

# STAT 32950 Assignment 3

Bin Yu

Apr 14, 2025

## Question 1

(a)

(i)

$$\bar{x} = \frac{1}{4} \sum_{j=1}^4 x_j = \begin{pmatrix} \frac{2+8+6+8}{4} \\ \frac{12+9+9+10}{4} \end{pmatrix} = \begin{pmatrix} \frac{24}{4} \\ \frac{40}{4} \end{pmatrix} = \begin{pmatrix} 6 \\ 10 \end{pmatrix}.$$

(ii)

the sample covariance matrix is given by

$$S = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T,$$

We are given the observations:

$$x_1 = \begin{pmatrix} 2 \\ 12 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 8 \\ 9 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 6 \\ 9 \end{pmatrix}, \quad x_4 = \begin{pmatrix} 8 \\ 10 \end{pmatrix},$$

and the sample mean vector as

$$\bar{x} = \begin{pmatrix} 6 \\ 10 \end{pmatrix}.$$

$$x_1 - \bar{x} = \begin{pmatrix} 2 - 6 \\ 12 - 10 \end{pmatrix} = \begin{pmatrix} -4 \\ 2 \end{pmatrix},$$

$$x_2 - \bar{x} = \begin{pmatrix} 8 - 6 \\ 9 - 10 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix},$$

$$x_3 - \bar{x} = \begin{pmatrix} 6 - 6 \\ 9 - 10 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

$$x_4 - \bar{x} = \begin{pmatrix} 8 - 6 \\ 10 - 10 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

For each deviation  $x_j - \bar{x}$  (a column vector), compute the outer product with its transpose:

$$(x_1 - \bar{x})(x_1 - \bar{x})^T = \begin{pmatrix} -4 \\ 2 \end{pmatrix} \begin{pmatrix} -4 & 2 \end{pmatrix} = \begin{pmatrix} (-4)(-4) & (-4)(2) \\ (2)(-4) & (2)(2) \end{pmatrix} = \begin{pmatrix} 16 & -8 \\ -8 & 4 \end{pmatrix},$$

$$(x_2 - \bar{x})(x_2 - \bar{x})^T = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \begin{pmatrix} 2 & -1 \end{pmatrix} = \begin{pmatrix} 2 \cdot 2 & 2 \cdot (-1) \\ (-1) \cdot 2 & (-1) \cdot (-1) \end{pmatrix} = \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix},$$

$$(x_3 - \bar{x})(x_3 - \bar{x})^T = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \begin{pmatrix} 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 \cdot 0 & 0 \cdot (-1) \\ (-1) \cdot 0 & (-1) \cdot (-1) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

$$(x_4 - \bar{x})(x_4 - \bar{x})^T = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \end{pmatrix} = \begin{pmatrix} 2 \cdot 2 & 2 \cdot 0 \\ 0 \cdot 2 & 0 \cdot 0 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus

$$\sum_{j=1}^4 (x_j - \bar{x})(x_j - \bar{x})^T = \begin{pmatrix} 16 & -8 \\ -8 & 4 \end{pmatrix} + \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}.$$

$$(1,1) \text{ entry: } 16 + 4 + 0 + 4 = 24,$$

$$(1,2) \text{ entry: } -8 + (-2) + 0 + 0 = -10,$$

$$(2,1) \text{ entry: } -8 + (-2) + 0 + 0 = -10,$$

$$(2,2) \text{ entry: } 4 + 1 + 1 + 0 = 6.$$

$$\sum_{j=1}^4 (x_j - \bar{x})(x_j - \bar{x})^T = \begin{pmatrix} 24 & -10 \\ -10 & 6 \end{pmatrix}.$$

Therefore, the sample covariance matrix is

$$S = \frac{1}{3} \begin{pmatrix} 24 & -10 \\ -10 & 6 \end{pmatrix} = \begin{pmatrix} 8 & -\frac{10}{3} \\ -\frac{10}{3} & 2 \end{pmatrix}.$$

(iii)

For a  $2 \times 2$  matrix

$$S = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

the inverse is given by

$$S^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Here,

$$a = 8, \quad b = -\frac{10}{3}, \quad c = -\frac{10}{3}, \quad d = 2.$$

$$\det(S) = 8 \cdot 2 - \left(-\frac{10}{3}\right) \left(-\frac{10}{3}\right) = 16 - \frac{100}{9} = \frac{144 - 100}{9} = \frac{44}{9}.$$

$$S^{-1} = \frac{1}{\frac{44}{9}} \begin{pmatrix} 2 & \frac{10}{3} \\ \frac{10}{3} & 8 \end{pmatrix} = \frac{9}{44} \begin{pmatrix} 2 & \frac{10}{3} \\ \frac{10}{3} & 8 \end{pmatrix}.$$

Thus,

$$S^{-1} = \begin{pmatrix} \frac{9}{22} & \frac{15}{22} \\ \frac{15}{22} & \frac{18}{11} \end{pmatrix} = \frac{1}{22} \begin{pmatrix} 9 & 15 \\ 15 & 36 \end{pmatrix}.$$

(iv)

