STAT 32950 Assignment 4

Bin Yu

Apr 21, 2025

Question 1

(a) Decomposition of the two variables

First compute

$$\bar{x} = \frac{1}{12}\sum_{t=1}^{3}\sum_{j} x_{tj} = (4,\ 5)',$$

and,

$$\bar{x}_{1\cdot} = (6,8)',\ \bar{x}_{2\cdot} = (2,4)',\ \bar{x}_{3\cdot} = (3,2)'.$$

Hence the treatment effects are

$$\bar{x}_{1\cdot} - \bar{x} = (2,3)', \quad \bar{x}_{2\cdot} - \bar{x} = (-2,-1)', \quad \bar{x}_{3\cdot} - \bar{x} = (-1,-3)'.$$

For Variable 1 (x_1):

$$\begin{bmatrix} 6 & 5 & 8 & 4 & 7 \\ 3 & 1 & 2 & & \\ 2 & 5 & 3 & 2 & \end{bmatrix} = \underbrace{\begin{bmatrix} 4 & 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & & \\ 4 & 4 & 4 & 4 & \end{bmatrix}}_{\bar{x}} + \underbrace{\begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ -2 & -2 & -2 & & \\ -1 & -1 & -1 & -1 & \end{bmatrix}}_{\bar{x}_{t\cdot}-\bar{x}} + \underbrace{\begin{bmatrix} 0 & -1 & 2 & -2 & 1 \\ 1 & -1 & 0 & & \\ -1 & 2 & 0 & -1 & \end{bmatrix}}_{x_{tj}-\bar{x}_{t\cdot}}.$$

For Variable 2 (x_2):

$$\begin{bmatrix} 7 & 9 & 6 & 9 & 9 \\ 3 & 6 & 3 & & \\ 3 & 1 & 1 & 3 & \end{bmatrix} = \underbrace{\begin{bmatrix} 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & & \\ 5 & 5 & 5 & 5 & \end{bmatrix}}_{\bar{x}} + \underbrace{\begin{bmatrix} 3 & 3 & 3 & 3 & 3 \\ -1 & -1 & -1 & & \\ -3 & -3 & -3 & -3 & \end{bmatrix}}_{\bar{x}_{t\cdot}-\bar{x}} + \underbrace{\begin{bmatrix} -1 & 1 & -2 & 1 & 1 \\ -1 & 2 & -1 & & \\ 1 & -1 & -1 & 1 & \end{bmatrix}}_{x_{tj}-\bar{x}_{t\cdot}}.$$

(b) One-way MANOVA table

First compute the overall and treatment means:

$$\bar{x} = \frac{1}{12}\sum_{t,j} x_{tj} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}, \quad \bar{x}_{1\cdot} = \begin{pmatrix} 6 \\ 8 \end{pmatrix},\ \bar{x}_{2\cdot} = \begin{pmatrix} 2 \\ 4 \end{pmatrix},\ \bar{x}_{3\cdot} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

Then the between-group and within-group matrix:

$$
\begin{aligned}
B &= \sum_{t=1}^{3} n_t \left(\bar{x}_{t\cdot} - \bar{x}\right)\left(\bar{x}_{t\cdot} - \bar{x}\right)' \\
&= 5 \begin{pmatrix} 2 \\ 3 \end{pmatrix} \begin{pmatrix} 2 & 3 \end{pmatrix} + 3 \begin{pmatrix} -2 \\ -1 \end{pmatrix} \begin{pmatrix} -2 & -1 \end{pmatrix} + 4 \begin{pmatrix} -1 \\ -3 \end{pmatrix} \begin{pmatrix} -1 & -3 \end{pmatrix} \\
&= 5 \begin{pmatrix} 4 & 6 \\ 6 & 9 \end{pmatrix} + 3 \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} + 4 \begin{pmatrix} 1 & 3 \\ 3 & 9 \end{pmatrix} \\
&= \begin{pmatrix} 20 & 30 \\ 30 & 45 \end{pmatrix} + \begin{pmatrix} 12 & 6 \\ 6 & 3 \end{pmatrix} + \begin{pmatrix} 4 & 12 \\ 12 & 36 \end{pmatrix} \\
&= \begin{pmatrix} 36 & 48 \\ 48 & 84 \end{pmatrix}.
\end{aligned}
$$

Thus,

$$
B = \sum_{t=1}^{3} n_t(\bar{x}_{t\cdot} - \bar{x})(\bar{x}_{t\cdot} - \bar{x})' = \begin{pmatrix} 36 & 48 \\ 48 & 84 \end{pmatrix},
$$

For the matrix W:

$$
W = \sum_{t=1}^{3} \sum_{j} (x_{tj} - \bar{x}_{t\cdot})(x_{tj} - \bar{x}_{t\cdot})' = W_1 + W_2 + W_3,
$$

$$
W_1 = \sum_{j=1}^{5} \begin{pmatrix} x_{1j} - 6 \\ y_{1j} - 8 \end{pmatrix} \begin{pmatrix} x_{1j} - 6 & y_{1j} - 8 \end{pmatrix} = \begin{pmatrix} 10 & -6 \\ -6 & 8 \end{pmatrix},
$$

$$
W_2 = \sum_{j=1}^{3} \begin{pmatrix} x_{2j} - 2 \\ y_{2j} - 4 \end{pmatrix} \begin{pmatrix} x_{2j} - 2 & y_{2j} - 4 \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ -3 & 6 \end{pmatrix},
$$

$$
W_3 = \sum_{j=1}^{4} \begin{pmatrix} x_{3j} - 3 \\ y_{3j} - 2 \end{pmatrix} \begin{pmatrix} x_{3j} - 3 & y_{3j} - 2 \end{pmatrix} = \begin{pmatrix} 6 & -4 \\ -4 & 4 \end{pmatrix},
$$

$$
W = W_1 + W_2 + W_3 = \begin{pmatrix} 10 & -6 \\ -6 & 8 \end{pmatrix} + \begin{pmatrix} 2 & -3 \\ -3 & 6 \end{pmatrix} + \begin{pmatrix} 6 & -4 \\ -4 & 4 \end{pmatrix} = \begin{pmatrix} 18 & -13 \\ -13 & 18 \end{pmatrix}.
$$

