2.
- In our pairwise comparison of Periods 1 and 3, we recover a T^2 statistic of approximately 0.231 and a corresponding p -value of $0.020 < 0.05$. As such, there *is* statistical evidence of significant differences in the four skull measurements between Egyptian males in Period 1 and those in Period 3.

- In our pairwise comparison of Periods 2 and 3, we recover a T^2 statistic of approximately 0.234 and a corresponding p -value of $0.019 < 0.05$. As such, there *is* statistical evidence of significant differences in the four skull measurements between Egyptian males in Period 2 and those in Period 3.

Thus, Egyptian males who lived during Period 3 appear to be the “odd ones out,” having significantly different skull measurements than their ancestors during Periods 1 and 2 (whose measurements were not significantly different from one another).

```
# Pairwise Comparison: Periods 1 and 2
periods12 = filter(skull, period != 3)
manova(cbind(periods12$x1, periods12$x2, periods12$x3, periods12$x4) ~
  period,
       data = periods12) %>%
  summary(test = "Hotelling")
```

	Df	Hotelling-Lawley approx F	num Df	den Df	Pr(>F)
period	1	0.028462	0.39135	4	55 0.8139
Residuals	58				

```
# Pairwise Comparison: Periods 1 and 3
periods13 = filter(skull, period != 2)
manova(cbind(periods13$x1, periods13$x2, periods13$x3, periods13$x4) ~
  period,
       data = periods13) %>%
  summary(test = "Hotelling")
```

	Df	Hotelling-Lawley approx F	num Df	den Df	Pr(>F)						
period	1	0.23087	3.1744	4	55 0.02035 *						
Residuals	58										

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

```
# Pairwise Comparison: Periods 2 and 3
periods23 = filter(skull, period != 1)
manova(cbind(periods23$x1, periods23$x2, periods23$x3, periods23$x4) ~
  period,
       data = periods23) %>%
  summary(test = "Hotelling")
```

	Df	Hotelling-Lawley approx F	num Df	den Df	Pr(>F)
period	1	0.2342	3.2203	4	55 0.01908 *

```

Residuals 58
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

1.3 Part (c)

Assume that you would like to construct confidence intervals for pairwise differences of component means simultaneously for all component pairs and all time periods.

1.3.1 Part (i)

How many simultaneous confidence intervals do you need to construct? (Show how you get the number.)

Since there are three time periods, there are $\binom{3}{2} = \frac{3!}{2!(3-2)!} = 3$ possible pairwise comparisons of time periods. Therefore, since there are four variables in total, we must construct $3 \times 4 = 12$ simultaneous confidence intervals.

1.3.2 Part (ii)

Based on Part (i), there are quite a few (simultaneous) confidence intervals. To save time, in this part you are asked to write out a subset of the simultaneous confidence intervals.

Assume that you construct 85% Bonferroni confidence intervals for pairwise differences of component means simultaneously for all component pairs and all time periods.

Write out the Bonferroni simultaneous confidence intervals for the difference of the component means between samples of Periods 1 and 2, only. Provide the details (including formula for the estimated variance, formula and numerical value of the multiplier).

Since we are interested in constructing 85% Bonferroni confidence intervals, we have $\alpha = 1 - 0.85 = 0.15$. Since we have to construct 12 confidence intervals in all, we have $p = 12$. And since each confidence interval is comparing measurements between two time periods, each of which has 30 skulls, we have $n_1 = n_2 = 30$.

Thus, for variable X_i , $i \in \{1, 2, 3, 4\}$, the confidence interval for the difference in means between Period 1 skulls and Period 2 skulls has the form

$$\begin{aligned}
& \left(\bar{x}_{i1} - \bar{x}_{i2} - t_{n_1+n_2-2, 1-\alpha/2p} \sqrt{s_i^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \bar{x}_{i1} - \bar{x}_{i2} + t_{n_1+n_2-2, 1-\alpha/2p} \sqrt{s_i^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right) \\
&= \left(\bar{x}_{i1} - \bar{x}_{i2} - t_{30+30-2, 1-0.15/2(12)} \sqrt{s_i^2 \left(\frac{1}{30} + \frac{1}{30} \right)}, \bar{x}_{i1} - \bar{x}_{i2} + t_{30+30-2, 1-0.15/2(12)} \sqrt{s_i^2 \left(\frac{1}{30} + \frac{1}{30} \right)} \right) \\
&= \left(\bar{x}_{i1} - \bar{x}_{i2} - t_{58, 0.99375} \sqrt{\frac{s_i^2}{15}}, \bar{x}_{i1} - \bar{x}_{i2} + t_{58, 0.99375} \sqrt{\frac{s_i^2}{15}} \right) \\
&\approx \left(\bar{x}_{i1} - \bar{x}_{i2} - 2.578 \sqrt{\frac{s_i^2}{15}}, \bar{x}_{i1} - \bar{x}_{i2} + 2.578 \sqrt{\frac{s_i^2}{15}} \right)
\end{aligned}$$

where \bar{x}_{it} denotes the mean of measurement i among skulls from Period t , and s_i^2 is the i^{th} diagonal element of the pooled sample covariance matrix $S_{pool} = \frac{(30-1)S_1+(30-1)S_2}{30+30-2} = \frac{29(S_1+S_2)}{58} = \frac{S_1+S_2}{2}$. Here, S_t is the sample covariance matrix for skulls from Period t , $S_t = \frac{1}{30-1} \sum_{j=1}^{30} (\mathbf{x}_{tj} - \bar{\mathbf{x}}_t)(\mathbf{x}_{tj} - \bar{\mathbf{x}}_t)^T$, where \mathbf{x}_{tj} is the 4-vector of skull measurements for skull j from Period t , and $\bar{\mathbf{x}}_t$ is the 4-vector of mean skull measurements for the skulls from Period t .

