Assignment 3

STAT 32950

Ki Hyun

Due: 09:00 (CT) 2023-04-11

Q1.

(a)

i)

$$\bar{\mathbf{x}} = \frac{1}{4}\sum_{i=1}^{4}\mathbf{x}_i = \frac{1}{4}\left(\begin{pmatrix}2\\12\end{pmatrix} + \begin{pmatrix}8\\9\end{pmatrix} + \begin{pmatrix}6\\9\end{pmatrix} + \begin{pmatrix}8\\10\end{pmatrix}\right) = \begin{pmatrix}6\\10\end{pmatrix}$$

ii)

$$\begin{aligned}
\mathbf{S} &= \frac{1}{4-1}\sum_{j=1}^{4}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T \\
&= \frac{1}{3}\sum_{j=1}^{4}\begin{pmatrix}x_{j,1}-6\\x_{j,2}-10\end{pmatrix}\begin{pmatrix}x_{j,1}-6\\x_{j,2}-10\end{pmatrix}^T \\
&= \frac{1}{3}\sum_{j=1}^{4}\begin{pmatrix}(x_{j,1}-6)^2 & (x_{j,1}-6)(x_{j,2}-10)\\(x_{j,2}-10)(x_{j,1}-6) & (x_{j,2}-10)^2\end{pmatrix} \\
&= \begin{pmatrix}\frac{1}{3}\sum_{j=1}^{4}(x_{j,1}-6)^2 & \frac{1}{3}\sum_{j=1}^{4}(x_{j,1}-6)(x_{j,2}-10)\\ \frac{1}{3}\sum_{j=1}^{4}(x_{j,2}-10)(x_{j,1}-6) & \frac{1}{3}\sum_{j=1}^{4}(x_{j,2}-10)^2\end{pmatrix} \\
&= \begin{pmatrix}8 & -\frac{10}{3}\\ -\frac{10}{3} & 2\end{pmatrix}
\end{aligned}$$

iii)

$$\begin{aligned}
\mathbf{S}^{-1} &= \begin{pmatrix}8 & -\frac{10}{3}\\ -\frac{10}{3} & 2\end{pmatrix}^{-1} \\
&= \frac{1}{8\cdot 2 - (-\frac{10}{3})^2}\begin{pmatrix}2 & \frac{10}{3}\\ \frac{10}{3} & 8\end{pmatrix} \\
&= \frac{9}{44}\begin{pmatrix}2 & \frac{10}{3}\\ \frac{10}{3} & 8\end{pmatrix} \\
&= \begin{pmatrix}\frac{9}{22} & \frac{15}{22}\\ \frac{15}{22} & \frac{18}{11}\end{pmatrix}
\end{aligned}$$

iv)

$$
\begin{aligned}
T^2 &= (\bar{\mathbf{x}} - \mu_{\mathbf{0}})^T \left(\frac{\mathbf{S}}{4}\right)^{-1} (\bar{\mathbf{x}} - \mu_{\mathbf{0}}) \\
&= (\bar{\mathbf{x}} - \mu_{\mathbf{0}})^T 4\mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu_{\mathbf{0}}) \\
&= \begin{pmatrix} -1 \\ -1 \end{pmatrix}^T \begin{bmatrix} \frac{18}{11} & \frac{30}{11} \\ \frac{30}{11} & \frac{72}{11} \end{bmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} \\
&= \begin{bmatrix} -\frac{48}{11} & -\frac{102}{11} \end{bmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} \\
&= \frac{150}{11}
\end{aligned}
$$

v)

$$
T^2 \sim \frac{4-2}{(4-1)\cdot 2} F_{2,4-2} = \frac{1}{3} F_{2,2}
$$

(b)

i)

$$
\begin{aligned}
\bar{\mathbf{y}} &= \frac{1}{4}\sum_{j=1}^{4}\mathbf{y_j} \\
&= \frac{1}{4}\sum_{j=1}^{4}C\mathbf{x_j} \\
&= \frac{1}{4}C\left(\sum_{j=1}^{4}\mathbf{x_j}\right) \\
&= C\left(\frac{1}{4}\sum_{j=1}^{4}\mathbf{x_j}\right) \\
&= C\bar{\mathbf{x}} \\
&= \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{pmatrix} 6 \\ 10 \end{pmatrix} \\
&= \begin{pmatrix} 16 \\ 4 \end{pmatrix}
\end{aligned}
$$

ii)

$$
\begin{aligned}
S_y &= C\mathbf{S}C^T \\
&= \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 8 & -\frac{10}{3} \\ -\frac{10}{3} & 2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{14}{3} & -\frac{4}{3} \\ -\frac{34}{3} & \frac{16}{3} \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{10}{3} & -6 \\ -6 & \frac{50}{3} \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
T^2 &= (\bar{\mathbf{y}} - \mu_{\mathbf{0}}^*)^T\left(\frac{S_y}{4}\right)^{-1}(\bar{\mathbf{y}} - \mu_{\mathbf{0}}^*) \\
&= \begin{pmatrix} -2 \\ 0 \end{pmatrix}^T\begin{bmatrix} \frac{5}{6} & -\frac{3}{2} \\ -\frac{3}{2} & \frac{25}{6} \end{bmatrix}^{-1}\begin{pmatrix} -2 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -2 \\ 0 \end{pmatrix}^T\frac{1}{\frac{5}{6}\cdot\frac{25}{6} - \left(-\frac{3}{2}\right)^2}\begin{bmatrix} \frac{25}{6} & \frac{3}{2} \\ \frac{3}{2} & \frac{5}{6} \end{bmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -2 \\ 0 \end{pmatrix}^T\frac{9}{11}\begin{bmatrix} \frac{25}{6} & \frac{3}{2} \\ \frac{3}{2} & \frac{5}{6} \end{bmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -2 \\ 0 \end{pmatrix}^T\begin{bmatrix} \frac{75}{22} & \frac{27}{22} \\ \frac{27}{22} & \frac{15}{22} \end{bmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -\frac{75}{11} & -\frac{27}{11} \end{pmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} \\
&= \frac{150}{11}
\end{aligned}
$$

(c)

First of all, let's note the three relationships that C links between the original and transformed data.

$$\bar{\mathbf{y}} = C\bar{\mathbf{x}} \ \ldots \ (1)$$

$$S_y = C\mathbf{S}C^T \ \ldots \ (2)$$

$$\mu_{\mathbf{0}}^* = C\mu_{\mathbf{0}} \ \ldots \ (3)$$

We know that the Hotelling's T^2 statistic gets evaluated under $H_0 : \bar{\mathbf{x}} = \mu_{\mathbf{0}}$ as:

$$T_x^2 = (\bar{\mathbf{x}} - \mu_{\mathbf{0}})^T \left(\frac{\mathbf{S}}{n}\right)^{-1} (\bar{\mathbf{x}} - \mu_{\mathbf{0}})$$