Thus

$$
W = \sum_{t=1}^{3} \sum_{j} (x_{tj} - \bar{x}_{t\cdot})(x_{tj} - \bar{x}_{t\cdot})' = \begin{pmatrix} 18 & -13 \\ -13 & 18 \end{pmatrix}.
$$

The total is $B + W$:

$$
T = \begin{pmatrix} 54 & 35 \\ 35 & 102 \end{pmatrix}.
$$

Source	Degrees of Freedom	Martix of Sum of Squares
Treatments	$g-1=2$	$B=\begin{pmatrix}36 & 48\\ 48 & 84\end{pmatrix}$
Error	$N-g=9$	$W=\begin{pmatrix}18 & -13\\ -13 & 18\end{pmatrix}$
Total	$N-1=11$	$T=\begin{pmatrix}54 & 35\\ 35 & 102\end{pmatrix}$

(c)

From

$$W=\begin{pmatrix}18 & -13\\ -13 & 18\end{pmatrix},\quad B=\begin{pmatrix}36 & 48\\ 48 & 84\end{pmatrix},\quad T=B+W=\begin{pmatrix}54 & 35\\ 35 & 102\end{pmatrix},$$

we compute

$$|W|=18\cdot 18-(-13)^2=324-169=155,$$
$$|B|=36\cdot 84-48^2=3024-2304=720,$$
$$|T|=54\cdot 102-35^2=5508-1225=4283.$$

Thus Wilks' lambda is

$$\Lambda^*=\frac{|W|}{|T|}=\frac{155}{4283}\approx 0.0362.$$

(d)

Under H_0, Bartlett's correction gives

$$\chi^2_{\text{obs}} \;=\; -\ln(\Lambda^*)\,\big(N-1-\tfrac{p+g}{2}\big),$$

with $N=12$, $p=2$, $g=3$, so

$$N-1-\tfrac{p+g}{2}=11-\tfrac{2+3}{2}=11-2.5=8.5.$$

Hence

$$\chi^2_{\text{obs}}=-\,8.5\,\ln(0.0362)\approx 28.21,$$

with $df=p(g-1)=2\cdot 2=4$. The corresponding p-value is

$$p=1-F_{\chi^2_4}(28.21)=0.0000113\;<\;0.0001.$$

(e)

Let $\mu^{(t)}$ be the p–vector of population means under treatment t. Then the hypotheses tested by the MANOVA are

$$\begin{aligned}H_0:&\quad \mu^{(1)}=\mu^{(2)}=\mu^{(3)},\\ H_a:&\quad \text{not all of } \mu^{(1)},\mu^{(2)},\mu^{(3)} \text{ are equal.}\end{aligned}$$

In words:

- H_0: There is no multivariate treatment effect; all three treatments have the same mean response vector.

- H_a: At least one treatment differs in its mean response vector from the others.

Question 2

(a)

Use the following R code to get the numbers:

```
p   <- ncol(skull) - 1                # number of variables (excluding 'period')
g   <- length(unique(skull$period))   # number of time-period groups
n_i <- table(skull$period)           # sample size for each period
p
g
n_i   # shows n1, n2, n3
[1] 4
[1] 3
 1  2  3
30 30 30
```

Thus

$$p = 4, \quad g = 3, \quad (n_1, n_2, n_3) = (30,\ 30,\ 30).$$

(b)

Let $\mu^{(t)} \in R^4$ be the true mean vector of $(X_1, \ldots, X_4)$ in period t. Then

$$H_0 : \ \mu^{(1)} = \mu^{(2)} = \mu^{(3)} \quad \text{vs.} \quad H_a : \ \text{not all } \mu^{(t)} \text{ are equal.}$$

(c)

Using the following R code:

```
skull$period <- as.factor(skull$period)
m <- manova(cbind(x1,x2,x3,x4) ~ period, data = skull)
sm <- summary(m, test = "Wilks")
print(sm)

Lambda <- sm$stats[1, "Wilks"]
Fval   <- sm$stats[1, "approx F"]
df1    <- sm$stats[1, "num Df"]
df2    <- sm$stats[1, "den Df"]
pval   <- sm$stats[1, "Pr(>F)"]

cat("Wilks' Lambda:", Lambda, "\n")
cat("Approx F(", df1, ",", df2, ") =", Fval,
    ", p =", pval, "\n")
if(pval < 0.05) {
  cat("→ Reject H0 at =0.05\n")
} else {
  cat("→ Fail to reject H0 at =0.05\n")
}
```

We get the following Results:

```
          Df  Wilks approx F num Df den Df  Pr(>F)
period     2 0.8301   2.0491      8    168 0.04358 *
Residuals 87
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Wilks’ Lambda: 0.8301027
Approx F( 8 , 168 ) = 2.049069 , p = 0.04358254
→ Reject H0 at =0.05
```

we have

$$\Lambda^* = 0.8301027, \qquad F_{\text{obs}} = 2.049069, \quad {}_1 = 8, \quad {}_2 = 168, \quad p\text{-value} = 0.04358254.$$

By Johnson & Wichern, the approximation formula is

$$F \;=\; \frac{\left(\sum_\ell n_\ell - p - 2\right)\left(1 - \sqrt{\Lambda^*}\right)}{p\sqrt{\Lambda^*}} \;\sim\; F_{2p,\, 2\left(\sum_\ell n_\ell - p - 2\right)}.$$

Here $p = 4$ and $\sum_\ell n_\ell = 90$, so the reference distribution is

$$F_{8,\,168}.$$

$$F_{8,168} = 2.0491, \quad p = 1 - \text{pf}(2.049069,\, 8,\, 168) \approx 0.04358 < 0.05,$$

hence we Reject H_0 at $\alpha = 0.05$.