Hotelling's statistic is defined by

$$T^2 = n(\bar{x} - \mu_0)^\top S^{-1}(\bar{x} - \mu_0),$$

We have

$$\bar{x} - \mu_0 = \begin{pmatrix} 6 \\ 10 \end{pmatrix} - \begin{pmatrix} 7 \\ 11 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

Then,

$$(\bar{x} - \mu_0)^\top S^{-1}(\bar{x} - \mu_0) = (-1 \quad -1) \left( \frac{1}{22} \begin{pmatrix} 9 & 15 \\ 15 & 36 \end{pmatrix} \right) \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

$$\begin{pmatrix} 9 & 15 \\ 15 & 36 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -9 - 15 \\ -15 - 36 \end{pmatrix} = \begin{pmatrix} -24 \\ -51 \end{pmatrix}.$$

$$(\bar{x} - \mu_0)^\top S^{-1}(\bar{x} - \mu_0) = \frac{1}{22} (-1 \quad -1) \begin{pmatrix} -24 \\ -51 \end{pmatrix} = \frac{1}{22} (24 + 51) = \frac{75}{22}.$$

Therefore,

$$T^2 = 4 \times \frac{75}{22} = \frac{150}{11}.$$

(v)

Under the null hypothesis  $H_0 : \mu = \mu_0$  and assuming a multivariate normal distribution, the statistic follows:

$$T^2 \sim \frac{(n-1)p}{n-p} F_{p, n-p}$$

Here,  $p = 2$  and  $n = 4$ , so

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

(b)

(i)

$$x_1 = \begin{pmatrix} 2 \\ 12 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 8 \\ 9 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 6 \\ 9 \end{pmatrix}, \quad x_4 = \begin{pmatrix} 8 \\ 10 \end{pmatrix}.$$

Applying  $C$  to each  $x_j$ :

$$y_j = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_{j1} \\ x_{j2} \end{pmatrix}.$$

$$y_1 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 12 \end{pmatrix} = \begin{pmatrix} 14 \\ 10 \end{pmatrix},$$

$$y_2 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 8 \\ 9 \end{pmatrix} = \begin{pmatrix} 17 \\ 1 \end{pmatrix},$$

$$y_3 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 6 \\ 9 \end{pmatrix} = \begin{pmatrix} 15 \\ 3 \end{pmatrix},$$

$$y_4 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 8 \\ 10 \end{pmatrix} = \begin{pmatrix} 18 \\ 2 \end{pmatrix}.$$

Thus

$$y_1 = \begin{pmatrix} 14 \\ 10 \end{pmatrix}, \quad y_2 = \begin{pmatrix} 17 \\ 1 \end{pmatrix}, \quad y_3 = \begin{pmatrix} 15 \\ 3 \end{pmatrix}, \quad y_4 = \begin{pmatrix} 18 \\ 2 \end{pmatrix}.$$

The sample mean vector of the  $\{y_j\}$  is

$$\bar{y} = \frac{1}{4} \sum_{j=1}^4 y_j = \frac{1}{4} \begin{pmatrix} 14 + 17 + 15 + 18 \\ 10 + 1 + 3 + 2 \end{pmatrix} = \begin{pmatrix} \frac{64}{4} \\ \frac{16}{4} \end{pmatrix} = \begin{pmatrix} 16 \\ 4 \end{pmatrix}.$$

(ii)

From part (a), the sample covariance matrix of  $\{x_j\}$  is

$$S = \begin{pmatrix} 8 & -\frac{10}{3} \\ -\frac{10}{3} & 2 \end{pmatrix}.$$

and,

$$S_y = C S C^T,$$

where

$$C = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad C^T = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

$$C S = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 8 & -\frac{10}{3} \\ -\frac{10}{3} & 2 \end{pmatrix} = \begin{pmatrix} 1 \cdot 8 + 1 \cdot (-\frac{10}{3}) & 1 \cdot (-\frac{10}{3}) + 1 \cdot 2 \\ (-1) \cdot 8 + 1 \cdot (-\frac{10}{3}) & (-1) \cdot (-\frac{10}{3}) + 1 \cdot 2 \end{pmatrix}.$$

$$= \begin{pmatrix} \frac{14}{3} & -\frac{4}{3} \\ -\frac{34}{3} & \frac{16}{3} \end{pmatrix}.$$

$$C S C^T = \begin{pmatrix} \frac{14}{3} & -\frac{4}{3} \\ -\frac{34}{3} & \frac{16}{3} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{14}{3} \cdot 1 + (-\frac{4}{3}) \cdot 1 & \frac{14}{3} \cdot (-1) + (-\frac{4}{3}) \cdot 1 \\ (-\frac{34}{3}) \cdot 1 + \frac{16}{3} \cdot 1 & (-\frac{34}{3}) \cdot (-1) + \frac{16}{3} \cdot 1 \end{pmatrix}.$$

$$S_y = \begin{pmatrix} \frac{14}{3} - \frac{4}{3} & -\frac{14}{3} - \frac{4}{3} \\ -\frac{34}{3} + \frac{16}{3} & \frac{34}{3} + \frac{16}{3} \end{pmatrix} = \begin{pmatrix} \frac{10}{3} & -\frac{18}{3} \\ -\frac{18}{3} & \frac{50}{3} \end{pmatrix} = \begin{pmatrix} \frac{10}{3} & -6 \\ -6 & \frac{50}{3} \end{pmatrix}.$$