Similarly, under $H_0 : \bar{\mathbf{y}} = \mu_{\mathbf{0}}^*$:

$$\begin{aligned}
T_y^2 &= (\bar{\mathbf{y}} - \mu_{\mathbf{0}}^*)^T \left(\frac{S_y}{n}\right)^{-1} (\bar{\mathbf{y}} - \mu_{\mathbf{0}}^*) \\
&= (C\bar{\mathbf{x}} - C\mu_{\mathbf{0}})^T \left(\frac{S_y}{n}\right)^{-1} (C\bar{\mathbf{x}} - C\mu_{\mathbf{0}}) \\
&(\because (1), (3)) \\
&= (C(\bar{\mathbf{x}} - \mu_{\mathbf{0}}))^T \left(\frac{S_y}{n}\right)^{-1} (C(\bar{\mathbf{x}} - \mu_{\mathbf{0}})) \\
&= (C(\bar{\mathbf{x}} - \mu_{\mathbf{0}}))^T \left(\frac{C\mathbf{S}C^T}{n}\right)^{-1} (C(\bar{\mathbf{x}} - \mu_{\mathbf{0}})) \\
&(\because (2)) \\
&= (C(\bar{\mathbf{x}} - \mu_{\mathbf{0}}))^T \left(C\frac{\mathbf{S}}{n}C^T\right)^{-1} (C(\bar{\mathbf{x}} - \mu_{\mathbf{0}})) \\
&= (\bar{\mathbf{x}} - \mu_{\mathbf{0}})^T C^T \left(C\frac{\mathbf{S}}{n}C^T\right)^{-1} C(\bar{\mathbf{x}} - \mu_{\mathbf{0}}) \\
&= (\bar{\mathbf{x}} - \mu_{\mathbf{0}})^T C^T (C^T)^{-1} \left(\frac{\mathbf{S}}{n}\right)^{-1} C^{-1}C(\bar{\mathbf{x}} - \mu_{\mathbf{0}}) \\
&(\because C \text{ is square and invertible}) \\
&= (\bar{\mathbf{x}} - \mu_{\mathbf{0}})^T I \left(\frac{\mathbf{S}}{n}\right)^{-1} I(\bar{\mathbf{x}} - \mu_{\mathbf{0}}) \\
&= (\bar{\mathbf{x}} - \mu_{\mathbf{0}})^T \left(\frac{\mathbf{S}}{n}\right)^{-1} (\bar{\mathbf{x}} - \mu_{\mathbf{0}}) \\
&= T_x^2
\end{aligned}$$

$$Q.E.D.$$

Q2.

(a)

First one would be a matrix in the form CC^T as discussed in lecture.

(Here, there is no dimensional difference between C^TC and CC^T)

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$$

Another one would be the **A** matrix:

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Another one would be the **B** matrix:

$$\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

(b)

Given the matrixes above, let's calculate the **A** matrix with the given information and decompose to eigen-values using R.

```
sigma_11 <- matrix(c(8, 2, 2, 5), nrow = 2)
sigma_22 <- matrix(c(6, -2, -2, 7), nrow = 2)
sigma_12 <- matrix(c(3, -1, 1, 3), nrow = 2)
sigma_21 <- t(sigma_12)

A = solve(sigma_11) %*% sigma_12 %*% solve(sigma_22) %*% sigma_21

rho_star <- eigen(A)$values
a <- eigen(A)$vectors
```

Therefore, $\rho_1^* \approx 0.3046268$ and $\rho_2^* \approx 0.2399638$

(c)

```
B = solve(sigma_22) %*% sigma_21 %*% solve(sigma_11) %*% sigma_12

b <- eigen(B)$vectors
```

Given the eigen-vectors, for ρ_1^*:

$$U_1 = \mathbf{a}_1^T\mathbf{X} = \begin{pmatrix} 0.658354 \\ -0.7527084 \end{pmatrix}^T \mathbf{X}$$

$$V_1 = \mathbf{b}_1^T\mathbf{Y} = \begin{pmatrix} 0.9676678 \\ -0.2522283 \end{pmatrix}^T \mathbf{Y}$$

Moreover, for ρ_2^*:

$$U_2 = \mathbf{a}_2^T \mathbf{X} = \begin{pmatrix} 0.5452882 \\ 0.8382486 \end{pmatrix}^T \mathbf{X}$$

$$V_2 = \mathbf{b}_2^T \mathbf{Y} = \begin{pmatrix} 0.5058921 \\ 0.8625968 \end{pmatrix}^T \mathbf{Y}$$

(d)

$$\begin{aligned}
E\left(\begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}\right) &= E\left(\begin{bmatrix} U_1 & U_2 \\ V_1 & V_2 \end{bmatrix}\right) \\
&= E\left(\begin{bmatrix} \begin{pmatrix} 0.658354 \\ -0.7527084 \end{pmatrix}^T \mathbf{X} & \begin{pmatrix} 0.5452882 \\ 0.8382486 \end{pmatrix}^T \mathbf{X} \\ \begin{pmatrix} 0.9676678 \\ -0.2522283 \end{pmatrix}^T \mathbf{Y} & \begin{pmatrix} 0.5058921 \\ 0.8625968 \end{pmatrix}^T \mathbf{Y} \end{bmatrix}\right) \\
&= \begin{bmatrix} \begin{pmatrix} 0.658354 \\ -0.7527084 \end{pmatrix}^T E(\mathbf{X}) & \begin{pmatrix} 0.5452882 \\ 0.8382486 \end{pmatrix}^T E(\mathbf{X}) \\ \begin{pmatrix} 0.9676678 \\ -0.2522283 \end{pmatrix}^T E(\mathbf{Y}) & \begin{pmatrix} 0.5058921 \\ 0.8625968 \end{pmatrix}^T E(\mathbf{Y}) \end{bmatrix} \\
&= \begin{bmatrix} \begin{pmatrix} 0.658354 \\ -0.7527084 \end{pmatrix}^T \mu_1 & \begin{pmatrix} 0.5452882 \\ 0.8382486 \end{pmatrix}^T \mu_1 \\ \begin{pmatrix} 0.9676678 \\ -0.2522283 \end{pmatrix}^T \mu_2 & \begin{pmatrix} 0.5058921 \\ 0.8625968 \end{pmatrix}^T \mu_2 \end{bmatrix} \\
&\approx \begin{bmatrix} -3.48 & 0.041 \\ -0.252 & 0.863 \end{bmatrix}
\end{aligned}$$

For the variance, we know the below

$$\begin{aligned}
&Var(U_k) = Var(V_k) = 1 \\
&Cov(U_k, U_l) = Cov(V_k, V_l) = 0, \ \forall (k \neq l) \\
&Cov(U_k, V_l) = 0, \ \forall (k \neq l) \\
&Cov(U_k, V_k) = \rho_k^*
\end{aligned}$$

Therefore,

$$Cov\left(\begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}\right) = \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} \rho_1^* & 0 \\ 0 & \rho_2^* \end{bmatrix} \\ \begin{bmatrix} \rho_1^* & 0 \\ 0 & \rho_2^* \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{bmatrix} \approx \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 0.552 & 0 \\ 0 & 0.49 \end{bmatrix} \\ \begin{bmatrix} 0.552 & 0 \\ 0 & 0.49 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{bmatrix}$$

(e)

Like mentioned in question `(d)`, the correlation within both **U** and **V** are 0. The correlation between **U** and **V** are described by the relevant canonical correlation.

Q3.

```
stiff = read.table('stiffness.DAT')
colnames(stiff) <- c("x1", "x2", "x3", "x4", "d")
```

(a)

```
X <- stiff %>%
  select(x1, x2)

Y <- stiff %>%
  select(x3, x4)

cancor(X, Y)
```

```
## $cor
## [1] 0.91291935 0.06805556
##
## $xcoef
##             [,1]         [,2]
## x1 -0.0006687933 -0.001237328
## x2  0.0001106253  0.001430402
##
## $ycoef
##             [,1]         [,2]
## x3 -0.0002497238  0.001573032
## x4 -0.0003515941 -0.001453802
##
## $xcenter
##       x1       x2
## 1906.100 1749.533
##
## $ycenter
##       x3       x4
## 1509.133 1724.967
```

(b)