(d)

To ensure that the approximate F–distribution of the Wilks' Λ statistic is valid, the following conditions must hold:

1. **Multivariate normality:** Within each treatment group t, the p–vector $\mathbf{X} = (X_1, \ldots, X_p)'$ is assumed to follow a multivariate normal distribution
$$\mathbf{X} \;\sim\; \mathcal{N}\left(\mu^{(t)}, \Sigma^{(t)}\right).$$
2. **Homogeneity of covariance matrices:** All treatment groups share the same covariance matrix,
$$\Sigma^{(1)} = \Sigma^{(2)} = \cdots = \Sigma^{(g)} = \Sigma.$$
3. **Independence:** Observations are assumed to be independent both within each group, and between different groups.

(e)

We wish to form confidence intervals for every pairwise difference of component means $\mu_j^{(i)} - \mu_j^{(k)}$, where

- $i, k \in \{1, \ldots, g\}$ with $i < k$, and
- $j \in \{1, \ldots, p\}$ indexes the p variables $(X_1, \ldots, X_p)$.

There are

$$\binom{g}{2} \;=\; \binom{3}{2} \;=\; 3$$

distinct pairs of periods (i, k), and $p = 4$ component variables, so the total number of intervals is

$$\binom{g}{2} \times p \;=\; 3 \;\times\; 4 \;=\; 12.$$

Thus we must construct 12 simultaneous confidence intervals to cover all component-wise, period-pair differences while controlling the family-wise error rate.

Question 3

(a)

We apply cmdscale() for $q = 3, 4, 5$ and obtain the site coordinates. The printed results are:

```
rm(list=ls())

mat <- matrix(0, 9, 9)
mat[row(mat) <= col(mat)] <- scan("/Users/yubin/Desktop/Multivariate Analysis/T12-13.DAT")
X = t(mat)
d = as.dist(X)

mds3 = cmdscale(d, k = 3)
mds4 = cmdscale(d, k = 4)
mds5 = cmdscale(d, k = 5)

> print(mds3)
           [,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
> print(mds4)
           [,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
> print(mds5)
           [,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)

Using the following code to get the stress plot:

```
max_q <- 5
stress <- numeric(max_q)
denom  <- sum(d^2)

for(q in 1:max_q) {
  # classical MDS in q dims
  coords_q <- cmdscale(d, k = q)
  # pairwise distances in q-dim solution
  d_q <- dist(coords_q)
  # stress numerator and normalize
  stress[q] <- sqrt( sum((d - d_q)^2) / denom )
}

# plot Stress vs q
plot(1:max_q, stress, type="b", pch=19,
     xlab = "Dimension q",
     ylab = "Stress(q)",
     main = "MDS Stress vs Dimension")
grid()
```

Output:

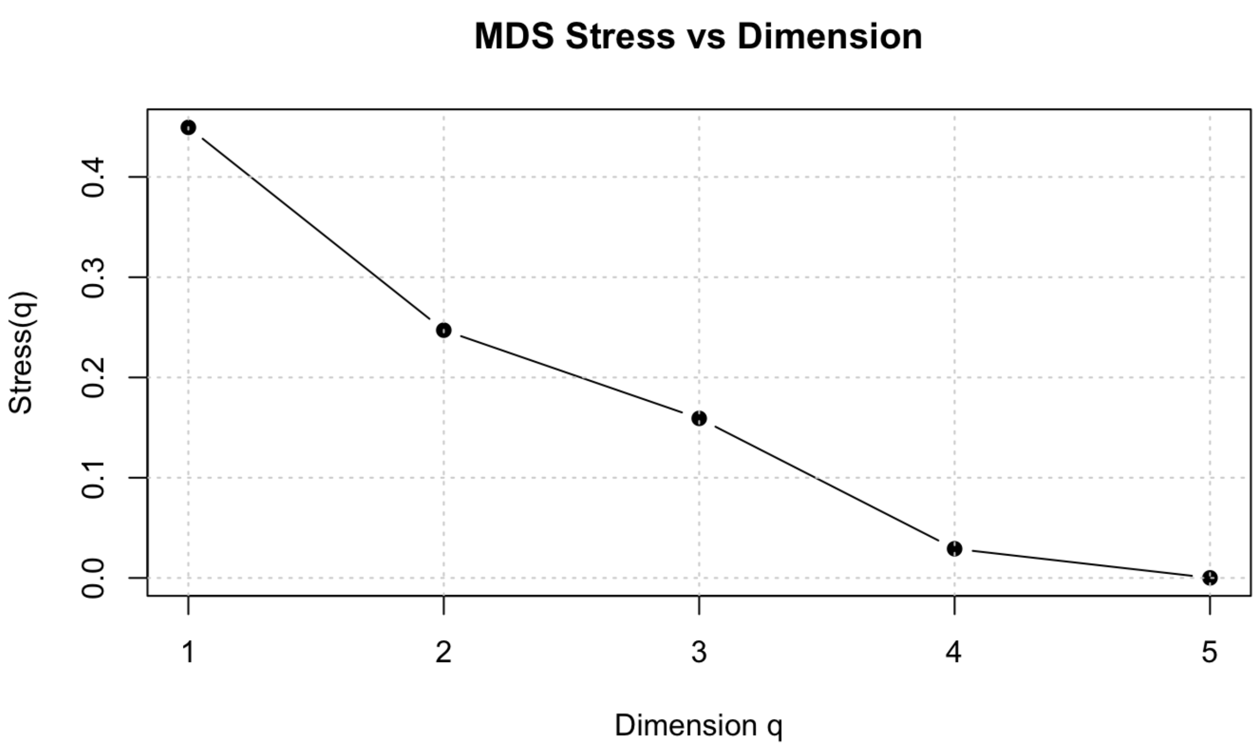

Figure 1: Stress Plot

Interpretation: The plot of Stress(q) shows a sharp decline as q increases from 1 to 4, then levels off. To be more specific:

- Moving from $q = 1$ to $q = 2$, stress drops sharply, representing a poor fit when $q = 1$.

- Going to $q = 3$ further reduces stress, and the largest marginal improvement occurs between $q = 3$ and $q = 4$, where stress falls to 0.028.
- By $q = 4$ the stress is low(below 0.05), indicating that a four-dimensional plot almost perfectly recovers the original distances.

Thus the "elbow" in the plot lies between $q = 3$ and $q = 4$. If one wants a low-dimensional display with stress under 0.05, a four-dimensional solution is preferred; for a more parsimonious graphical display (with tolerable distortion of about 16%), three dimensions may suffice.

(c)

We take the first two dimensions of the $q = 5$ classical MDS solution and label each point by its archaeological date (A.D. year):