```
# Period 1 and 2 means for each variable
periods12means = periods12 %>%
  group_by(period) %>%
  summarize(m1 = mean(x1),
            m2 = mean(x2),
            m3 = mean(x3),
            m4 = mean(x4))

# Period 1, Period 2, and Pooled sample covariance matrices
S1 = cov(select(filter(periods12, period == 1), select = -period))
S2 = cov(select(filter(periods12, period == 2), select = -period))
Sp = 0.5*S1 + 0.5*S2
```

In particular, for $X_1 = \text{maximum breadth of skull}$, we have $\bar{x}_{11} \approx 131.367$, $\bar{x}_{12} \approx 132.367$, and $s_1^2 \approx 24.723$, so our confidence interval is approximately $(-4.310, 2.310) \ni 0$.

```
c(periods12means$m1[1] - periods12means$m1[2] - qt(1-0.00625,
  ↵ 58)*sqrt(Sp[1,1]/15),
  ↵ periods12means$m1[1] - periods12means$m1[2] + qt(1-0.00625,
  ↵ 58)*sqrt(Sp[1,1]/15))
```

[1] -4.309679 2.309679

For X_2 = basibregmatic height of skull, we have $\bar{x}_{21} = 133.6$, $\bar{x}_{22} = 132.7$, and $s_2^2 \approx 20.784$, so our confidence interval is approximately $(-2.135, 3.935) \ni 0$.

```
c(periods12means$m2[1] - periods12means$m2[2] - qt(1-0.00625,
  ↵ 58)*sqrt(Sp[2,2]/15),
  periods12means$m2[1] - periods12means$m2[2] + qt(1-0.00625,
  ↵ 58)*sqrt(Sp[2,2]/15))
```

[1] -2.134624 3.934624

For X_3 = basialveolar length of skull, we have $\bar{x}_{31} \approx 99.167$, $\bar{x}_{32} \approx 99.067$, and $s_3^2 \approx 26.759$, so our confidence interval is approximately $(-3.343, 3.543) \ni 0$.

```
c(periods12means$m3[1] - periods12means$m3[2] - qt(1-0.00625,
  ↵ 58)*sqrt(Sp[3,3]/15),
  periods12means$m3[1] - periods12means$m3[2] + qt(1-0.00625,
  ↵ 58)*sqrt(Sp[3,3]/15))
```

[1] -3.343276 3.543276

For X_4 = nasal height of skull, we have $\bar{x}_{41} \approx 50.533$, $\bar{x}_{42} \approx 50.233$, and $s_4^2 \approx 8.187$, so our confidence interval is approximately $(-1.605, 2.205) \ni 0$.

```
c(periods12means$m4[1] - periods12means$m4[2] - qt(1-0.00625,
  ↵ 58)*sqrt(Sp[4,4]/15),
  periods12means$m4[1] - periods12means$m4[2] + qt(1-0.00625,
  ↵ 58)*sqrt(Sp[4,4]/15))
```

[1] -1.604548 2.204548

1.3.3 Part (iii)

Determine which mean components differ statistically significantly at 85% confidence level between Periods 1 and 2.

Since all four confidence intervals in Part (ii) contained 0, we conclude that none of the four mean skull components differ statistically significantly at the 85% confidence level between Periods 1 and 2. This result coheres with our finding in Part (b) that the mean vectors of Periods 1 and 2 were not statistically significantly different from one another at the $\alpha = 0.05$ level.

2 Exercise 3: Multidimensional Scaling

The table in T12-13.DAT gives the “distances” between certain archaeological sites from different periods, based upon the frequencies of different types of potsherds found at the sites. The dates of the periods 1, 2, …, 9 correspond to A.D. years 918, 1131, 960, 987, 1024, 1005, 945, 1137, and 1062, respectively.

```
potsherds = matrix(0, 9, 9)
potsherds[row(potsherds) <= col(potsherds)] =
  scan("C:/Users/rewin/OneDrive/Documents/STAT_32950/Homework/HW4/T12-13.DAT")
potsherds = t(potsherds)
```

2.1 Part (a)

Given these distances, determine the coordinates of the sites in $q = 3, 4, 5$ dimensions using (classical metric) multidimensional scaling.

The coordinates of the nine sites in $q = 3$ dimensions are as follows:

```
# q = 3
cmdscale(as.dist(potsherds), k = 3) %>% round(3)
```

```
[,1]   [,2]   [,3]
[1,]  0.512 -0.278  0.242
[2,] -1.318  0.692  0.623
[3,]  0.470 -0.071  0.186
[4,]  0.387  0.088  0.049
[5,]  0.234  0.296 -0.325
[6,]  0.469  0.137 -0.219
[7,]  0.581 -0.349  0.457
[8,] -1.118 -1.122 -0.316
[9,] -0.216  0.608 -0.697
```

The coordinates of the nine sites in $q = 4$ dimensions are as follows:

```
# q = 4
cmdscale(as.dist(potsherds), k = 4) %>% round(3)
```

```
[,1]   [,2]   [,3]   [,4]
[1,]  0.512 -0.278  0.242  0.676
[2,] -1.318  0.692  0.623  0.050
[3,]  0.470 -0.071  0.186 -0.302
```

```
[4,]  0.387  0.088  0.049 -0.344
[5,]  0.234  0.296 -0.325 -0.052
[6,]  0.469  0.137 -0.219  0.139
[7,]  0.581 -0.349  0.457 -0.178
[8,] -1.118 -1.122 -0.316 -0.052
[9,] -0.216  0.608 -0.697  0.062
```

The coordinates of the nine sites in $q = 5$ dimensions are as follows:

```
# q = 5
cmdscale(as.dist(potsherds), k = 5) %>% round(3)
```

```
[,1]   [,2]   [,3]   [,4]   [,5]
[1,]  0.512 -0.278  0.242  0.676  0.119
[2,] -1.318  0.692  0.623  0.050 -0.024
[3,]  0.470 -0.071  0.186 -0.302  0.059
[4,]  0.387  0.088  0.049 -0.344  0.102
[5,]  0.234  0.296 -0.325 -0.052  0.121
[6,]  0.469  0.137 -0.219  0.139 -0.281
[7,]  0.581 -0.349  0.457 -0.178 -0.102
[8,] -1.118 -1.122 -0.316 -0.052 -0.005
[9,] -0.216  0.608 -0.697  0.062  0.010
```

2.2 Part (b)

The $\text{Stress}(q)$ can be calculated as

$$\text{Stress}(q) = \left[\frac{\sum_{j < i} (x_{ij} - d_{ij}^{(q)})^2}{\sum_{j < i} x_{ij}^2} \right]^{1/2}$$

where $d_{ij}^{(q)}$ is the distance between sites i and j in q -dimensional representation by using (classical metric) multidimensional scaling method, and x_{ij} is the corresponding distance matrix X given by the data. Plot (the minimum) $\text{Stress}(q)$ versus q and interpret the graph.

Below, we plot $\text{Stress}(q)$ against q for $q \in \{1, \dots, 7\}$ (the $q = N - 1 = 8$ configuration is identical to the $q = 7$ configuration, so is not included in the plot). For these data, we see that as the dimension of our map increases, the stress associated with the map generally decreases. (It is difficult to see in the plot, but the stress actually increases by a very small amount from $q = 6$ to $q = 7$.) We also see that once $q = 5$ dimensions are used, the q -dimensional map almost perfectly captures the true distances between the sites, such that moving into 6- or 7-dimensional representations of the sites barely improves the fit. Using the rule of

thumb that a value of *Stress* less than 0.05 is “good”, even a 4-dimensional representation of the data is good enough, as the associated *Stress* is approximately 0.029.

```

stress = rep(0, 7)
for(i in 1:7){
  fit = cmdscale(as.dist(potsherds), k = i)
  diffS2 = sum((as.dist(potsherds) - dist(fit))^2)
  dist2 = sum((as.dist(potsherds))^2)
  stress[i] = sqrt(diffS2/dist2)
}

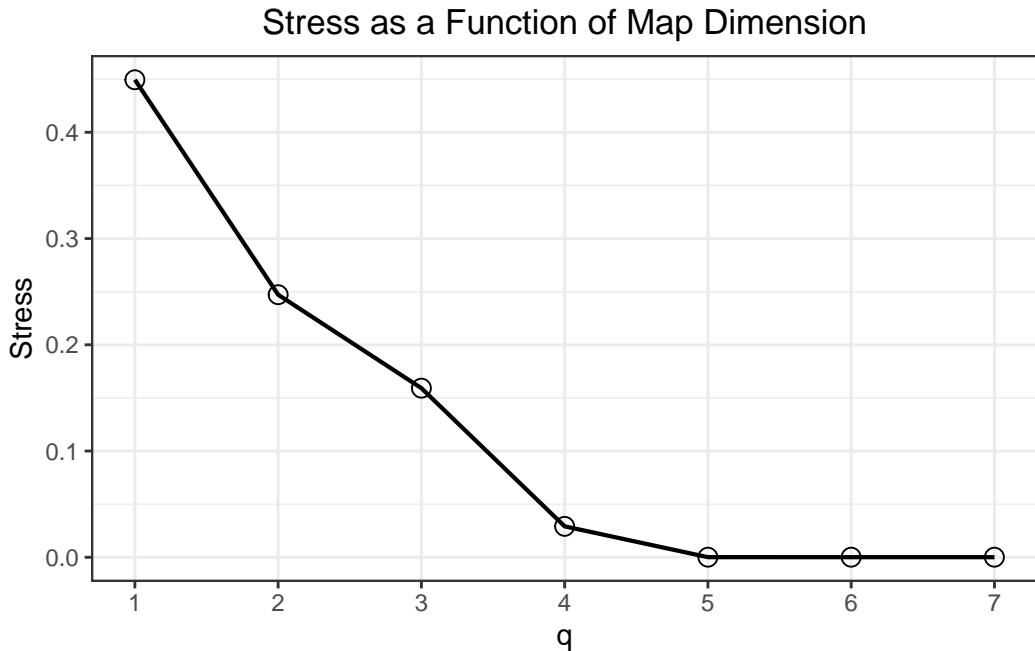
stress = stress %>% as.data.frame()
stress = rename(stress, Stress = .) %>%
  mutate(q = row_number())

```