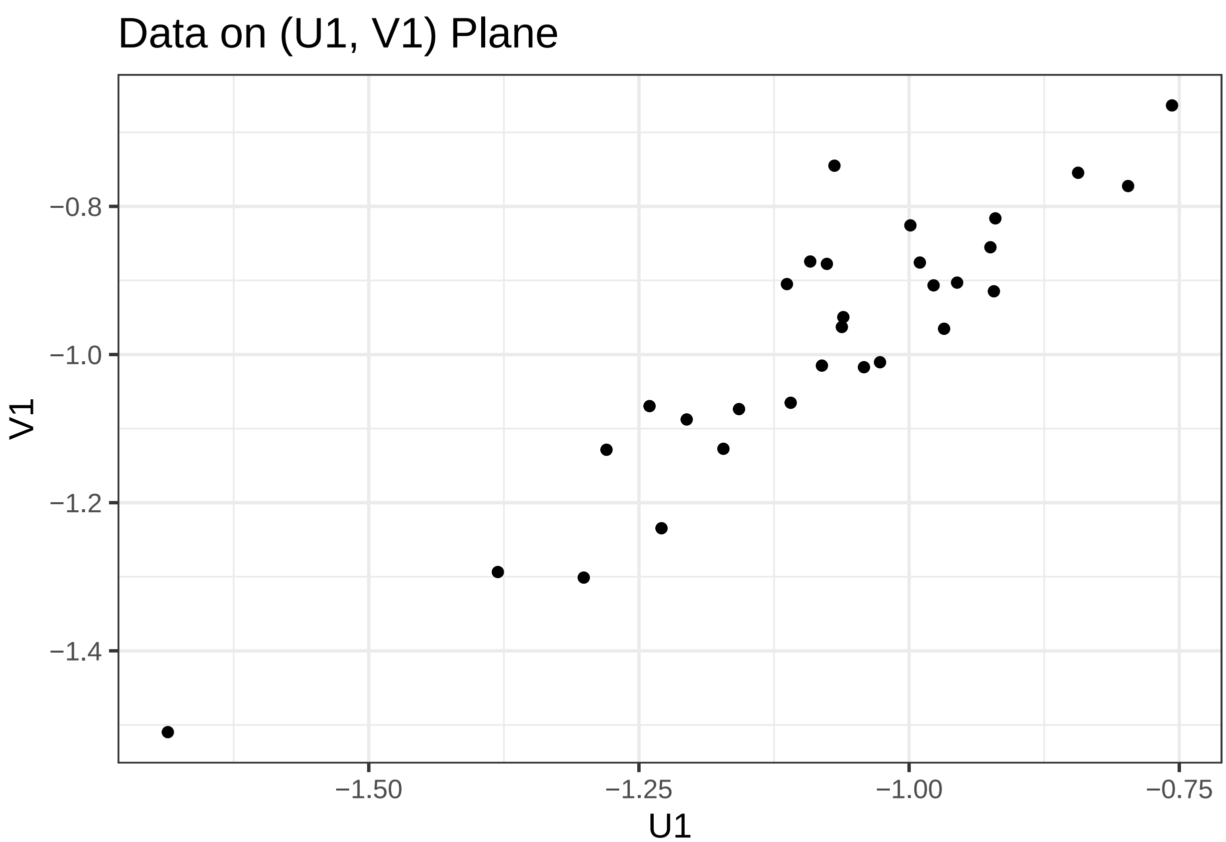
$$U_1 \approx \begin{pmatrix} -0.000669 \\ 0.000111 \end{pmatrix}^T \mathbf{X} + \begin{pmatrix} -0.000250 \\ -0.000352 \end{pmatrix}^T \mathbf{Y}$$

$$V_1 \approx \begin{pmatrix} -0.00124 \\ 0.00143 \end{pmatrix}^T \mathbf{X} + \begin{pmatrix} 0.00157 \\ -0.00145 \end{pmatrix}^T \mathbf{Y}$$

(c)

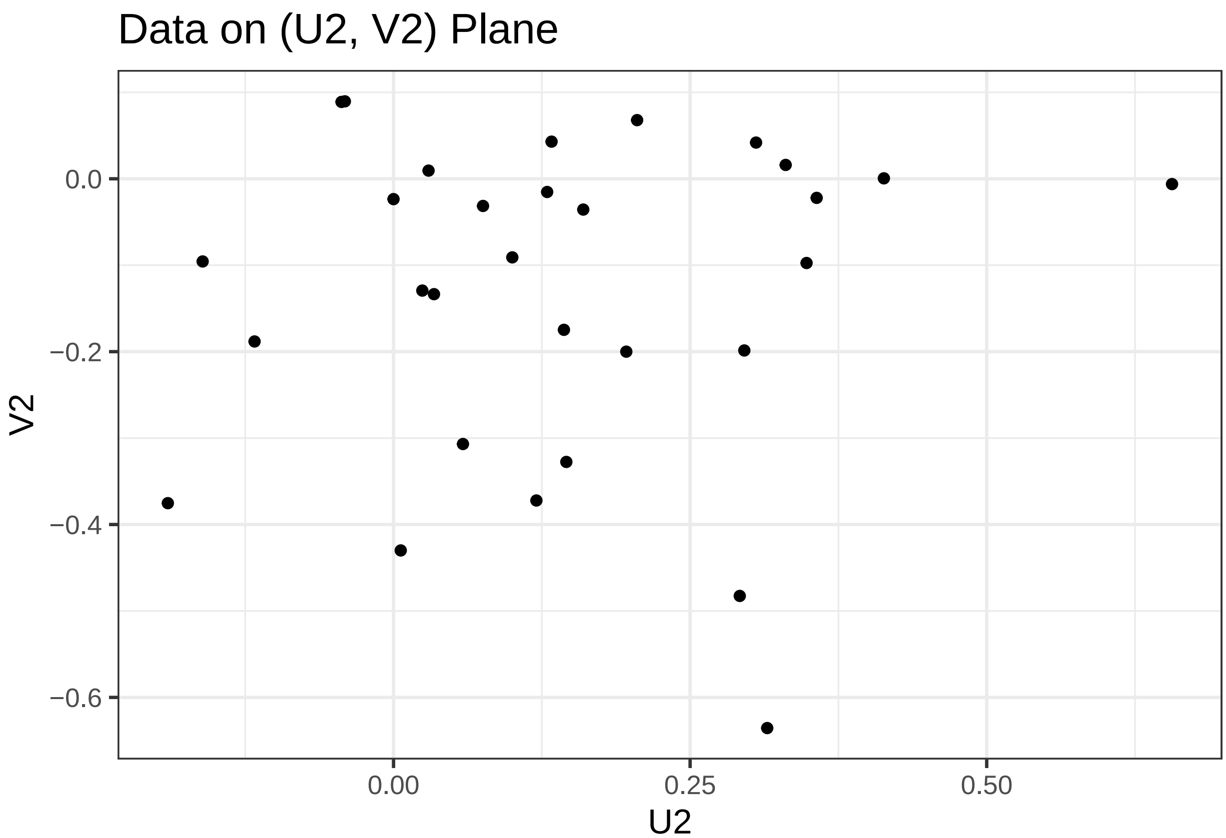
```
U_1 <- X$x1 * cancor(X, Y)$xcoef[1,1] + X$x2 * cancor(X, Y)$xcoef[2,1]
V_1 <- Y$x3 * cancor(X, Y)$ycoef[1,1] + Y$x4 * cancor(X, Y)$ycoef[2,1]

ggplot(tibble(u = U_1, v = V_1)) +
  geom_point(mapping = aes(x = u, y = v)) +
  xlab(expression(U1)) +
  ylab(expression(V1)) +
  labs(title = "Data on (U1, V1) Plane") +
  theme_bw(base_size = 13)
```




```
U_2 <- X$x1 * cancor(X, Y)$xcoef[1,2] + X$x2 * cancor(X, Y)$xcoef[2,2]
V_2 <- Y$x3 * cancor(X, Y)$ycoef[1,2] + Y$x4 * cancor(X, Y)$ycoef[2,2]

ggplot(tibble(u = U_2, v = V_2)) +
  geom_point(mapping = aes(x = u, y = v)) +
  xlab(expression(U2)) +
  ylab(expression(V2)) +
  labs(title = "Data on (U2, V2) Plane") +
  theme_bw(base_size = 13)
```



(d)

The plots and the canonical correlation values agree with each other in that (U_1, V_1) pair resembles a strong positive correlation. On the other hand, (U_2, V_2) pair resembles a weak correlation.