```
mds5   <- cmdscale(d, k = 5)
coords <- mds5
years  <- c(918,1131,960,987,1024,1005,945,1137,1062)

plot(coords[,1], coords[,2],
     type="n",
     xlab="MDS Dimension 1",
     ylab="MDS Dimension 2",
     main="MDS Configuration (q=2)",
     cex.axis=0.7)
text(coords[,1], coords[,2],
     labels=years,
     cex=0.8, col="darkred")
```

Output:

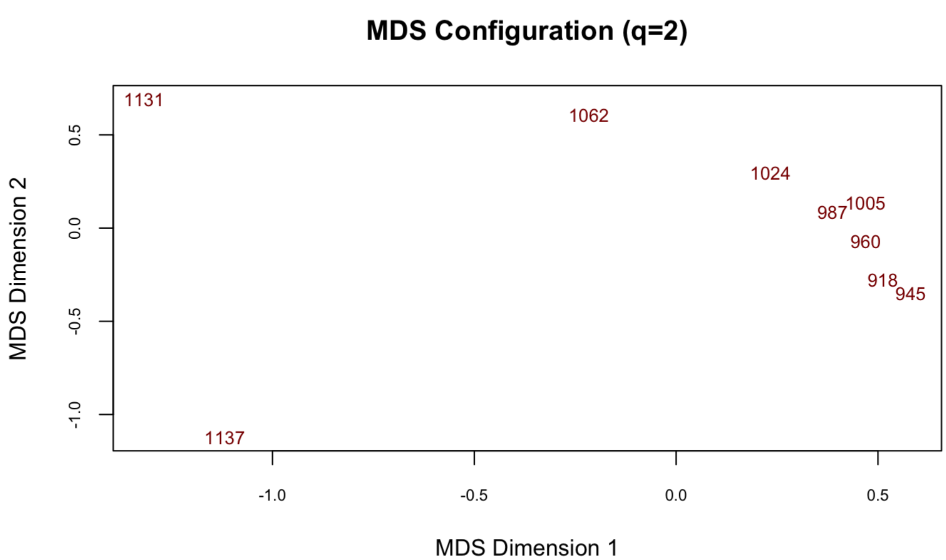

Figure 2: MDS map

Interpretation

The two-dimensional plot shows a rough ordering of sites by date: early period sites cluster at one end, middle period sites lie in the right side, and late period sites cluster at the opposite end. This suggests a broad temporal gradient in the potsherd compositions.

To be more specific:

- The latest periods (A.D. 1137 and 1131) lie far to the left.
- Mid-range dates (A.D. 1062, 1024, 1005, and 987) cluster at the right hand side of the map.
- The earliest sites (A.D. 960, 945, 918) appear on the right.

Archaeologists might interpret this as evidence of a directional change in potsherd composition through time.

Question 4

(a)

Let x_{ij} be the count of students with hair color i and eye color j, and $n = 592$. We use the following R code to compute cell percentages, row percentages and column percentages: **Cell percentages:**

```
data <- read.table("/Users/yubin/Desktop/Multivariate Analysis/HairEyeAll.txt")
rownames(data) <- c("Black","Brown","Red","Blond")   # Hair
colnames(data) <- c("Brown","Blue","Hazel","Green")  # Eye

data = as.matrix(data)
# Total sample size
n <- sum(data)

p = data / n
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
```

Row percentages

```
data_row = data%*%c(1,1,1,1)
round(diag(c(1/data_row))%*%data,2)

     Brown Blue Hazel Green
[1,]  0.63 0.19  0.14  0.05
[2,]  0.42 0.29  0.19  0.10
[3,]  0.37 0.24  0.20  0.20
[4,]  0.06 0.74  0.08  0.13
```

Column percentages

```
data_col = c(1,1,1,1)%*%data
round(data%*%diag(c(1/data_col)),2)
```

```
      [,1] [,2] [,3] [,4]
Black 0.31 0.09 0.16 0.08
Brown 0.54 0.39 0.58 0.45
Red   0.12 0.08 0.15 0.22
Blond 0.03 0.44 0.11 0.25
```

(b)

Under H_0: "Hair" and "Eye" are independent, the expected count in cell (i, j) is

$$E_{ij} = \frac{n_{i\bullet}\, n_{\bullet j}}{n},$$

where $n_{i\bullet} = \sum_j x_{ij}$, $n_{\bullet j} = \sum_i x_{ij}$, and $n = \sum_{i,j} x_{ij} = 592$.