```

ggplot(data = stress, aes(x=q, y=Stress)) +
  theme_bw() +
  geom_line(linewidth = 0.75) +
  geom_point(shape = 1, size = 3) +
  scale_x_continuous(breaks = scales::pretty_breaks(n=7),
                     minor_breaks = NULL) +
  ggtitle("Stress as a Function of Map Dimension") +
  theme(plot.title = element_text(hjust = 0.5))

```



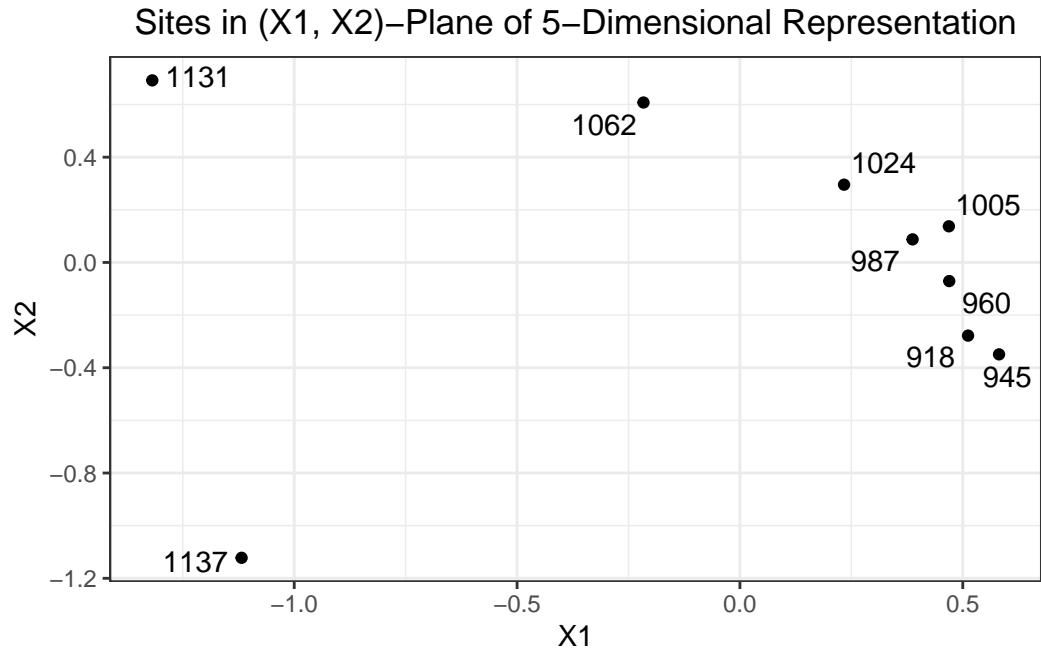
2.3 Part (c)

Plot the nine sites (treated as variables) in two dimensions using the first two coordinates for $q = 5$ dimensional solutions. Note the periods associated with the sites. Archaeologists would say that the two dimensional configuration contains a time trend or movement pattern. Do you see it (admittedly the pattern is very vague)?

Below, we plot the nine sites using the first two coordinates of their 5-dimensional representations. In general, it appears that more recent sites have progressively smaller X_1 coordinates and progressively larger X_2 coordinates, though this is not a perfect rule (for example, the most recent site, dated 1137 A.D., actually has the smallest X_2 coordinate). The imperfect quality of this pattern may be because there is additional information captured in the third through fifth coordinates of each site that are not visualized.

```
q3dim5 = cmdscale(as.dist(potsherds), k = 5) %>%
  as.data.frame() %>%
  mutate(year = c(918, 1131, 960, 987, 1024, 1005, 945, 1137, 1062))

ggplot(q3dim5, aes(x=V1, y=V2, label = year)) +
  theme_bw() +
  geom_point() +
  geom_text_repel() +
  xlab("X1") +
  ylab("X2") +
  ggtitle("Sites in (X1, X2)-Plane of 5-Dimensional Representation") +
  theme(plot.title = element_text(hjust = 0.5))
```



3 Exercise 4: Correspondence Analysis

A sample of 592 students is cross-classified according to hair colors and eye colors in the table in HairEyeAll.txt.

```
haireyeall =
  ↵  read.table("C:/Users/rewin/OneDrive/Documents/STAT_32950/Homework/HW4/HairEyeAll.txt")
rownames(haireyeall) = c("Black", "Brown", "Red", "Blond") # for variable
  ↵  Hair
colnames(haireyeall) = c("Brown", "Blue", "Hazel", "Green") # for variable
  ↵  Eye
```

3.1 Part (a)

Obtain the tables of cell percentages ($p_{ij} = x_{ij}/n$), row percentages, and column percentages.

First, we display a table of cell percentages:

```
round(as.matrix(haireyeall)/sum(as.matrix(haireyeall)), 3)
```

	Brown	Blue	Hazel	Green
Black	0.115	0.034	0.025	0.008
Brown	0.201	0.142	0.091	0.049
Red	0.044	0.029	0.024	0.024
Blond	0.012	0.159	0.017	0.027

Next, we display a table of within-row percentages (i.e., a row profile matrix):