(iii)

Under the null hypothesis  $H_0 : \mu_y = \mu_o^*$  where  $\mu_o^* = \begin{pmatrix} 18 \\ 4 \end{pmatrix}$ , the Hotelling's  $T^2$  statistic for the transformed data is

$$T_y^2 = n (\bar{y} - \mu_o^*)^T S_y^{-1} (\bar{y} - \mu_o^*),$$

with  $n = 4$ . We already have

$$\bar{y} = \begin{pmatrix} 16 \\ 4 \end{pmatrix}, \quad \mu_o^* = \begin{pmatrix} 18 \\ 4 \end{pmatrix}.$$

Hence

$$\bar{y} - \mu_o^* = \begin{pmatrix} 16 - 18 \\ 4 - 4 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

and we have:

$$S_y = \begin{pmatrix} \frac{10}{3} & -6 \\ -6 & \frac{50}{3} \end{pmatrix}.$$

$$\det(S_y) = \left(\frac{10}{3}\right) \left(\frac{50}{3}\right) - (-6) \cdot (-6) = \frac{500}{9} - 36 = \frac{500}{9} - \frac{324}{9} = \frac{176}{9}.$$

By the formula for the inverse matrix,

$$S_y^{-1} = \frac{1}{\frac{176}{9}} \begin{pmatrix} \frac{50}{3} & 6 \\ 6 & \frac{10}{3} \end{pmatrix} = \frac{9}{176} \begin{pmatrix} \frac{50}{3} & 6 \\ 6 & \frac{10}{3} \end{pmatrix} = \begin{pmatrix} \frac{9}{176} \cdot \frac{50}{3} & \frac{9}{176} \cdot 6 \\ \frac{9}{176} \cdot 6 & \frac{9}{176} \cdot \frac{10}{3} \end{pmatrix}.$$

$$S_y^{-1} = \begin{pmatrix} \frac{75}{88} & \frac{27}{88} \\ \frac{27}{88} & \frac{15}{88} \end{pmatrix}.$$

Now compute:

$$(\bar{y} - \mu_o^*)^T S_y^{-1} (\bar{y} - \mu_o^*),$$

with

$$\bar{y} - \mu_o^* = \begin{pmatrix} -2 \\ 0 \end{pmatrix} \quad \text{and} \quad S_y^{-1} = \begin{pmatrix} \frac{75}{88} & \frac{27}{88} \\ \frac{27}{88} & \frac{15}{88} \end{pmatrix}.$$

$$S_y^{-1} \begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{75}{88} & \frac{27}{88} \\ \frac{27}{88} & \frac{15}{88} \end{pmatrix} \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

$$= \begin{pmatrix} -\frac{150}{88} \\ -\frac{54}{88} \end{pmatrix} = \begin{pmatrix} -\frac{75}{44} \\ -\frac{27}{44} \end{pmatrix}$$

$$(\bar{y} - \mu_o^*)^T = (-2 \quad 0).$$

$$(\bar{y} - \mu_o^*)^T S_y^{-1} (\bar{y} - \mu_o^*) = (-2 \quad 0) \begin{pmatrix} -\frac{75}{44} \\ -\frac{27}{44} \end{pmatrix}.$$

$$(-2) \left(-\frac{75}{44}\right) + (0) \left(-\frac{27}{44}\right) = \frac{150}{44} + 0 = \frac{150}{44} = \frac{75}{22}.$$

Since  $n = 4$ , the Hotelling's  $T^2$  statistic for the transformed data is given by

$$T_y^2 = n \times [(\bar{y} - \mu_o^*)^T S_y^{-1} (\bar{y} - \mu_o^*)] = 4 \times \frac{75}{22} = \frac{300}{22} = \frac{150}{11}.$$

Therefore,

$$T_y^2 = \frac{150}{11}.$$

(c)

Suppose  $C$  is a  $p \times p$  invertible matrix and the transformed data are defined by

$$y_j = Cx_j, \quad j = 1, \dots, n.$$

Then:

The sample mean of the transformed data is

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j = \frac{1}{n} \sum_{j=1}^n Cx_j = C \bar{x}.$$

The transformed hypothesized mean is

$$\mu_y^* = C \mu_o.$$

Since  $C$  is not random, the sample covariance matrix of the transformed data

$$S_y = \frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y})(y_j - \bar{y})^T.$$

Given the linear transformation

$$y_j = C x_j,$$

the difference  $y_j - \bar{y}$  can be written as

$$y_j - \bar{y} = C x_j - C \bar{x} = C(x_j - \bar{x}).$$

$$S_y = \frac{1}{n-1} \sum_{j=1}^n [C(x_j - \bar{x})] [C(x_j - \bar{x})]^T.$$

For any matrices  $A$  and  $B$ , the transpose satisfies  $(AB)^T = B^T A^T$ . Thus,

$$[C(x_j - \bar{x})]^T = (x_j - \bar{x})^T C^T.$$

Now we have:

$$S_y = \frac{1}{n-1} \sum_{j=1}^n C(x_j - \bar{x})(x_j - \bar{x})^T C^T.$$

Since the matrix  $C$  does not depend on  $j$ , it can be factored out of the summation:

$$S_y = C \left[ \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T \right] C^T.$$