```
E = data_row%*%data_col/ n
round(E,2)
```

```
       Brown   Blue Hazel Green
Black  40.14  39.22 16.97 11.68
Brown 106.28 103.87 44.93 30.92
Red    26.39  25.79 11.15  7.68
Blond  47.20  46.12 19.95 13.73
```

(c)

H_0: "Hair" and "Eye" are independent

The contributions of each cell are

$$m_{ij} = \frac{(x_{ij} - E_{ij})^2}{E_{ij}},$$

where $E_{ij} = n_{i\bullet}\, n_{\bullet j}/n$.

In R we compute:

```
chisq.test(data)

chi2 = sum((data-E)^2/E)
inertia = chi2 / n
inertia
```

```
Pearson's Chi-squared test

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

[1] 0.2335977
```

We have Pearson chi-square statistic,

$$\chi^2 = \sum_{i,j} m_{ij} \;=\; 138.29,$$

with degrees of freedom

$$df = (r-1)(c-1) = (4-1)(4-1) = 9,$$

and since p-value $< 2.2 \times 10^{-16}$.

Thus we reject H_0 of independence.

The total inertia in correspondence analysis is defined as

$$\text{Inertia} = \frac{\chi^2}{n} = \frac{138.29}{592} \approx 0.23360.$$

We can check that this equals the sum of the CA principal inertias (squared singular values) by

```
library(ca)
res.ca <- ca(data)
total_inertia <- sum(res.ca$sv^2)
print(total_inertia)
```

```
[1] 0.2335977
```

Under independence, the Pearson χ^2 statistic is

$$\chi^2 \;=\; \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(x_{ij} - E_{ij})^2}{E_{ij}},$$

and the total inertia is

$$\text{Inertia} = \frac{\chi^2}{n} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(x_{ij} - E_{ij})^2}{n\,E_{ij}}.$$

Since $p_{ij} = x_{ij}/n$, $r_i = \sum_j p_{ij}$, $c_j = \sum_i p_{ij}$, and $E_{ij} = n\,r_i\,c_j$, one shows

$$\text{Inertia} = \sum_{i,j} \frac{(p_{ij} - r_i\,c_j)^2}{r_i\,c_j}.$$

In correspondence analysis, define

$$P = (p_{ij})_{i=1,\dots,I}^{j=1,\dots,J}, \quad D_r = \mathrm{diag}(r_1, \dots, r_I), \quad D_c = \mathrm{diag}(c_1, \dots, c_J).$$

The standardized residual matrix is

$$S = D_r^{-1/2}\,(P - r\,c^T)\,D_c^{-1/2},$$

and its singular value decomposition

$$S = U\,\Sigma\,V^T, \qquad \Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_k)$$

yields the principal inertias $\sigma_1^2, \dots, \sigma_k^2$. It follows that

$$\sum_{k=1}^{k} \sigma_k^2 = \mathrm{trace}(S^T S) = \sum_{i,j} s_{ij}^2 = \sum_{i,j} \frac{(p_{ij} - r_i c_j)^2}{r_i c_j} = \frac{\chi^2}{n}.$$

Hence computing χ^2/n from the contingency table and summing the eigenvalues σ_k^2 from the CA decomposition both give the same total inertia, confirming the required check.

(d)

```
ca(data)
plot(ca(data),
     map  = "symbiplot",
     main = "Correspondence Analysis Biplot\n(Hair vs. Eye Colors)")

Principal inertias (eigenvalues):
           1        2        3
Value      0.208773 0.022227 0.002598
Percentage 89.37%   9.52%    1.11%


 Rows:
            Black     Brown       Red    Blond
Mass     0.182432  0.483108  0.119932 0.214527
ChiDist  0.551192  0.159461  0.354770 0.838397
Inertia  0.055425  0.012284  0.015095 0.150793
Dim. 1  -1.104277 -0.324463 -0.283473 1.828229
Dim. 2   1.440917 -0.219111 -2.144015 0.466706