```
rowtotals = as.matrix(haireyeall) %*% c(1,1,1,1)
rowprofile = round(diag(c(1/rowtotals)) %*% as.matrix(haireyeall), 3)
rownames(rowprofile) = c("Black", "Brown", "Red", "Blond")
rowprofile %>%
  as.data.frame() %>%
  mutate(Total = Brown + Blue + Hazel + Green) %>%
  as.matrix()
```

	Brown	Blue	Hazel	Green	Total
Black	0.630	0.185	0.139	0.046	1.000
Brown	0.416	0.294	0.189	0.101	1.000
Red	0.366	0.239	0.197	0.197	0.999
Blond	0.055	0.740	0.079	0.126	1.000

Finally, we display a table of within-column percentages (i.e., a column profile matrix):

```
coltotals = c(1,1,1,1) %*% as.matrix(haireyeall)
colprofile = round(as.matrix(haireyeall) %*% diag(c(1/coltotals)), 3)
colnames(colprofile) = c("Brown", "Blue", "Hazel", "Green")
colprofile %>% t() %>%
  as.data.frame() %>%
  mutate(Total = Black + Brown + Red + Blond) %>%
  as.matrix() %>% t()
```

	Brown	Blue	Hazel	Green
Black	0.309	0.093	0.161	0.078
Brown	0.541	0.391	0.581	0.453
Red	0.118	0.079	0.151	0.219
Blond	0.032	0.437	0.108	0.250
Total	1.000	1.000	1.001	1.000

3.2 Part (b)

Obtain the table of expected cell counts (i.e., expected cell frequencies) E_{ij} if the variables *Hair* and *Eye* were independent (for the given data).

If *Hair* color and *Eye* color were independent, we would expect to observe the following counts:

```
E = (as.matrix(haireyeall) %*% c(1,1,1,1)) %*%
  ~ (c(1,1,1,1) %*% as.matrix(haireyeall)) /
  sum(haireyeall)
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

3.3 Part (c)

Obtain the table of cell mass $= \frac{(x_{ij} - E_{ij})^2}{E_{ij}}$. Conduct a Pearson's chi-square test (What is the hypothesis? What is the degrees of freedom of the χ^2 ?) and check that the χ^2 -statistic/ n = total inertia.

First, we present the table of cell masses:

```
((as.matrix(haireyeall) - E)^2/E) %>% round(3)
```

	Brown	Blue	Hazel	Green
Black	19.346	9.421	0.228	3.817
Brown	1.521	3.800	1.831	0.119
Red	0.006	2.993	0.726	5.211
Blond	34.234	49.697	4.963	0.375

We now test the null hypothesis H_0 : hair color is independent of eye color using a Pearson's chi-square test. The χ^2 statistic is the sum of the entries in the mass table above: $X^2 \approx 138.29$. Moreover, under H_0 , the chi-square statistic $X^2 \sim_{H_0} \chi^2_{(4-1)(4-1)} = \chi^2_9$. As such, we recover a p -value of $p = \mathbb{P}(X^2 > 138.29) < 2.2 \times 10^{-16}$, and we reject the null hypothesis that hair color and eye color are independent in favor of the alternative hypothesis that there is a relationship between an individual's hair color and their eye color.

```
chisq.test(haireyeall)
```

Pearson's Chi-squared test

```
data: haireyeall  
X-squared = 138.29, df = 9, p-value < 2.2e-16
```

The total inertia is $\frac{X^2}{n} \approx \frac{138.29}{592} \approx 0.234$. To confirm our calculation, we can use the `ca` function in R, which returns “principal inertias” that total approximately $0.209 + 0.022 + 0.003 \approx 0.234$ (as shown in Part (d) below).

3.4 Part (d)

Perform a correspondence analysis of the data. What percentage of variation in the data is captured in the 2-dimensional CA plot? Are there any associations of Eye category and Hair category reflected in the plot?

We first summarize the results of a correspondence analysis below. Notice that the first two dimensions of the output cumulatively capture 98.9% of the variation in the data.

```
ca(haireyeall) %>% summary()
```

Principal inertias (eigenvalues):

dim	value	%	cum%	scree	plot
1	0.208773	89.4	89.4	*****	*****
2	0.022227	9.5	98.9	**	
3	0.002598	1.1	100.0		
----- -----					
Total: 0.233598 100.0					

Rows:

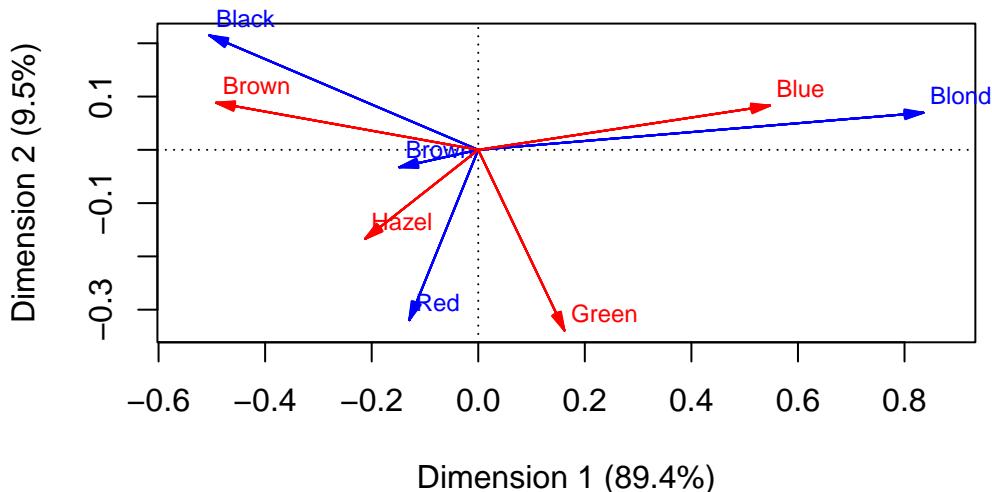
	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	Blck	182	990	237	-505	838	222	215	152	379
2	Brwn	483	906	53	-148	864	51	-33	42	23
3	Red	120	945	65	-130	133	10	-320	812	551
4	Blnd	215	1000	646	835	993	717	70	7	47