Q4.

```r
fly = read.table('fly.dat')
colnames(stiff) <- c("x1", "x2", "x3", "x4", "d")
```

(a)

i)

```r
# data cleaning for Af species
Af <- fly %>%
  filter(Species == "Af") %>%
  mutate(
    Y1 = Ant.Length + Wing.Length,
    Y2 = Wing.Length
  ) %>%
  select(4, 5)

Af_bar = colMeans(Af)

n1 = nrow(Af)

# data cleaning for Apf species
Apf <- fly %>%
  filter(Species == "Apf") %>%
  mutate(
    Y1 = Ant.Length + Wing.Length,
    Y2 = Wing.Length
  ) %>%
  select(4, 5)

Apf_bar = colMeans(Apf)

n2 = nrow(Apf)

# combining data
p = 2

diffmean = Af_bar - Apf_bar

S_pool = (n1 - 1)/(n1 + n2 - 2)*cov(Af) + (n2 - 1)/(n1 + n2 - 2)*cov(Apf)

T2 = t(diffmean) %*% solve((1/n1 + 1/n2)*S_pool)%*%diffmean

T2
```

```
##          [,1]
## [1,] 55.8807
```

```r
p_val <- 1 - pf((n1 + n2 - p - 1) * T2 / (p* (n1 + n2 - 2)),
                df1 = p, df2 = n1 + n2 - p - 1)

p_val
```

```
##              [,1]
## [1,] 4.519337e-05
```

The hypothesis of equality of the means will be rejected at significance levels $\alpha > 4.5193373 \times 10^{-5}$.

ii)

Yes.

If we let $\mathbf{y}_j$ be the jth observation where:

$$\mathbf{y}_j = \begin{pmatrix} y_{1,j} \\ y_{2,j} \end{pmatrix}$$

and similarly:

$$\mathbf{x}_j = \begin{pmatrix} x_{1,j} \\ x_{2,j} \end{pmatrix}$$

we know from the definition that:

$$\mathbf{y}_j = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \mathbf{x}_j$$

Here, we should note that $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is 2×2 and invertible:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

Therefore, as discussed in the proof of Question 1 (c), (Y_1, Y_2) should yield the same T^2-statistic as the original (X_1, X_2).

(b)

```r
fly <- fly %>%
  mutate(
    Y1 = Ant.Length + Wing.Length,
    Y2 = Wing.Length
  )
```

Y_1 and Y_2, by the given definition, are definitely not independent. With the independence assumption broken, we do not know how the univariate two-sample t-test would agree with the Hotelling's T^2-statistic.

Below is the results of the t-test

First for Y_1:

```r
# t-test for Y1
t1 <- t.test(Y1 ~ Species, data = fly)
t1
```

```
##
##  Welch Two Sample t-test
##
## data:  Y1 by Species
## t = 0.7132, df = 12.943, p-value = 0.4884
## alternative hypothesis: true difference in means between group Af and group Apf is not equal to 0
## 95 percent confidence interval:
##  -0.1308543  0.2597432
## sample estimates:
##  mean in group Af mean in group Apf
##          3.217778          3.153333
```

The t-test fails to reject the null-hypothesis (H_0) at both significance levels $\alpha = 0.05$ and $\alpha = 0.01$.

```r
# t-test for Y2
t2 <- t.test(Y2 ~ Species, data = fly)
t2
```

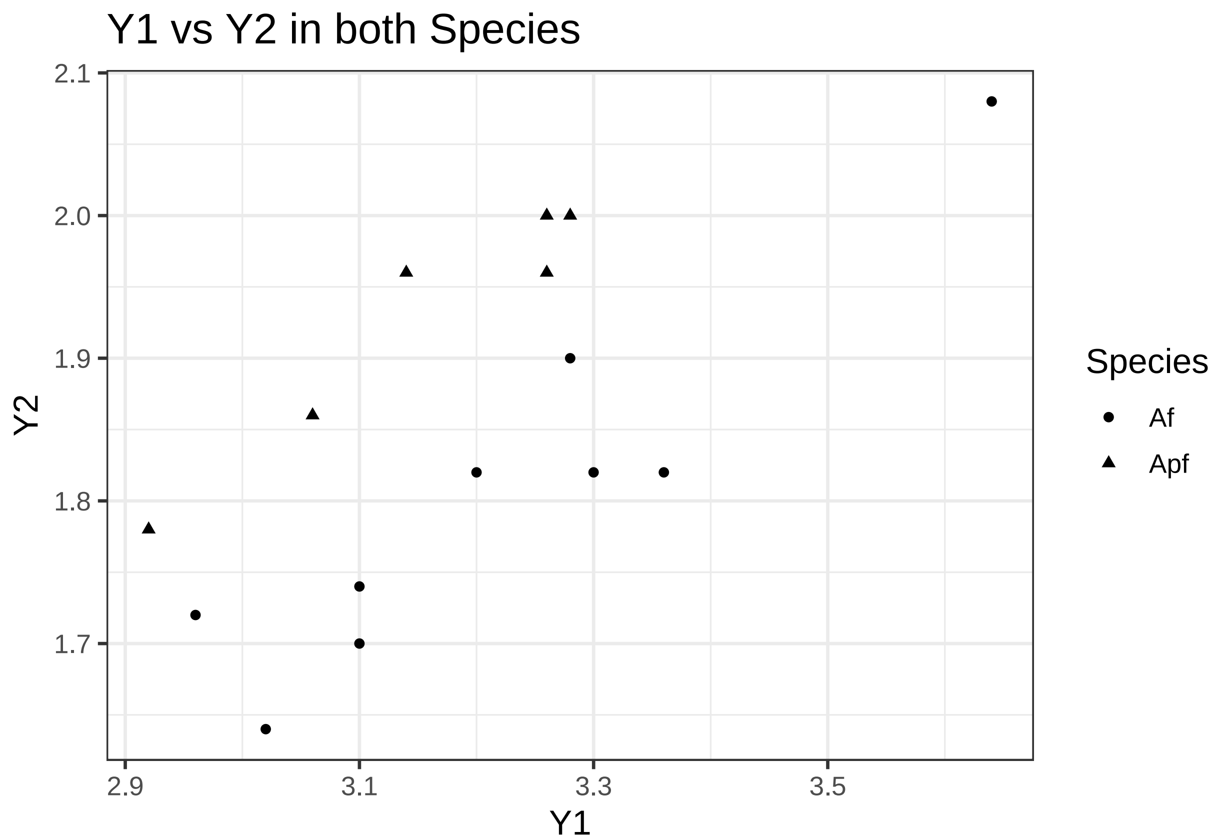
```
##
##  Welch Two Sample t-test
##
## data:  Y2 by Species
## t = -2.1697, df = 12.967, p-value = 0.0492
## alternative hypothesis: true difference in means between group Af and group Apf is not equal to 0
## 95 percent confidence interval:
##  -0.2439471978 -0.0004972466
## sample estimates:
##  mean in group Af mean in group Apf
##          1.804444          1.926667
```

The t-test rejects the null-hypothesis (H_0) at significance level $\alpha = 0.05$, but fails to reject at level $\alpha = 0.01$.

This is different from the Hotelling's T^2-statistic test, which rejects the null-hypothesis (H_0) at both significance levels.