 Columns:
            Brown     Blue     Hazel     Green
Mass     0.371622 0.363176  0.157095  0.108108
ChiDist  0.500487 0.553684  0.288654  0.385727
Inertia  0.093086 0.111337  0.013089  0.016085
Dim. 1  -1.077128 1.198061 -0.465286  0.354011
Dim. 2   0.592420 0.556419 -1.122783 -2.274122
```

Output:

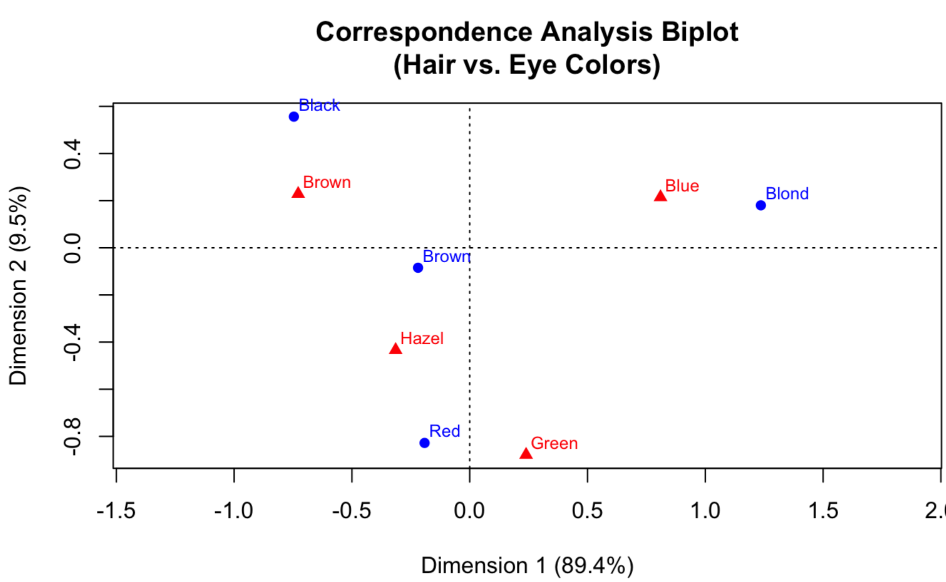

Figure 3: CA Plot

From the CA of the 4×4 Hair x Eye table we obtain the principal inertias (percent variance explained):

$$\text{Dimension 1:} \quad 89.37\%\,, \qquad \text{Dimension 2:} \quad 9.52\%\,.$$

Thus the two-dimensional biplot captures

$$89.37\% + 9.52\% = 98.89\%$$

of the total inertia (variation) in the data.

Associations The CA biplot makes clear that hair color and eye color are not independent: certain combinations occur far more often than expected under independence.

In particular,

- Blond hair and blue eyes lie very close together, indicating a strong positive association—blond-haired people are much more likely to have blue eyes.
- Brown hair and brown eyes also cluster together, showing that brown hair and brown eyes tend to co-occur.
- red hair tends to appear near green eyes, and black hair near hazel eyes.

These proximity patterns accord with common experience (for example, fair hair often comes with light eyes), and confirm a clear correlation between hair and eye color in the sample.

Question 5

(a)

Consider a general case:

Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}\Big).$$

That is, (X, Y) is jointly normal with mean vector $\mu = (\mu_x, \mu_y)$ and covariance matrix Σ as above.

Lemma: From the formula of block matrix, if we have a partitioned matrix

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}.$$

A standard result for its inverse (assuming the necessary inverses exist) is

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{xx.y}^{-1} & -\Sigma_{xx.y}^{-1}\,\Sigma_{xy}\,\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\,\Sigma_{yx}\,\Sigma_{xx.y}^{-1} & \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\,\Sigma_{yx}\,\Sigma_{xx.y}^{-1}\,\Sigma_{xy}\,\Sigma_{yy}^{-1} \end{pmatrix},$$

where

$$\Sigma_{xx.y} = \Sigma_{xx} - \Sigma_{xy}\,\Sigma_{yy}^{-1}\,\Sigma_{yx}.$$

Marginally,

$$X \sim \mathcal{N}(\mu_x,\ \Sigma_{xx}), \quad Y \sim \mathcal{N}(\mu_y,\ \Sigma_{yy}).$$

The joint density is

$$f_{X,Y}(x, y) = \frac{1}{(2\pi)^{\frac{n}{2}}\,|\Sigma|^{\frac{1}{2}}}\,\exp\Big(-\tfrac{1}{2}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}^T \Sigma^{-1}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}\Big).$$

The conditional density of X given $Y = y$ is

$$f_{X|Y}(x \mid y) \;=\; \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

and the marginal density of Y as

$$f_Y(y) \;\propto\; \exp\!\Big(-\tfrac{1}{2}\,(y-\mu_y)^T\,\Sigma_{yy}^{-1}\,(y-\mu_y)\Big).$$

Hence,

$$f_{X|Y}(x \mid y) \;=\; \frac{f_{X,Y}(x,y)}{f_Y(y)} \;\propto\; \exp\!\Big(-\tfrac{1}{2}\,\big[(x-\mu_x),(y-\mu_y)\big]\,\Sigma^{-1}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix} \;+\; \tfrac{1}{2}\,(y-\mu_y)^T\,\Sigma_{yy}^{-1}\,(y-\mu_y)\Big).$$

Calculate the exponential term and factor out $-\frac{1}{2}$:

$$\Big(\big[(x-\mu_x),(y-\mu_y)\big]\,\Sigma^{-1}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix} \;-\; (y-\mu_y)^T\,\Sigma_{yy}^{-1}\,(y-\mu_y)\Big)$$

$$=\; (x-\mu_x)^T\,\Sigma_{(x,x)}^{-1}\,(x-\mu_x) \,+\, (x-\mu_x)^T\,\Sigma_{(x,y)}^{-1}\,(y-\mu_y) \,+\, (y-\mu_y)^T\,\Sigma_{(y,x)}^{-1}\,(x-\mu_x) \,+\, (y-\mu_y)^T\big[\Sigma_{(y,y)}^{-1} \,-\, \Sigma_{yy}^{-1}\big]\,(y-\mu_y).$$

$$= (x-\mu_x)^T\Sigma xx.y^{-1}(x-\mu_x) + (x-\mu_x)^T(-\Sigma_{xx.y}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y) + (y-\mu_y)^T(-\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1})(x-\mu_x)$$
$$+(y-\mu_y)^T(\Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} - \Sigma_{yy}^{-1})(y-\mu_y)$$

$$= (x-\mu_x)^T\Sigma_{xx.y}^{-1}(x-\mu_x) + (x-\mu_x)^T(-\Sigma_{xx.y}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y) + (y-\mu_y)^T(-\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1})(x-\mu_x)$$