The expression inside is the sample covariance matrix  $S$  of the original data:

$$S = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T.$$

Thus,

$$S_y = C S C^T.$$

The Hotelling's  $T^2$  statistic for the original data is

$$T_x^2 = n (\bar{x} - \mu_o)^T S^{-1} (\bar{x} - \mu_o).$$

For the transformed data, the statistic is defined as

$$T_y^2 = n (\bar{y} - C \mu_o)^T (S_y)^{-1} (\bar{y} - C \mu_o).$$

Since  $\bar{y} = C \bar{x}$ , we can write

$$\bar{y} - C \mu_o = C(\bar{x} - \mu_o).$$

Substitute this into  $T_y^2$ :

$$T_y^2 = n [C(\bar{x} - \mu_o)]^T (S_y)^{-1} [C(\bar{x} - \mu_o)] = n (\bar{x} - \mu_o)^T C^T (S_y)^{-1} C (\bar{x} - \mu_o)$$

Using the fact that  $S_y = C S C^T$  and the invertibility of  $C$  we have

$$(S_y)^{-1} = (C S C^T)^{-1} = (C^T)^{-1} S^{-1} C^{-1}.$$

Therefore,

$$\begin{aligned} T_y^2 &= n (\bar{x} - \mu_o)^T C^T (C^T)^{-1} S^{-1} C^{-1} C (\bar{x} - \mu_o) \\ &= n (\bar{x} - \mu_o)^T S^{-1} (\bar{x} - \mu_o) \\ &= T_x^2. \end{aligned}$$

This completes the proof that Hotelling's  $T^2$  is invariant under any invertible linear transformation.

## Question 2

(a)

Let

$$C = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2},$$

and then the square roots of the eigenvalues of  $CC^T$  are the canonical correlations; the largest such singular value is the largest canonical correlation  $\rho_1^*$ .

By definition,

$$CC^T = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T \Sigma_{11}^{-1/2}.$$

Since

$$\Sigma_{12} = r J, \quad \text{with } J = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

first compute  $\Sigma_{22}^{-1}$ . For

$$\Sigma_{22} = \begin{pmatrix} q & 1 \\ 1 & q \end{pmatrix},$$

its inverse is

$$\Sigma_{22}^{-1} = \frac{1}{q^2 - 1} \begin{pmatrix} q & -1 \\ -1 & q \end{pmatrix}.$$

Thus,

$$\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T = r J \Sigma_{22}^{-1} r J = r^2 J \Sigma_{22}^{-1} J.$$

$$\Sigma_{22}^{-1} J = \frac{1}{q^2 - 1} \begin{pmatrix} q & -1 \\ -1 & q \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{q^2 - 1} \begin{pmatrix} q - 1 & q - 1 \\ -1 + q & -1 + q \end{pmatrix} = \frac{q - 1}{q^2 - 1} J = \frac{1}{q + 1} J$$

Then,

$$J \Sigma_{22}^{-1} J = J \frac{1}{q + 1} J = \frac{1}{q + 1} J^2.$$

$$J^2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = 2J.$$

Thus,

$$J \Sigma_{22}^{-1} J = \frac{2}{q + 1} J,$$

$$\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T = r^2 J \Sigma_{22}^{-1} J = \frac{2r^2}{q + 1} J.$$

we now have

$$CC^T = \frac{2r^2}{q + 1} \Sigma_{11}^{-1/2} J \Sigma_{11}^{-1/2}.$$

Since

$$\Sigma_{11} = \begin{pmatrix} 1 & p \\ p & 1 \end{pmatrix},$$

To compute  $\Sigma_{11}^{-1/2}$ , we first perform an eigen-decomposition of  $\Sigma_{11}$ .

The eigenvalues  $\lambda$  satisfy

$$\det \left( \begin{pmatrix} 1 - \lambda & p \\ p & 1 - \lambda \end{pmatrix} \right) = (1 - \lambda)^2 - p^2 = 0.$$

Thus,

$$(1 - \lambda)^2 = p^2 \implies 1 - \lambda = \pm p,$$

which gives

$$\lambda_1 = 1 + p \quad \text{and} \quad \lambda_2 = 1 - p.$$

The corresponding normalized eigenvectors are:

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (\text{for } \lambda_1 = 1 + p), \quad v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (\text{for } \lambda_2 = 1 - p).$$

We can express  $\Sigma_{11}$  as

$$\Sigma_{11} = V \Lambda V^T,$$

where

$$V = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 1+p & 0 \\ 0 & 1-p \end{pmatrix}.$$

Since

$$\Sigma_{11}^{-1/2} = V \Lambda^{-1/2} V^T,$$

we have

$$\Lambda^{-1/2} = \begin{pmatrix} \frac{1}{\sqrt{1+p}} & 0 \\ 0 & \frac{1}{\sqrt{1-p}} \end{pmatrix}.$$

Thus,

$$\begin{aligned} \Sigma_{11}^{-1/2} &= \frac{1}{\sqrt{1+p}} v_1 v_1^T + \frac{1}{\sqrt{1-p}} v_2 v_2^T. \\ v_1 v_1^T &= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad v_2 v_2^T = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \end{aligned}$$