(c)

```r
fly %>%
  ggplot() +
  geom_point(mapping = aes(x = Y1, y = Y2, shape = Species)) +
  xlab(expression(Y1)) +
  ylab(expression(Y2)) +
  labs(title = "Y1 vs Y2 in both Species") +
  theme_bw(base_size = 13)
```



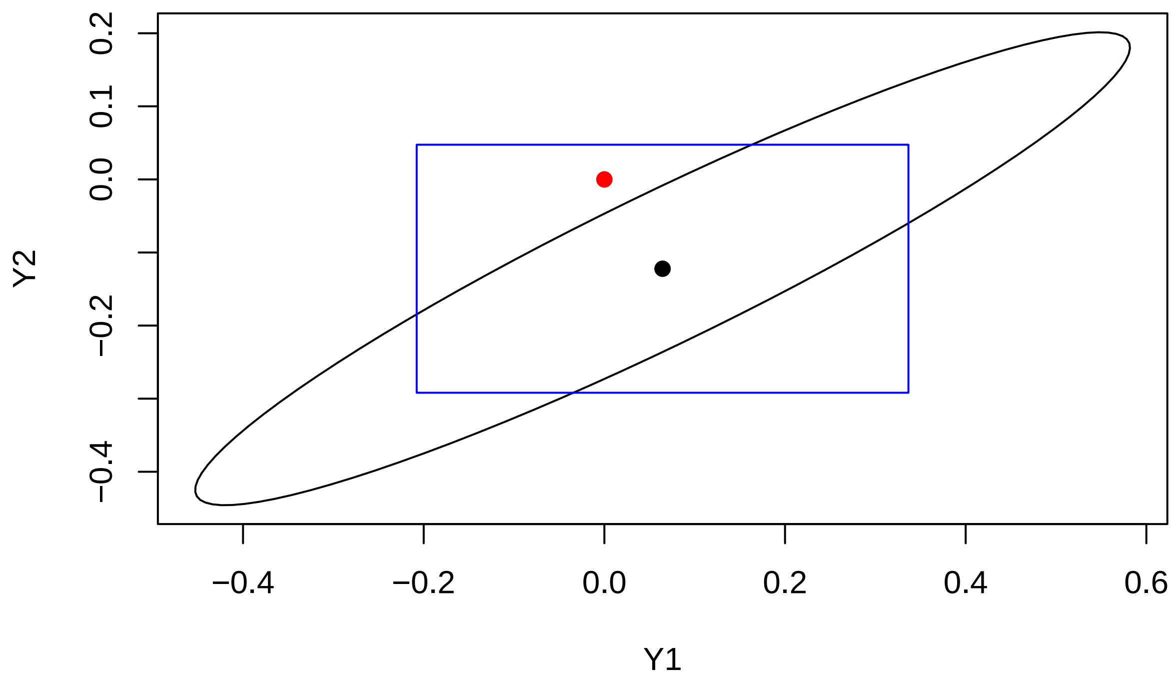
On the (Y_1, Y_2) plane, the scatter plot exhibits a clear division between the two species. However, on each axis, it is not clear that the two species have a clear difference in mean.

(d)

i) and ii)

```
# t-distribution
## Standard error
se_Y1 <- sqrt(var(Af$Y1) / n1 + var(Apf$Y1) / n2)
se_Y2 <- sqrt(var(Af$Y2) / n1 + var(Apf$Y2) / n2)
## critical value
t_crit <- qt(1 - 0.01/2, n1 + n2 - 2)
## lines for the rectangle
x_left <- diffmean[1] - t_crit * se_Y1
x_right <- diffmean[1] + t_crit * se_Y1
y_bottom <- diffmean[2] - t_crit * se_Y2
y_top <- diffmean[2] + t_crit * se_Y2


plot(ellipse(S_pool, level = 0.98, center = diffmean),
     type = "l", xlab = "Y1", ylab = "Y2")
rect(x_left, y_bottom, x_right, y_top, border = 'blue')
points(diffmean[1], diffmean[2], col = "black", pch = 19)
points(0, 0, col = "red", pch = 19)
```



iii)

Since there are two variables (Y_1 and Y_2), a Bonferroni method's adjusted significance level of $\alpha = 0.2$, which is linked with individual significance level of $\frac{0.2}{2} = 0.1$.

Another way to explain is that there is a 99 probability that Type-I error will not result from each of the t-tests for Y_1 and Y_2. This would mean that, assuming Type-I error independence, there is a $0.99 \times 0.99 = 0.9801 \approx 0.98$ probability that a Type-I error will not occur across the tests.

Both of the explanations align with the 98% confidence region by Bonferroni method.

iv)

The zero vector, shown as a red dot in the plot, is not included in the ellipse region; however, it is included in the rectangle region.

Given the correlation and, thus, no independence between Y_1 and Y_2, the ellipse region would be a better region to rule on a hypothesis test for the same significance level.

Q5.

The 2 x 1 random vectors X and Y have joint covariance matrix Σ,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

with

$$\Sigma_{11} = \begin{bmatrix} 1 & \rho_x \\ \rho_x & 1 \end{bmatrix},\ \Sigma_{22} = \begin{bmatrix} 1 & \rho_y \\ \rho_y & 1 \end{bmatrix},\ \Sigma_{12} = \Sigma_{21} = \begin{bmatrix} r & r \\ r & r \end{bmatrix}$$

(a)

If we let:

$$\mathbf{A} = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

then the square root of the largest eigen-value of $\mathbf{A}$ would be the largest canonical correlation between $\mathbf{X}$ and $\mathbf{Y}$.

$$\begin{aligned}
\mathbf{A} &= \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\
&= \begin{bmatrix} 1 & \rho_x \\ \rho_x & 1 \end{bmatrix}^{-1} \begin{bmatrix} r & r \\ r & r \end{bmatrix} \begin{bmatrix} 1 & \rho_y \\ \rho_y & 1 \end{bmatrix}^{-1} \begin{bmatrix} r & r \\ r & r \end{bmatrix} \\
&= \frac{1}{1-\rho_x^2} \begin{bmatrix} 1 & -\rho_x \\ -\rho_x & 1 \end{bmatrix} \begin{bmatrix} r & r \\ r & r \end{bmatrix} \frac{1}{1-\rho_y^2} \begin{bmatrix} 1 & -\rho_y \\ -\rho_y & 1 \end{bmatrix} \begin{bmatrix} r & r \\ r & r \end{bmatrix} \\
&= \frac{r - r\rho_x}{1-\rho_x^2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \frac{r - r\rho_y}{1-\rho_y^2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
&= \frac{r^2(1-\rho_x)(1-\rho_y)}{(1-\rho_x^2)(1-\rho_y^2)} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
&= \frac{2r^2}{(1+\rho_x)(1+\rho_y)} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}
\end{aligned}$$

If we let λ be the eigen-value of $\mathbf{A}$ then by definition

$$\begin{aligned}
&|\mathbf{A} - \lambda\mathbf{I}| = 0 \\
\Leftrightarrow &\left(\frac{2r^2}{(1+\rho_x)(1+\rho_y)} - \lambda\right)^2 - \left(\frac{2r^2}{(1+\rho_x)(1+\rho_y)}\right)^2 = 0 \\
\Leftrightarrow &\lambda\left(\lambda - \frac{4r^2}{(1+\rho_x)(1+\rho_y)}\right) = 0
\end{aligned}$$

Since $\lambda \neq 0$, there is only one eigen-value, and by default the largest eigen-value.

$$\therefore (\rho_1^*)^2 = \frac{4r^2}{(1+\rho_x)(1+\rho_y)}$$

In other words,

$$\rho_1^* = \frac{2r}{\sqrt{(1+\rho_x)(1+\rho_y)}}\ (\because \rho_1^* \geq 0)$$

(b)

Similar to the definition of $\mathbf{A}$, if we let $\mathbf{B}$ be:

$$\mathbf{B} = \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

We know that $\mathbf{B}$ would share the same eigen-value $(\rho_1^*)^2$ as $\mathbf{A}$.