$$+(y-\mu_y)^T(\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y)$$

$$= (x-\mu_x)^T\Sigma_{xx.y}^{-1}[(x-\mu_x) - (\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y)] + (y-\mu_y)^T(\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx.y}^{-1})[(\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y) - (x-\mu_x)]$$

$$= [(x-\mu_x)^T - (y-\mu_y)^T(\Sigma_{yy}^{-1}\Sigma_{yx})]\Sigma_{xx.y}^{-1}[(x-\mu_x) - (\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y)]$$

Since

$$\Sigma_{yx}^T \;=\; \Sigma_{xy}, \quad (\Sigma_{yy}^{-1})^T \;=\; \Sigma_{yy}^{-1}, \quad \text{and} \quad (AB)^T = B^T A^T.$$

Specifically,

$$\begin{aligned}
(y-\mu_y)^T\,(\Sigma_{yy}^{-1}\,\Sigma_{yx}) &= \big[(\Sigma_{yy}^{-1}\,\Sigma_{yx})^T\,(y-\mu_y)\big]^T && \text{(since it is a scalar, equals its transpose)}\\
&= \big[\Sigma_{yx}^T\,(\Sigma_{yy}^{-1})^T\,(y-\mu_y)\big]^T && \text{(transpose of a product reverses the order)}\\
&= \big[\Sigma_{xy}\,\Sigma_{yy}^{-1}\,(y-\mu_y)\big]^T && \text{(using } \Sigma_{yx}^T = \Sigma_{xy} \text{ and } \Sigma_{yy}^{-1} \text{ is symmetric)}\\
&= \big[\Sigma_{xy}\,\Sigma_{yy}^{-1}\,(y-\mu_y)\big]^T.
\end{aligned}$$

Therefore,

$$(x-\mu_x)^T \;-\; (y-\mu_y)^T\,(\Sigma_{yy}^{-1}\,\Sigma_{yx}) \;=\; (x-\mu_x)^T \;-\; \big[\Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\big]^T.$$

For vectors a and b, we know $a^T - b^T = (a-b)^T$. Hence the above difference can be written as

$$\left[(x-\mu_x) \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\right]^T.$$

Therefore,

$$\begin{aligned}&\Big\{(x-\mu_x)^T \;-\; (y-\mu_y)^T(\Sigma_{yy}^{-1}\,\Sigma_{yx})\Big\}\,\Sigma_{xx.y}^{-1}\,\Big\{(x-\mu_x) \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\Big\}\\ &= \left[x-\mu_x-\Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\right]^T\Sigma_{xx.y}^{-1}\left[x-\mu_x-\Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\right].\end{aligned}$$

Therefore,

$$f_{X|Y}(x\mid y) \propto \exp\!\Big(-\tfrac{1}{2}\left[x-\mu_x-\Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\right]^T\Sigma_{xx.y}^{-1}\left[x-\mu_x-\Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\right],$$

we see that the exponent is the usual quadratic form

$$-\tfrac{1}{2}\left[x-\mu_x \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\right]^T\Sigma_{xx.y}^{-1}\left[x-\mu_x \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y)\right].$$

This indicates that $f(x \mid y)$ has the kernel of a multivariate normal density in x with shifted mean

$$\mu_x + \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y) \quad \text{and covariance} \quad \Sigma_{xx.y} \;=\; \Sigma_{xx} \;-\; \Sigma_{xy}\,\Sigma_{yy}^{-1}\,\Sigma_{yx}.$$

Hence, putting the normalizing constant back in, we conclude

$$(X \mid Y=y) \;\sim\; \mathcal{N}\Big(\mu_x+\Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y),\; \Sigma_{xx}-\Sigma_{xy}\,\Sigma_{yy}^{-1}\,\Sigma_{yx}\Big).$$

In other words, the conditional distribution $X \mid Y = y$ is still Gaussian.

Thus, we have:

(i) Conditional Expectation

$$E(X \mid Y=y) = \mu_x + \Sigma_{xy}\,\Sigma_{yy}^{-1}(y-\mu_y).$$

(ii) Conditional Variance

$$\mathrm{Var}(X \mid Y=y) = \Sigma_{xx} - \Sigma_{xy}\,\Sigma_{yy}^{-1}\,\Sigma_{yx}.$$

In this specific sample: Write

$$X = X_1, \quad Y = \begin{pmatrix} X_2 \\ X_3 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_Y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{1Y} \\ \Sigma_{Y1} & \Sigma_{YY} \end{pmatrix},$$

where

$$\Sigma_{1Y} = \begin{pmatrix} \Sigma_{12} & \Sigma_{13} \end{pmatrix}, \qquad \Sigma_{YY} = \begin{pmatrix} \Sigma_{22} & 0 \\ 0 & \Sigma_{33} \end{pmatrix}.$$

By the formula above,

$$E[X \mid Y=y] = \mu_1 + \Sigma_{1Y}\,\Sigma_{YY}^{-1}(y-\mu_Y).$$

Since

$$\Sigma_{YY}^{-1} = \begin{pmatrix} \Sigma_{22}^{-1} & 0 \\ 0 & \Sigma_{33}^{-1} \end{pmatrix}, \quad y-\mu_Y = \begin{pmatrix} x_2-\mu_2 \\ x_3-\mu_3 \end{pmatrix},$$

$$\Sigma_{1Y}\,\Sigma_{YY}^{-1}\,(y-\mu_Y) = \begin{pmatrix} \Sigma_{12} & \Sigma_{13} \end{pmatrix} \underbrace{\begin{pmatrix} \Sigma_{22}^{-1} & 0 \\ 0 & \Sigma_{33}^{-1} \end{pmatrix}}_{\Sigma_{YY}^{-1}} \underbrace{\begin{pmatrix} x_2-\mu_2 \\ x_3-\mu_3 \end{pmatrix}}_{y-\mu_Y}$$