Columns:

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	Brwn	372	998	398	-492	967	431	88	31	130
2	Blue	363	1000	477	547	977	521	83	22	112
3	Hazl	157	879	56	-213	542	34	-167	336	198
4	Gren	108	948	69	162	176	14	-339	773	559

Now, we plot the coordinates of each hair and eye color using the standard symmetric scaling of the corresponding coordinate systems. As implied by the table above, this plot captures 98.9% of the variation in the data. In the upper-right corner of the plot, we see that blond hair and blue eyes are both fairly far from the origin and have a small angle between them; this suggests a strong association between having blond hair and having blue eyes. Similarly, in the upper-left corner of the plot, we see that black hair and brown eyes are both fairly far from the origin and have a fairly small angle between them; this suggests a strong association between having black hair and having brown eyes.

```
plot(ca(haireyeall), arrows = c(T,T))
```



5. $X_1 \in \mathbb{R}^{d_1}, X_2 \in \mathbb{R}^{d_2}, X_3 \in \mathbb{R}^{d_3}$ ($d_i \geq 2$) are jointly MVN, w/r

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & 0 \\ \Sigma_{31} & 0 & \Sigma_{33} \end{bmatrix},$$

where Σ_{ii} are positive definite.

Sol(i) Derive $\mathbb{E}[X_1 | X_2 = x_2, X_3 = x_3]$.

We partition $\mu + \Sigma$ as follows:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \leftarrow \mu_A \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & 0 \\ \Sigma_{31} & 0 & \Sigma_{33} \end{bmatrix}$$

$\downarrow \Sigma_A \quad \downarrow \Sigma_B$
 $\uparrow \Sigma_C \quad \uparrow \Sigma_D$

Then

$$\begin{aligned} \mathbb{E}[X_1 | X_2 = x_2, X_3 = x_3] &= \mu_A + \Sigma_B \Sigma_D^{-1} \left(\begin{bmatrix} x_2 \\ x_3 \end{bmatrix} - \mu_B \right) \\ &= \mu_1 + [\Sigma_{12} \ \Sigma_{13}] \begin{bmatrix} \Sigma_{22} & 0 \\ 0 & \Sigma_{33} \end{bmatrix}^{-1} \left(\begin{bmatrix} x_2 \\ x_3 \end{bmatrix} - \begin{bmatrix} \mu_2 \\ \mu_3 \end{bmatrix} \right) \end{aligned}$$

Since Σ_{22}, Σ_{33} are positive definite, they are invertible:

$\Sigma_{22}^{-1}, \Sigma_{33}^{-1}$ exist. So,

$$\begin{aligned} \begin{bmatrix} \Sigma_{22} & 0 \\ 0 & \Sigma_{33} \end{bmatrix}^{-1} &= \begin{bmatrix} \Sigma_{22}^{-1} + \Sigma_{22}^{-1} O (\Sigma_{33} - O \Sigma_{22}^{-1} O)^{-1} O \Sigma_{22}^{-1} & -\Sigma_{22}^{-1} O (\Sigma_{33} - O \Sigma_{22}^{-1} O)^{-1} \\ -(\Sigma_{33} - O \Sigma_{22}^{-1} O)^{-1} O \Sigma_{22}^{-1} & (\Sigma_{33} - O \Sigma_{22}^{-1} O)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{22}^{-1} + \Sigma_{22}^{-1} O (\Sigma_{33})^{-1} O \Sigma_{22}^{-1} & -\Sigma_{22}^{-1} O (\Sigma_{33})^{-1} \\ -(\Sigma_{33})^{-1} O \Sigma_{22}^{-1} & (\Sigma_{33})^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{22}^{-1} & 0 \\ 0 & \Sigma_{33}^{-1} \end{bmatrix} \end{aligned}$$

So,

$$\mathbb{E}[X_1 | X_2 = x_2, X_3 = x_3] = \mu_1 + [\Sigma_{12} \ \Sigma_{13}] \begin{bmatrix} \Sigma_{22}^{-1} & 0 \\ 0 & \Sigma_{33}^{-1} \end{bmatrix} \begin{bmatrix} x_2 - \mu_2 \\ x_3 - \mu_3 \end{bmatrix}$$

$$= \mu_1 + [\Sigma_{12} \quad \Sigma_{13}] \begin{bmatrix} \Sigma_{22}^{-1} (x_2 - \mu_2) \\ \Sigma_{33}^{-1} (x_3 - \mu_3) \end{bmatrix}$$

$$\Rightarrow E[X_1 | X_2 = x_2, X_3 = x_3] = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2) + \Sigma_{13} \Sigma_{33}^{-1} (x_3 - \mu_3)$$