Hence,

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

now define

$$A = \frac{1}{\sqrt{1+p}}, \quad B = \frac{1}{\sqrt{1-p}},$$

so that

$$\Sigma_{11}^{-1/2} = \frac{1}{2} \begin{pmatrix} A+B & A-B \\ A-B & A+B \end{pmatrix}.$$

$$CC^T = \frac{2r^2}{q+1} \Sigma_{11}^{-1/2} J \Sigma_{11}^{-1/2}.$$



$$\begin{aligned}
\Sigma_{11}^{-1/2} J &= \frac{1}{2} \begin{pmatrix} A+B & A-B \\ A-B & A+B \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} (A+B) + (A-B) & (A+B) + (A-B) \\ (A-B) + (A+B) & (A-B) + (A+B) \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} 2A & 2A \\ 2A & 2A \end{pmatrix} \\
&= \begin{pmatrix} A & A \\ A & A \end{pmatrix} \\
&= A J.
\end{aligned}$$

$$X = \Sigma_{11}^{-1/2} J \Sigma_{11}^{-1/2} = (\Sigma_{11}^{-1/2} J) \Sigma_{11}^{-1/2} = A J \Sigma_{11}^{-1/2}.$$

Since  $\Sigma_{11}^{-1/2}$  is symmetric (because  $\Sigma_{11}$  is symmetric),

$$J \Sigma_{11}^{-1/2} = \Sigma_{11}^{-1/2} J.$$

Hence,

$$J \Sigma_{11}^{-1/2} = A J.$$

Therefore, by  $A = \frac{1}{\sqrt{1+p}}$ ,

$$\begin{aligned}
\Sigma_{11}^{-1/2} J \Sigma_{11}^{-1/2} &= \frac{1}{1+p} J. \\
CC^T &= \frac{2r^2}{q+1} \cdot \frac{1}{1+p} J = \frac{2r^2}{(q+1)(1+p)} J.
\end{aligned}$$

The matrix  $J = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$  has eigenvalues:

$$\lambda_1(J) = 2 \quad \text{and} \quad \lambda_2(J) = 0.$$

Thus, the eigenvalues of  $CC^T$  are

$$\lambda_1(CC^T) = \frac{2r^2}{(q+1)(1+p)} \times 2 = \frac{4r^2}{(q+1)(1+p)} \quad \text{and} \quad \lambda_2(CC^T) = 0.$$

Taking square roots, the largest singular value (which is the largest canonical correlation) is

$$\rho_1^* = \sqrt{\frac{4r^2}{(q+1)(1+p)}} = \frac{2r}{\sqrt{(q+1)(1+p)}}.$$

(b)

From part (a) that we have defined

$$CC^T = \frac{2r^2}{(q+1)(1+p)} J,$$

with

$$J = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Since the eigenvalues of  $J$  are 2 (with normalized eigenvector proportional to  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ) and 0, the eigenvector corresponding to the largest eigenvalue of  $CC^T$  is

$$e_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

We already have

$$\Sigma_{11}^{-1/2} = \frac{1}{2} \begin{pmatrix} \frac{1}{\sqrt{1+p}} + \frac{1}{\sqrt{1-p}} & \frac{1}{\sqrt{1+p}} - \frac{1}{\sqrt{1-p}} \\ \frac{1}{\sqrt{1+p}} - \frac{1}{\sqrt{1-p}} & \frac{1}{\sqrt{1+p}} + \frac{1}{\sqrt{1-p}} \end{pmatrix}.$$

let

$$A = \frac{1}{\sqrt{1+p}}, \quad B = \frac{1}{\sqrt{1-p}},$$

$$\Sigma_{11}^{-1/2} = \frac{1}{2} \begin{pmatrix} A+B & A-B \\ A-B & A+B \end{pmatrix}.$$

using the following formula:

$$a_1 = \Sigma_{11}^{-1/2} e_1.$$

$$a_1 = \frac{1}{2} \begin{pmatrix} A+B & A-B \\ A-B & A+B \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$a_1 = \frac{1}{2\sqrt{2}} \begin{pmatrix} (A+B) + (A-B) \\ (A-B) + (A+B) \end{pmatrix} = \frac{1}{2\sqrt{2}} \begin{pmatrix} 2A \\ 2A \end{pmatrix} = \frac{A}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

we have  $A = 1/\sqrt{1+p}$ :

$$a_1 = \frac{1}{\sqrt{2(1+p)}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

This  $a_1$  satisfies the normalization  $a_1^T \Sigma_{11} a_1 = 1$ .

For  $b_1$ , since the covariance matrix

$$\Sigma_{22} = \begin{pmatrix} q & 1 \\ 1 & q \end{pmatrix}$$

has eigenvalues  $q+1$  and  $q-1$  with the normalized eigenvector corresponding to  $q+1$  being

$$w_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Thus, following the same procedure as for  $\Sigma_{11}$ , we have

$$\Sigma_{22}^{-1/2} = \frac{1}{2} \begin{pmatrix} D+E & D-E \\ D-E & D+E \end{pmatrix},$$

where

$$D = \frac{1}{\sqrt{q+1}}, \quad E = \frac{1}{\sqrt{q-1}}.$$

Then,

$$b_1 = \Sigma_{22}^{-1/2} w_1 = \frac{1}{2} \begin{pmatrix} D+E & D-E \\ D-E & D+E \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$b_1 = \frac{1}{2\sqrt{2}} \begin{pmatrix} (D+E) + (D-E) \\ (D-E) + (D+E) \end{pmatrix} = \frac{1}{2\sqrt{2}} \begin{pmatrix} 2D \\ 2D \end{pmatrix} = \frac{D}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Since  $D = 1/\sqrt{q+1}$ , we obtain

$$b_1 = \frac{1}{\sqrt{2(q+1)}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

This satisfies  $b_1^T \Sigma_{22} b_1 = 1$ .