$$\begin{pmatrix} \Sigma_{22}^{-1} & 0 \\ 0 & \Sigma_{33}^{-1} \end{pmatrix}\begin{pmatrix} x_2-\mu_2 \\ x_3-\mu_3 \end{pmatrix} = \begin{pmatrix} \Sigma_{22}^{-1}(x_2-\mu_2) \\ \Sigma_{33}^{-1}(x_3-\mu_3) \end{pmatrix}.$$

$$\begin{pmatrix}\Sigma_{12} & \Sigma_{13}\end{pmatrix}\begin{pmatrix}\Sigma_{22}^{-1}(x_2-\mu_2)\\ \Sigma_{33}^{-1}(x_3-\mu_3)\end{pmatrix} = \Sigma_{12}\,\Sigma_{22}^{-1}\,(x_2-\mu_2) \;+\; \Sigma_{13}\,\Sigma_{33}^{-1}\,(x_3-\mu_3).$$

$$\Sigma_{1Y}\,\Sigma_{YY}^{-1}\,(y-\mu_Y) = \Sigma_{12}\,\Sigma_{22}^{-1}\,(x_2-\mu_2) + \Sigma_{13}\,\Sigma_{33}^{-1}\,(x_3-\mu_3),$$

Thus,

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

(b)

Again by the general formula above:

$$Var(X \mid Y = y) = \Sigma_{11} \;-\; \Sigma_{1Y}\,\Sigma_{YY}^{-1}\,\Sigma_{Y1}.$$

$$\Sigma_{1Y}\,\Sigma_{YY}^{-1}\,\Sigma_{Y1} = (\Sigma_{12},\Sigma_{13})\begin{pmatrix}\Sigma_{22}^{-1} & 0\\ 0 & \Sigma_{33}^{-1}\end{pmatrix}\begin{pmatrix}\Sigma_{21}\\ \Sigma_{31}\end{pmatrix} = \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21} + \Sigma_{13}\,\Sigma_{33}^{-1}\,\Sigma_{31}.$$

Hence

$$Var\big(X_1 \mid X_2 = x_2,\, X_3 = x_3\big) = \Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21} - \Sigma_{13}\,\Sigma_{33}^{-1}\,\Sigma_{31}.$$

(c)

Let

$$Y \;=\; X_2 + X_3.$$

Since $\mathrm{Cov}(X_2, X_3) = 0$, X_2 and X_3 are independent normals. Hence

$$Y \;\sim\; \mathcal{N}\big(\mu_2+\mu_3,\; \Sigma_{22}+\Sigma_{33}\big).$$

Now form the vector

$$\begin{pmatrix}X_1\\ Y\end{pmatrix}$$

whose mean and covariance are

$$E\begin{pmatrix}X_1\\ Y\end{pmatrix} = \begin{pmatrix}\mu_1\\ \mu_2+\mu_3\end{pmatrix}, \qquad \mathrm{Cov}\begin{pmatrix}X_1\\ Y\end{pmatrix} = \begin{pmatrix}\Sigma_{11} & \mathrm{Cov}(X_1, Y)\\ \mathrm{Cov}(Y, X_1) & \Sigma_{22}+\Sigma_{33}\end{pmatrix}.$$

By linearity of covariance:

$$\mathrm{Cov}(X_1, Y) = \mathrm{Cov}\big(X_1, X_2 + X_3\big) = \mathrm{Cov}\big(X_1, X_2\big) + \mathrm{Cov}\big(X_1 + X_3\big) = \Sigma_{12} + \Sigma_{13},$$

and $(\Sigma_{12} + \Sigma_{13})^T = \Sigma_{21} + \Sigma_{31}$:

$$\begin{pmatrix}X_1\\ X_2+X_3\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\mu_1\\ \mu_2+\mu_3\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}+\Sigma_{13}\\ \Sigma_{21}+\Sigma_{31} & \Sigma_{22}+\Sigma_{33}\end{pmatrix}\right)$$

(d)

Let

$$Y \;=\; X_2 + X_3, \qquad Z \;=\; \begin{pmatrix}X_1\\ Y\end{pmatrix}.$$

From part (c), we have

$$E[Z] = \begin{pmatrix}\mu_1\\ \mu_2+\mu_3\end{pmatrix}, \qquad \mathrm{Cov}(Z) = \begin{pmatrix}\Sigma_{11} & \Sigma_{1Y}\\ \Sigma_{Y1} & \Sigma_{YY}\end{pmatrix},$$

where

$$\Sigma_{1Y} = \mathrm{Cov}\big(X_1,\ X_2 + X_3\big) = \Sigma_{12} + \Sigma_{13}, \qquad \Sigma_{YY} = \mathrm{Cov}\big(X_2 + X_3,\ X_2 + X_3\big) = \Sigma_{22} + \Sigma_{33}.$$

By the block-Gaussian conditioning formula,

$$E\big[X_1 \mid Y = x_0\big] = \mu_1 + \Sigma_{1Y}\,\Sigma_{YY}^{-1}\,\big(x_0 - (\mu_2 + \mu_3)\big),$$

$$Var\big(X_1 \mid Y\big) = \Sigma_{11} - \Sigma_{1Y}\,\Sigma_{YY}^{-1}\,\Sigma_{Y1}.$$

Substituting $\Sigma_{1Y} = \Sigma_{12} + \Sigma_{13}$, $\Sigma_{YY} = \Sigma_{22} + \Sigma_{33}$ and $\mu_Y = \mu_2 + \mu_3$ gives:

$$\Sigma_{1Y}\,\Sigma_{YY}^{-1}\,\big(x_0 - (\mu_2 + \mu_3)\big) = \big(\Sigma_{12}\ +\ \Sigma_{13}\big)(\Sigma_{22} + \Sigma_{33})^{-1}\big(x_0 - (\mu_2 + \mu_3)\big)$$

Similary,

$$\Sigma_{1Y}\,\Sigma_{YY}^{-1}\,\Sigma_{Y1} = (\Sigma_{12} + \Sigma_{13})\,(\Sigma_{22} + \Sigma_{33})^{-1}\,(\Sigma_{21} + \Sigma_{31}).$$

$$E\big[X_1 \mid X_2 + X_3 = x_0\big] = \mu_1 + (\Sigma_{12} + \Sigma_{13})\,(\Sigma_{22} + \Sigma_{33})^{-1}\,\big(x_0 - (\mu_2 + \mu_3)\big),$$

$$Var\big(X_1 \mid X_2 + X_3\big) = \Sigma_{11} - (\Sigma_{12} + \Sigma_{13})\,(\Sigma_{22} + \Sigma_{33})^{-1}\,(\Sigma_{21} + \Sigma_{31}).$$

Thus,

$$\big(X_1 \mid X_2{+}X_3 = x_0\big) \sim \mathcal{N}\Big(\mu_1{+}(\Sigma_{12}{+}\Sigma_{13})(\Sigma_{22}{+}\Sigma_{33})^{-1}(x_0{-}\mu_2{-}\mu_3),\ \Sigma_{11}\ -\ (\Sigma_{12}{+}\Sigma_{13})(\Sigma_{22}{+}\Sigma_{33})^{-1}(\Sigma_{21}{+}\Sigma_{31})\Big).$$