If we let the corresponding eigen-vector of $\mathbf{A}$ to be e_1 and the corresponding eigen-vector of $\mathbf{B}$ to be f_1, then the $\mathbf{a}_1$ and $\mathbf{b}_1$ vectors that satisfy the normalization constraints would be:

$$\mathbf{a}_1 = \Sigma_{11}^{-1/2}e_1,\ \mathbf{b}_1 = \Sigma_{22}^{-1/2}f_1$$

First of all, we know that:

$$\begin{aligned}
\mathbf{B} &= \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\
&= \begin{bmatrix} 1 & \rho_y \\ \rho_y & 1 \end{bmatrix}^{-1} \begin{bmatrix} r & r \\ r & r \end{bmatrix} \begin{bmatrix} 1 & \rho_x \\ \rho_x & 1 \end{bmatrix}^{-1} \begin{bmatrix} r & r \\ r & r \end{bmatrix} \\
&= \frac{1}{1-\rho_y^2} \begin{bmatrix} 1 & -\rho_y \\ -\rho_y & 1 \end{bmatrix} \begin{bmatrix} r & r \\ r & r \end{bmatrix} \frac{1}{1-\rho_x^2} \begin{bmatrix} 1 & -\rho_x \\ -\rho_x & 1 \end{bmatrix} \begin{bmatrix} r & r \\ r & r \end{bmatrix} \\
&= \frac{r - r\rho_y}{1-\rho_y^2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \frac{r - r\rho_x}{1-\rho_x^2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
&= \frac{r^2(1-\rho_x)(1-\rho_y)}{(1-\rho_x^2)(1-\rho_y^2)} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
&= \frac{2r^2}{(1+\rho_x)(1+\rho_y)} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}
\end{aligned}$$

Which makes us conclude that $\mathbf{A} = \mathbf{B}$ and therefore $e_1 = f_1$.

Secondly,

$$\begin{aligned}
\Sigma_{11}^{-1/2} &= \begin{bmatrix} 1 & \rho_x \\ \rho_x & 1 \end{bmatrix}^{-1/2} \\
&= \left(\begin{bmatrix} 1 & \rho_x \\ \rho_x & 1 \end{bmatrix}^{1/2} \right)^{-1} \\
&= \left(\begin{bmatrix} \frac{\sqrt{1+\rho_x}+\sqrt{1-\rho_x}}{2} & \frac{\sqrt{1+\rho_x}-\sqrt{1-\rho_x}}{2} \\ \frac{\sqrt{1+\rho_x}-\sqrt{1-\rho_x}}{2} & \frac{\sqrt{1+\rho_x}+\sqrt{1-\rho_x}}{2} \end{bmatrix} \right)^{-1} \\
&= \begin{bmatrix} \frac{1}{2\sqrt{1+\rho_x}} + \frac{1}{2\sqrt{1-\rho_x}} & \frac{1}{2\sqrt{1+\rho_x}} - \frac{1}{2\sqrt{1-\rho_x}} \\ \frac{1}{2\sqrt{1+\rho_x}} - \frac{1}{2\sqrt{1-\rho_x}} & \frac{1}{2\sqrt{1+\rho_x}} + \frac{1}{2\sqrt{1-\rho_x}} \end{bmatrix}
\end{aligned}$$

Similarly,

$$\Sigma_{22}^{-1/2} = \begin{bmatrix} \frac{1}{2\sqrt{1+\rho_y}} + \frac{1}{2\sqrt{1-\rho_y}} & \frac{1}{2\sqrt{1+\rho_y}} - \frac{1}{2\sqrt{1-\rho_y}} \\ \frac{1}{2\sqrt{1+\rho_y}} - \frac{1}{2\sqrt{1-\rho_y}} & \frac{1}{2\sqrt{1+\rho_y}} + \frac{1}{2\sqrt{1-\rho_y}} \end{bmatrix}$$

Finally, to find the eigen-vector:

$$\frac{2r^2}{(1+\rho_x)(1+\rho_y)}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}e_1 = \frac{4r^2}{(1+\rho_x)(1+\rho_y)}e_1$$
$$\Leftrightarrow \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}e_1 = 2e_1$$
$$\therefore e_1 = f_1 = \begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}$$

Together, we are able to derive that

$$\begin{aligned}\mathbf{a}_1 =& \Sigma_{11}^{-1/2}e_1\\ =& \begin{bmatrix}\frac{1}{2\sqrt{1+\rho_x}}+\frac{1}{2\sqrt{1-\rho_x}} & \frac{1}{2\sqrt{1+\rho_x}}-\frac{1}{2\sqrt{1-\rho_x}}\\ \frac{1}{2\sqrt{1+\rho_x}}-\frac{1}{2\sqrt{1-\rho_x}} & \frac{1}{2\sqrt{1+\rho_x}}+\frac{1}{2\sqrt{1-\rho_x}}\end{bmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}\\ =& \begin{pmatrix}\frac{1}{\sqrt{2(1+\rho_x)}}\\ \frac{1}{\sqrt{2(1+\rho_x)}}\end{pmatrix}\end{aligned}$$

moreover

$$\begin{aligned}\mathbf{b}_1 =& \Sigma_{22}^{-1/2}e_1\\ =& \begin{bmatrix}\frac{1}{2\sqrt{1+\rho_y}}+\frac{1}{2\sqrt{1-\rho_y}} & \frac{1}{2\sqrt{1+\rho_y}}-\frac{1}{2\sqrt{1-\rho_y}}\\ \frac{1}{2\sqrt{1+\rho_y}}-\frac{1}{2\sqrt{1-\rho_y}} & \frac{1}{2\sqrt{1+\rho_y}}+\frac{1}{2\sqrt{1-\rho_y}}\end{bmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}\\ =& \begin{pmatrix}\frac{1}{\sqrt{2(1+\rho_y)}}\\ \frac{1}{\sqrt{2(1+\rho_y)}}\end{pmatrix}\end{aligned}$$

Ultimately, the canonical variate paris are:

$$U_1 = \begin{pmatrix}\frac{1}{\sqrt{2(1+\rho_x)}}\\ \frac{1}{\sqrt{2(1+\rho_x)}}\end{pmatrix}^T \mathbf{X},$$
$$V_1 = \begin{pmatrix}\frac{1}{\sqrt{2(1+\rho_y)}}\\ \frac{1}{\sqrt{2(1+\rho_y)}}\end{pmatrix}^T \mathbf{Y}$$

Q6.

(a)