The canonical variate pair is given by

$$U_1 = a_1^T \mathbf{X} \quad \text{and} \quad V_1 = b_1^T \mathbf{Y}.$$

That is,

$$U_1 = \frac{X_1 + X_2}{\sqrt{2(1+p)}}, \quad V_1 = \frac{Y_1 + Y_2}{\sqrt{2(q+1)}}$$

## Question 3

(a)

The following code gives the result of canonical correlation analysis of these data:

```
stiff = read.table("/Users/yubin/Desktop/Multivariate Analysis/stiffness.DAT")

X <- stiff[, 1:2]
Y <- stiff[, 3:4]

cc <- cancel(X, Y)
print("Canonical Correlation Analysis Results:")
print(cc)
```

The result is:

```
[1] "Canonical Correlation Analysis Results:"
$cor
[1] 0.91291935 0.06805556

$xcoef
      [,1]      [,2]
V1 -0.0006687933 -0.001237328
V2  0.0001106253  0.001430402

$ycoef
      [,1]      [,2]
V3 -0.0002497238  0.001573032
V4 -0.0003515941 -0.001453802

$xcenter
      V1      V2
1906.100 1749.533

$ycenter
      V3      V4
1509.133 1724.967
```

These result shows that the first canonical correlation is approximately 0.913, which suggests a strong linear relationship between the dynamic and static measures of stiffness. The second canonical correlation is only about 0.068, implying that the second pair of canonical variates has little association.

(b)

The first canonical variates  $U_1$  and  $V_1$  are defined as the linear combinations of the components of  $X$  and  $Y$ , respectively, using the canonical coefficients obtained from the canonical correlation analysis. That is,

$$U_1 = a_{11}X_1 + a_{21}X_2 \quad \text{and} \quad V_1 = b_{11}Y_1 + b_{21}Y_2,$$

then the canonical coefficients (for the first canonical variate) from the output are:

For  $X$  (dynamic measures):

$$a_{11} = -0.0006687933, \quad a_{21} = 0.0001106253,$$

and for  $Y$  (static measures):

$$b_{11} = -0.0002497238, \quad b_{21} = -0.0003515941.$$

Thus, the first canonical variates can be written as:

$$U_1 = -0.0006687933 X_1 + 0.0001106253 X_2$$

and

$$V_1 = -0.0002497238 Y_1 - 0.0003515941 Y_2$$

(c)

The following code produce the scatterplot needed:

```
cc <- cancel(X, Y)

U <- as.matrix(X) %*% cc$xcoef
V <- as.matrix(Y) %*% cc$ycoef

# Plot for the first canonical variate pair (U1, V1)
plot(U[,1], V[,1], type = "n",
     xlab = "U1 (Dynamic Measures)",
     ylab = "V1 (Static Measures)",
     cex.lab = 0.8, cex.axis = 0.5)
text(U[,1], V[,1], labels = rownames(X), cex = 0.6, col = rainbow(nrow(X)))
title(cex.main = 0.9, main = "Stiffness Data: Observations by Canonical Variates (U1, V1)")

# Plot for the second canonical variate pair (U2, V2)
plot(U[,2], V[,2], type = "n",
     xlab = "U2 (Dynamic Measures)",
     ylab = "V2 (Static Measures)",
     cex.lab = 0.8, cex.axis = 0.5)
text(U[,2], V[,2], labels = rownames(X), cex = 0.6, col = rainbow(nrow(X)))
title(cex.main = 0.9, main = "Stiffness Data: Observations by Canonical Variates (U2, V2)")
```

Output:

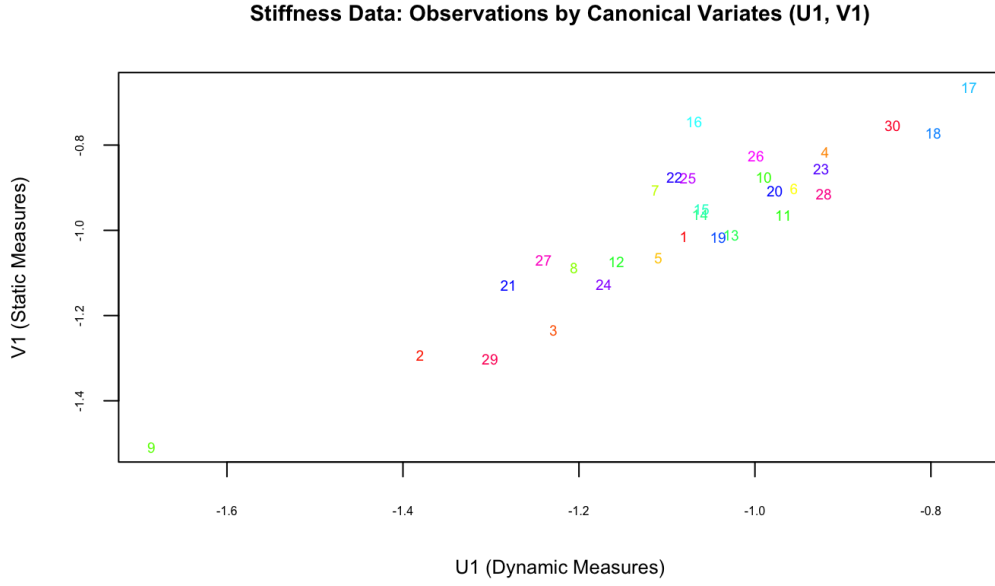


Figure 1: Stiffness Data: Observations by Canonical Variates (U1, V1)

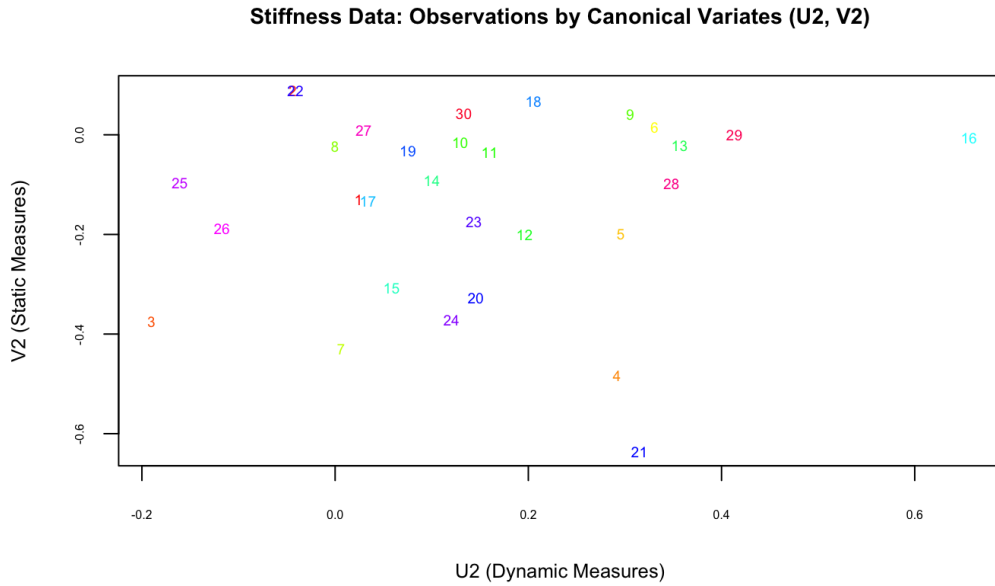


Figure 2: Stiffness Data: Observations by Canonical Variates (U2, V2)

(d)

From the canonical correlation analysis and plot, we obtain two canonical variate pairs:

$$(U_1, V_1) \quad \text{and} \quad (U_2, V_2).$$

Let  $\rho_1^*$  and  $\rho_2^*$  be the corresponding canonical correlations, with

$$\rho_1^* > \rho_2^*.$$

#### First Canonical Variate Pair $(U_1, V_1)$ .

- **High Correlation:** The first pair typically explains the largest portion of the linear relationship between the dynamic stiffness measures  $(X_1, X_2)$  and the static measures  $(X_3, X_4)$ .

In our plots, the scatter of points in  $(U_1, V_1)$  coordinates appears roughly linear, indicating a strong positive correlation. Numerically,  $\rho_1^*$  is close to 0.91, confirming a substantial linear relationship.

- **Interpretation:** Because  $(U_1, V_1)$  captures the dominant mode of covariance between  $(X_1, X_2)$  and  $(X_3, X_4)$ , a board's position along  $(U_1, V_1)$  can be interpreted as reflecting how dynamic and static stiffness are jointly related in their first principal direction of correlation. Boards that lie far out on one side in the  $(U_1, V_1)$  plane likely have higher levels of both dynamic and static measures (or vice versa).

#### Second Canonical Variate Pair $(U_2, V_2)$ .

- **Smaller Correlation:** The second pair explains the remaining variability after accounting for the first pair. Its correlation,  $\rho_2^*$ , is smaller than  $\rho_1^*$ . In our example, if  $\rho_1^* \approx 0.91$ , the second correlation is only 0.068. In the  $(U_2, V_2)$  plane, the points are scattered, suggesting that there is minimal leftover linear association between  $(X_1, X_2)$  and  $(X_3, X_4)$  beyond what was already captured by the first variates. This is consistent with the small canonical correlation for the second pair.

#### Overall Comment on the Correlation Structure.

- The **first canonical correlation** is high ( $\rho_1^*$  near 0.91), indicating that most of the linear association between the dynamic (shock/vibration) and static (two static tests) stiffness measures is explained by the first pair  $(U_1, V_1)$ .
- The **second canonical correlation** is substantially smaller ( $\rho_2^* \approx 0.07$ ), which implies that after removing the first mode of correlation, there is very little additional linear relationship between the dynamic and static variables. The second scatter plot  $(U_2, V_2)$  shows points with little to no obvious linear trend.
- Consequently, almost all of the correlation between the dynamic and static measures is “captured” by the first canonical variate pair.