5a(i) Derive $\text{Var}(X_1 | X_2 = x_2, X_3 = x_3)$.

$$\begin{aligned} \text{Var}(X_1 | X_2 = x_2, X_3 = x_3) &= \bar{\Sigma}_A - \bar{\Sigma}_B \bar{\Sigma}_D^{-1} \bar{\Sigma}_C \\ &= \bar{\Sigma}_{11} - [\Sigma_{12} \quad \Sigma_{13}] \begin{bmatrix} \Sigma_{22} & 0 \\ 0 & \Sigma_{33} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_{21} \\ \Sigma_{31} \end{bmatrix} \\ &= \bar{\Sigma}_{11} - [\Sigma_{12} \quad \Sigma_{13}] \begin{bmatrix} \Sigma_{22}^{-1} & 0 \\ 0 & \Sigma_{33}^{-1} \end{bmatrix} \begin{bmatrix} \Sigma_{21} \\ \Sigma_{31} \end{bmatrix} \\ &= \bar{\Sigma}_{11} - [\Sigma_{12} \quad \Sigma_{13}] \begin{bmatrix} \Sigma_{22}^{-1} \Sigma_{21} + 0 \Sigma_{31} \\ 0 \Sigma_{21} + \Sigma_{33}^{-1} \Sigma_{31} \end{bmatrix} \\ &= \bar{\Sigma}_{11} - [\Sigma_{12} \quad \Sigma_{13}] \begin{bmatrix} \Sigma_{22}^{-1} \Sigma_{21} \\ \Sigma_{33}^{-1} \Sigma_{31} \end{bmatrix} \end{aligned}$$

$$\Rightarrow \text{Var}(X_1 | X_2 = x_2, X_3 = x_3) = \bar{\Sigma}_{11} - \bar{\Sigma}_{12} \bar{\Sigma}_{22}^{-1} \bar{\Sigma}_{21} - \bar{\Sigma}_{13} \bar{\Sigma}_{33}^{-1} \bar{\Sigma}_{31}$$

5b Derive conditional distribution of X_1 given $X_2 + X_3 = x_0$.

Start by finding distribution of $X_2 + X_3$.

- Let $Y = X_2 + X_3$.

- Write $X_1 = (X_{11}, \dots, X_{1d_1})$, $X_2 = (X_{21}, \dots, X_{2d_2})$, $X_3 = (X_{31}, \dots, X_{3d_3})$ (since $d_3 = d_2$ necessarily).

- Since (X_1, X_2, X_3) are jointly MVN, any linear combination $a_1 X_{11} + \dots + a_{d_1} X_{1d_1} + b_1 X_{21} + \dots + b_{d_2} X_{2d_2} + c_1 X_{31} + \dots + c_{d_2} X_{3d_2}$ has normal distribution $\forall a_i, b_i, c_i \in \mathbb{R}$. In particular, if $a_i = 0 \forall i = 1, \dots, d_1$, then $b_1 X_{21} + \dots + b_{d_2} X_{2d_2} + c_1 X_{31} + \dots + c_{d_2} X_{3d_2}$ has normal distribution.

$$\begin{aligned} Y = X_2 + X_3 &= (X_{21}, \dots, X_{2d_2}) + (X_{31}, \dots, X_{3d_2}) \\ &= (X_{21} + X_{31}, \dots, X_{2d_2} + X_{3d_2}) \end{aligned}$$

So, linear combinations of the components of γ have the form

$$e_1(X_{21} + X_{31}) + \dots + e_{d_2}(X_{2d_2} + X_{3d_2}) \\ = e_1 X_{21} + \dots + e_{d_2} X_{2d_2} + e_1 X_{31} + \dots + e_{d_2} X_{3d_2}$$

for some $e_1, \dots, e_{d_2} \in \mathbb{R}$. But we've already shown that all linear combinations of $X_{21}, \dots, X_{2d_2}, X_{31}, \dots, X_{3d_2}$ have normal distribution, which means that all linear combinations of the components of γ have normal distribution. Thus, γ has multivariate normal distribution.

In particular,

$$\begin{aligned} \mathbb{E}[\gamma] &= \mathbb{E}[X_2 + X_3] \\ &= \mathbb{E}[X_2] + \mathbb{E}[X_3] \\ &= \mu_2 + \mu_3 \end{aligned}$$