$$
\begin{aligned}
\mathbf{S} &= \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})^T \\
&= \frac{1}{n-1}\sum_{j=1}^{n}\begin{pmatrix} (x_{j,1}-\bar{x}_1)^2 & (x_{j,1}-\bar{x}_1)(x_{j,2}-\bar{x}_2) & \cdots & (x_{j,1}-\bar{x}_1)(x_{j,p}-\bar{x}_p) \\ (x_{j,1}-\bar{x}_1)(x_{j,2}-\bar{x}_2) & (x_{j,2}-\bar{x}_2)^2 & & \vdots \\ \vdots & & \ddots & \\ (x_{j,1}-\bar{x}_1)(x_{j,p}-\bar{x}_p) & \cdots & & (x_{j,p}-\bar{x}_p)^2 \end{pmatrix} \\
&= \frac{1}{n-1}\begin{pmatrix} \sum_{j=1}^{n}(x_{j,1}-\bar{x}_1)^2 & \sum_{j=1}^{n}(x_{j,1}-\bar{x}_1)(x_{j,2}-\bar{x}_2) & \cdots & \sum_{j=1}^{n}(x_{j,1}-\bar{x}_1)(x_{j,p}-\bar{x}_p) \\ \sum_{j=1}^{n}(x_{j,1}-\bar{x}_1)(x_{j,2}-\bar{x}_2) & \sum_{j=1}^{n}(x_{j,2}-\bar{x}_2)^2 & & \vdots \\ \vdots & & \ddots & \\ \sum_{j=1}^{n}(x_{j,1}-\bar{x}_1)(x_{j,p}-\bar{x}_p) & \cdots & & \sum_{j=1}^{n}(x_{j,p}-\bar{x}_p)^2 \end{pmatrix} \\
&= \frac{1}{n-1}\begin{pmatrix} \sum_{j=1}^{n}(x_{j,1}^2)-n\bar{x}_1^2 & \sum_{j=1}^{n}(x_{j,1}\cdot x_{j,2})-n\bar{x}_1\bar{x}_2 & \cdots & \sum_{j=1}^{n}(x_{j,1}\cdot x_{j,p})-n\bar{x}_1\bar{x}_p \\ \sum_{j=1}^{n}(x_{j,1}\cdot x_{j,2})-n\bar{x}_1\bar{x}_2 & \sum_{j=1}^{n}(x_{j,2}^2)-n\bar{x}_2^2 & & \vdots \\ \vdots & & \ddots & \\ \sum_{j=1}^{n}(x_{j,1}\cdot x_{j,p})-n\bar{x}_1\bar{x}_p & \cdots & & \sum_{j=1}^{n}(x_{j,p}^2)-n\bar{x}_p^2 \end{pmatrix} \\
&= \frac{1}{n-1}\begin{pmatrix} \sum_{j=1}^{n}(x_{j,1}^2) & \sum_{j=1}^{n}(x_{j,1}\cdot x_{j,2}) & \cdots & \sum_{j=1}^{n}(x_{j,1}\cdot x_{j,p}) \\ \sum_{j=1}^{n}(x_{j,1}\cdot x_{j,2}) & \sum_{j=1}^{n}(x_{j,2}^2) & & \vdots \\ \vdots & & \ddots & \\ \sum_{j=1}^{n}(x_{j,1}\cdot x_{j,p}) & \cdots & & \sum_{j=1}^{n}(x_{j,p}^2) \end{pmatrix} \\
&\quad - \frac{1}{n-1}\begin{pmatrix} n\bar{x}_1^2 & n\bar{x}_1\bar{x}_2\cdots & n\bar{x}_1\bar{x}_p & \\ n\bar{x}_1\bar{x}_2 & n\bar{x}_2^2 & & \vdots \\ \vdots & & \ddots & \\ n\bar{x}_1\bar{x}_p & \cdots & & n\bar{x}_p^2 \end{pmatrix} \\
&= \frac{1}{n-1}\mathbf{X}'\mathbf{X} - \frac{1}{n-1}\cdot\frac{1}{n}\begin{pmatrix} (\sum_{j=1}^{n}x_{j,1})^2 & (\sum_{j=1}^{n}x_{j,1})(\sum_{j=1}^{n}x_{j,2}) & \cdots & (\sum_{j=1}^{n}x_{j,1})(\sum_{j=1}^{n}x_{j,p}) \\ (\sum_{j=1}^{n}x_{j,1})(\sum_{j=1}^{n}x_{j,2}) & (\sum_{j=1}^{n}x_{j,2})^2 & & \vdots \\ \vdots & & \ddots & \\ (\sum_{j=1}^{n}x_{j,1})(\sum_{j=1}^{n}x_{j,p}) & \cdots & & (\sum_{j=1}^{n}x_{j,p})^2 \end{pmatrix} \\
&= \frac{1}{n-1}\mathbf{X}'\mathbf{X} - \frac{1}{n-1}\cdot\frac{1}{n}\begin{pmatrix} \sum_{j=1}^{n}x_{j,1} \\ \sum_{j=1}^{n}x_{j,2} \\ \vdots \\ \sum_{j=1}^{n}x_{j,p} \end{pmatrix}\begin{pmatrix} \sum_{j=1}^{n}x_{j,1} \\ \sum_{j=1}^{n}x_{j,2} \\ \vdots \\ \sum_{j=1}^{n}x_{j,p} \end{pmatrix}^T \\
&= \frac{1}{n-1}\mathbf{X}'\mathbf{X} - \frac{1}{n-1}\cdot\frac{1}{n}(\mathbf{X}'\mathbf{1}_n)(\mathbf{X}'\mathbf{1}_n)^T \\
&= \frac{1}{n-1}\mathbf{X}'\mathbf{X} - \frac{1}{n-1}\cdot\frac{1}{n}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X} \\
&= \frac{1}{n-1}\mathbf{X}'\mathbf{I}_n\mathbf{X} - \frac{1}{n-1}\mathbf{X}'\frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\mathbf{X} \\
&= \frac{1}{n-1}\mathbf{X}'\left(\mathbf{I}_n\mathbf{X} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\right) = \frac{1}{n-1}\mathbf{X}'\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\mathbf{X}
\end{aligned}
$$

(b)

$$\begin{aligned}\mathbf{W} &= \mathbf{AY} + c \\ &= \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kp} \end{bmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} + \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} \\ &= \begin{pmatrix} c_1 + \sum_{i=1}^p a_{1,i} y_i \\ \vdots \\ c_k + \sum_{i=1}^p a_{k,i} y_i \end{pmatrix}\end{aligned}$$

Now if we analyze $\mathbf{W}$ by it's row components:

$$\begin{aligned}Var(W_j) =& Var(c_j + \sum_{i=1}^p a_{j,i} y_i) = \sum_{i=1}^p a_{j,i}^2 Var(y_i) + \sum_{i=1}^p \sum_{k \neq i} a_{j,i} a_{j,k} Cov(y_i, y_k) \\ =& \sum_{i=1}^p \sum_{k=1}^p a_{1,i} a_{1,k} Cov(y_i, y_k)\end{aligned}$$

$$\begin{aligned}Cov(W_j, W_k) =& Cov(c_j + \sum_{i=1}^p a_{j,i} y_i,\ c_k + \sum_{i=1}^p a_{k,i} y_i) \\ =& Cov(\sum_{i=1}^p a_{j,i} y_i,\ \sum_{i=1}^p a_{k,i} y_i) \\ =& \sum_{l=1}^p a_{j,l} Cov(y_l,\ \sum_{i=1}^p a_{k,i} y_i) \\ =& \sum_{l=1}^p a_{j,l} \sum_{m=1}^p a_{k,m} Cov(y_l,\ y_m) \\ =& \sum_{l=1}^p \sum_{m=1}^p a_{j,l} a_{k,m} Cov(y_l,\ y_m)\end{aligned}$$

If we put them together, the covariance matrix would look like:

$$\begin{aligned}Cov(\mathbf{W}) &= \begin{bmatrix} \sum_{i=1}^p \sum_{j=1}^p a_{1,i} a_{1,j} Cov(y_i, y_j) & \sum_{i=1}^p \sum_{j=1}^p a_{1,i} a_{2,j} Cov(y_i, y_j) & \cdots & \sum_{i=1}^p \sum_{j=1}^p a_{1,i} a_{k,j} Cov(y_i, y_j) \\ \sum_{i=1}^p \sum_{j=1}^p a_{1,i} a_{2,j} Cov(y_i, y_j) & \sum_{i=1}^p \sum_{j=1}^p a_{2,i} a_{2,j} Cov(y_i, y_j) & & \vdots \\ \vdots & & \ddots & \\ \sum_{i=1}^p \sum_{j=1}^p a_{1,i} a_{k,j} Cov(y_i, y_j) & \cdots & & \sum_{i=1}^p \sum_{j=1}^p a_{k,i} a_{k,j} Cov(y_i, y_j) \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^p a_{1,i} \sum_{j=1}^p a_{1,j} Cov(y_i, y_j) & \sum_{i=1}^p a_{1,i} \sum_{j=1}^p a_{2,j} Cov(y_i, y_j) & \cdots & \sum_{i=1}^p a_{1,i} \sum_{j=1}^p a_{k,j} Cov(y_i, y_j) \\ \sum_{i=1}^p a_{2,i} \sum_{j=1}^p a_{1,j} Cov(y_i, y_j) & \sum_{i=1}^p a_{2,i} \sum_{j=1}^p a_{2,j} Cov(y_i, y_j) & & \vdots \\ \vdots & & \ddots & \\ \sum_{i=1}^p a_{k,i} \sum_{j=1}^p a_{1,j} Cov(y_i, y_j) & \cdots & & \sum_{i=1}^p a_{k,i} \sum_{j=1}^p a_{k,j} Cov(y_i, y_j) \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{k1} & \cdots & & a_{kk} \end{bmatrix} \begin{bmatrix} \sum_{j=1}^p a_{1,j} Cov(y_1, y_j) & \sum_{j=1}^p a_{2,j} Cov(y_1, y_j) & \cdots & \sum_{j=1}^p a_{k,j} Cov(y_1, y_j) \\ \sum_{j=1}^p a_{1,j} Cov(y_2, y_j) & \sum_{j=1}^p a_{2,j} Cov(y_2, y_j) & & \vdots \\ \vdots & & \ddots & \\ \sum_{j=1}^p a_{1,j} Cov(y_k, y_j) & \cdots & & \sum_{j=1}^p a_{k,j} Cov(y_k, y_j) \end{bmatrix}\end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{k1} & \cdots & & a_{kk} \end{bmatrix} \left(\begin{bmatrix} Cov(y_1,y_1) & Cov(y_1,y_2) & \cdots & Cov(y_1,y_k) \\ Cov(y_2,y_1) & Cov(y_2,y_2) & & \vdots \\ \vdots & & \ddots & \\ Cov(y_k,y_1) & \cdots & & Cov(y_k,y_k) \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{1k} & \cdots & & a_{kk} \end{bmatrix} \right) \\
&= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{k1} & \cdots & & a_{kk} \end{bmatrix} \begin{bmatrix} Cov(y_1,y_1) & Cov(y_1,y_2) & \cdots & Cov(y_1,y_k) \\ Cov(y_2,y_1) & Cov(y_2,y_2) & & \vdots \\ \vdots & & \ddots & \\ Cov(y_k,y_1) & \cdots & & Cov(y_k,y_k) \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{1k} & \cdots & & a_{kk} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{k1} & \cdots & & a_{kk} \end{bmatrix} \begin{bmatrix} Cov(y_1,y_1) & Cov(y_1,y_2) & \cdots & Cov(y_1,y_k) \\ Cov(y_2,y_1) & Cov(y_2,y_2) & & \vdots \\ \vdots & & \ddots & \\ Cov(y_k,y_1) & \cdots & & Cov(y_k,y_k) \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{k1} & \cdots & & a_{kk} \end{bmatrix}^T \\
&= \mathbf{A}\ Cov(\mathbf{Y})\ \mathbf{A}'
\end{aligned}
$$

Q.E.D.

(c)

First let:

$$Cov(\mathbf{Y}, \mathbf{W}) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Here, Σ_{11} would be a $p \times p$ matrix and Σ_{22} would be a $q \times q$ matrix.

Moreover, $\Sigma_{12} = \Sigma_{21}'$ and Σ_{12} is a $p \times q$ matrix.

The specifics of each entry is below:

$$
\begin{aligned}
\Sigma_{11} &= \begin{bmatrix} Cov(Y_1,Y_1) & \cdots & Cov(Y_1,Y_p) \\ \vdots & \ddots & \vdots \\ Cov(Y_p,Y_1) & \cdots & Cov(Y_p,Y_p) \end{bmatrix},\ \Sigma_{22} = \begin{bmatrix} Cov(W_1,W_1) & \cdots & Cov(W_1,W_q) \\ \vdots & \ddots & \vdots \\ Cov(W_q,W_1) & \cdots & Cov(W_q,W_q) \end{bmatrix}, \\
\Sigma_{12} &= \begin{bmatrix} Cov(Y_1,W_1) & \cdots & Cov(Y_1,W_q) \\ \vdots & \ddots & \vdots \\ Cov(Y_p,W_1) & \cdots & Cov(Y_p,W_q) \end{bmatrix}
\end{aligned}
$$

Now if we let:

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix},\ \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_q \end{pmatrix}$$

Then we can now express the covariance as:

$$\begin{aligned}
Cov(\mathbf{a}'\mathbf{Y}, \mathbf{b}'\mathbf{W}) &= Cov(\sum_{i=1}^{p} a_i Y_i, \sum_{j=1}^{q} b_i W_j) \\
&= \sum_{i=1}^{p} a_i Cov(Y_i, \sum_{j=1}^{q} b_i W_j) \\
&= \sum_{i=1}^{p} a_i \sum_{j=1}^{q} b_i Cov(Y_i, W_j) \\
&= \sum_{i=1}^{p} a_i \begin{pmatrix} Cov(Y_i, W_1) \\ Cov(Y_i, W_2) \\ \vdots \\ Cov(Y_i, W_q) \end{pmatrix}^T \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_q \end{pmatrix} \\
&= \left(\sum_{i=1}^{p} a_i \begin{pmatrix} Cov(Y_i, W_1) \\ Cov(Y_i, W_2) \\ \vdots \\ Cov(Y_i, W_q) \end{pmatrix}^T \right) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_q \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{p} a_i Cov(Y_i, W_1) & \cdots & \sum_{i=1}^{p} a_i Cov(Y_i, W_q) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_q \end{pmatrix} \\
&= \begin{pmatrix} a_1 & \cdots & a_p \end{pmatrix} \begin{bmatrix} Cov(Y_1, W_1) & \cdots & Cov(Y_1, W_q) \\ \vdots & \ddots & \vdots \\ Cov(Y_p, W_1) & \cdots & Cov(Y_p, W_q) \end{bmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_q \end{pmatrix} \\
&= \mathbf{a}'\Sigma_{12}\mathbf{b}
\end{aligned}$$

Another way to look at it is:

$$\begin{aligned}
Cov(\mathbf{a}'\mathbf{Y}, \mathbf{b}'\mathbf{W}) &= Cov(\sum_{i=1}^{p} a_i Y_i, \sum_{j=1}^{q} b_i W_j) \\
&= \sum_{j=1}^{q} b_i Cov(\sum_{i=1}^{p} a_i Y_i, W_j) \\
&= \sum_{j=1}^{q} b_i \sum_{i=1}^{p} a_i Cov(Y_i, W_j) \\
&= \sum_{j=1}^{q} b_j \begin{pmatrix} Cov(Y_1, W_j) \\ Cov(Y_2, W_j) \\ \vdots \\ Cov(Y_p, W_j) \end{pmatrix}^T \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \\
&= \left(\sum_{j=1}^{q} b_j \begin{pmatrix} Cov(Y_1, W_j) \\ Cov(Y_2, W_j) \\ \vdots \\ Cov(Y_p, W_j) \end{pmatrix}^T \right) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^{q} b_j Cov(Y_1, W_j) & \cdots & \sum_{j=1}^{q} b_j Cov(Y_p, W_j) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \\
&= \begin{pmatrix} b_1 & \cdots & b_q \end{pmatrix} \begin{bmatrix} Cov(Y_1, W_1) & \cdots & Cov(Y_p, W_1) \\ \vdots & \ddots & \vdots \\ Cov(Y_1, W_p) & \cdots & Cov(Y_p, W_q) \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \\
&= \mathbf{b}'\Sigma_{21}\mathbf{a}
\end{aligned}$$

$$Q.E.D.$$