And, $\text{Var}(\gamma) = \text{Var}(X_2 + X_3)$

$$\begin{aligned} &= \mathbb{E}[(X_2 + X_3)(X_2 + X_3)^T] - \mathbb{E}[X_2 + X_3]\mathbb{E}[X_2 + X_3]^T \\ &= \mathbb{E}[(X_2 + X_3)(X_2^T + X_3^T)] - (\mu_2 + \mu_3)(\mu_2 + \mu_3)^T \\ &= \mathbb{E}[X_2 X_2^T + X_2 X_3^T + X_3 X_2^T + X_3 X_3^T] - (\mu_2 + \mu_3)(\mu_2^T + \mu_3^T) \\ &= \mathbb{E}[X_2 X_2^T] + \mathbb{E}[X_2 X_3^T] + \mathbb{E}[X_3 X_2^T] + \mathbb{E}[X_3 X_3^T] \\ &\quad - \mu_2 \mu_2^T - \mu_2 \mu_3^T - \mu_3 \mu_2^T - \mu_3 \mu_3^T \\ &= (\mathbb{E}[X_2 X_2^T] - \mu_2 \mu_2^T) + (\mathbb{E}[X_2 X_3^T] - \mu_2 \mu_3^T) \\ &\quad + (\mathbb{E}[X_3 X_2^T] - \mu_3 \mu_2^T) + (\mathbb{E}[X_3 X_3^T] - \mu_3 \mu_3^T) \\ &= \text{Var}(X_2) + \text{Cov}(X_2, X_3) + \text{Cov}(X_3, X_2) + \text{Var}(X_3) \\ &= \Sigma_{22} + \Sigma_{23} + \Sigma_{32} + \Sigma_{33} \\ &= \Sigma_{22} + 0 + 0 + \Sigma_{33} \\ &= \Sigma_{22} + \Sigma_{33} \end{aligned}$$

Thus,

$$\gamma = X_2 + X_3 \sim N_{d_2}(\mu_2 + \mu_3, \Sigma_{22} + \Sigma_{33})$$

- Note that linear combinations of the components of $X_1 + Y$ have the form

$$a_1 X_{11} + \dots + a_{d_1} X_{1d_1} + e_1 (X_{21} + X_{31}) + \dots + e_{d_2} (X_{2d_2} + X_{3d_2})$$

$$= a_1 X_{11} + \dots + a_{d_1} X_{1d_1} + e_1 X_{21} + \dots + e_{d_2} X_{2d_2} + e_1 X_{31} + \dots + e_{d_2} X_{3d_2}$$

for some a_i 's, e_j 's $\in \mathbb{R}$. But we've already shown that all linear combinations of $X_{11}, \dots, X_{1d_1}, X_{21}, \dots, X_{2d_2}, X_{31}, \dots, X_{3d_2}$ have normal distribution, which means that all linear combinations of the components of $X_1 + Y$ have normal distribution. Thus, (X_1, Y) are jointly multivariate normal.

- $\mathbb{E} \begin{bmatrix} X_1 \\ Y \end{bmatrix} = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[Y] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 + \mu_3 \end{bmatrix}$

- $\text{Var} \begin{pmatrix} X_1 \\ Y \end{pmatrix} = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, Y) \\ \text{Cov}(Y, X_1) & \text{Var}(Y) \end{bmatrix}$
 $= \begin{bmatrix} \Sigma_{11} & \text{Cov}(X_1, Y) \\ \text{Cov}(Y, X_1) & \Sigma_{22} + \Sigma_{33} \end{bmatrix}$

- $\text{Cov}(X_1, Y) = \mathbb{E}[X_1 Y^T] - \mathbb{E}[X_1] \mathbb{E}[Y]^T$
 $= \mathbb{E}[X_1 (X_2 + X_3)^T] - \mu_1 \mathbb{E}[X_2 + X_3]^T$
 $= \mathbb{E}[X_1 (X_2^T + X_3^T)] - \mu_1 (\mu_2 + \mu_3)^T$
 $= \mathbb{E}[X_1 X_2^T + X_1 X_3^T] - \mu_1 (\mu_2^T + \mu_3^T)$
 $= (\mathbb{E}[X_1 X_2^T] - \mu_1 \mu_2^T) + (\mathbb{E}[X_1 X_3^T] - \mu_1 \mu_3^T)$
 $= \text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3)$
 $= \Sigma_{12} + \Sigma_{13}$

- $\text{Cov}(Y, X_1) = \text{Cov}(X_1, Y)^T$
 $= (\Sigma_{12} + \Sigma_{13})^T$
 $= \Sigma_{12}^T + \Sigma_{13}^T$
 $= \Sigma_{21} + \Sigma_{31}$

- So $\text{Var} \begin{pmatrix} X_1 \\ Y \end{pmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} + \Sigma_{13} \\ \Sigma_{21} + \Sigma_{31} & \Sigma_{22} + \Sigma_{33} \end{bmatrix}$

That is,

$$\begin{bmatrix} X_1 \\ Y \end{bmatrix} \sim N_{d_1+d_2} \left(\begin{bmatrix} \mu_u \\ \mu_v \end{bmatrix}, \begin{bmatrix} \Sigma_Q & \Sigma_R \\ \Sigma_S & \Sigma_T \end{bmatrix} \right)$$

$\downarrow \mu_u$ $\downarrow \Sigma_Q$ $\downarrow \Sigma_R$
 $\downarrow \mu_v$ $\downarrow \Sigma_S$ $\downarrow \Sigma_T$

Thus,

$$X_1 | Y = x_0 \sim N_{d_1} (\mu_u + \Sigma_R \Sigma_T^{-1} (x_0 - \mu_v), \Sigma_Q - \Sigma_R \Sigma_T^{-1} \Sigma_S)$$

$$\Rightarrow X_1 | X_2 + X_3 = x_0 \sim N_{d_1} \left(\mu_u + (\Sigma_{12} + \Sigma_{13})(\Sigma_{22} + \Sigma_{33})^{-1} (x_0 - \mu_2 - \mu_3), \right.$$

$$\left. \Sigma_{11} - (\Sigma_{12} + \Sigma_{13})(\Sigma_{22} + \Sigma_{33})^{-1} (\Sigma_{21} + \Sigma_{31}) \right)$$