## Question 4

(a)

(i)

Using the following R code to perform Hotelling's T test:

```
fly$Y1 <- fly$Ant.Length + fly$Wing.Length
fly$Y2 <- fly$Wing.Length
flyDataY <- data.frame(Y1 = fly$Y1, Y2 = fly$Y2, Species = fly$Species)
resY <- manova(cbind(Y1, Y2) ~ Species, data = flyDataY)
summary(resY, test = "Hotelling")
```

Outcome:

```

      Df Hotelling-Lawley approx F num Df den Df      Pr(>F)
Species    1           4.2985  25.791      2    12 4.519e-05 ***
Residuals 13
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

This result indicates that the corresponding p-value is  $4.519 \times 10^{-5}$ , which is much smaller than the significance level ( $\alpha = 0.05$ ).

Thus, we reject the null hypothesis of equality of the mean vectors between the two species (based on  $(Y_1, Y_2)$ ). In other words, there is a statistically significant difference in the mean vectors between the two species.

(ii)

Yes, we obtain the same test results when using the original variables  $(X_1, X_2)$  instead of  $(Y_1, Y_2)$ :

```

flyDataX <- data.frame(X1 = fly$Ant.Length, X2 = fly$Wing.Length, Species = fly$Species)
resX <- manova(cbind(X1, X2) ~ Species, data = flyDataX)
summary(resX, test = "Hotelling")
      Df Hotelling-Lawley approx F num Df den Df      Pr(>F)
Species    1           4.2985  25.791      2    12 4.519e-05 ***
Residuals 13
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

**Proof:** If we define

$$Y_1 = X_1 + X_2 \quad \text{and} \quad Y_2 = X_2,$$

this transformation can be written in matrix form as:

$$Y = AX, \quad \text{with} \quad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Suppose we have two samples with means  $\bar{X}_1$  and  $\bar{X}_2$  and a common covariance matrix  $S_X$  for the variables  $X$ . The Hotelling's  $T^2$  statistic for testing the null hypothesis  $H_0 : \mu_1 = \mu_2$  (using the original variables) is:

$$T_X^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)^T S_X^{-1} (\bar{X}_1 - \bar{X}_2).$$

Now, assume that the variables  $Y$  are defined by a nonsingular linear transformation of  $X$ :

$$Y = AX,$$

with  $A$  being an invertible  $2 \times 2$  matrix. Then the mean vectors and the sample covariance matrix for the  $Y$ -data are given by:

$$\bar{Y}_j = A \bar{X}_j \quad (j = 1, 2) \quad \text{and} \quad S_Y = A S_X A^T.$$

The Hotelling's  $T^2$  statistic based on  $Y$  becomes:

$$T_Y^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{Y}_1 - \bar{Y}_2)^T S_Y^{-1} (\bar{Y}_1 - \bar{Y}_2).$$

Substitute  $\bar{Y}_1 - \bar{Y}_2 = A(\bar{X}_1 - \bar{X}_2)$  and  $S_Y^{-1} = (A S_X A^T)^{-1} = (A^T)^{-1} S_X^{-1} A^{-1}$  into the expression:

$$\begin{aligned} T_Y^2 &= \frac{n_1 n_2}{n_1 + n_2} [A(\bar{X}_1 - \bar{X}_2)]^T (A^T)^{-1} S_X^{-1} A^{-1} [A(\bar{X}_1 - \bar{X}_2)] \\ &= \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)^T A^T (A^T)^{-1} S_X^{-1} A^{-1} A (\bar{X}_1 - \bar{X}_2) \\ &= \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)^T S_X^{-1} (\bar{X}_1 - \bar{X}_2) \\ &= T_X^2. \end{aligned}$$

Thus, the Hotelling's  $T^2$  statistic remains unchanged under the nonsingular linear transformation, and the conclusion regarding the equality of the mean vectors will be the same whether the test is performed on  $(X_1, X_2)$  or on  $(Y_1, Y_2)$ .

(b)

We conducted univariate two-sample t-tests on the variables  $Y_1$  and  $Y_2$  using two different assumptions on the variance (equal and unequal).

### (1) Under the Equal Variance Assumption

```
> t_Y1 <- t.test(Y1 ~ Species, data = flyDataY, var.equal = TRUE)
> print("Two-sample t-test for Y1 ( = 0.05):")
[1] "Two-sample t-test for Y1 ( = 0.05):"
> print(t_Y1)
```

Two Sample t-test

```
data: Y1 by Species
t = 0.66097, df = 13, p-value = 0.5202
alternative hypothesis: true difference in means between group Af and group Apf is not equal to 0
95 percent confidence interval:
 -0.1461906  0.2750795
sample estimates:
mean in group Af mean in group Apf
      3.217778      3.153333
```

```
> t_Y2 <- t.test(Y2 ~ Species, data = flyDataY, var.equal = TRUE)
> print("Two-sample t-test for Y2 ( = 0.05):")
[1] "Two-sample t-test for Y2 ( = 0.05):"
> print(t_Y2)
```

Two Sample t-test

```
data: Y2 by Species
t = -2.0047, df = 13, p-value = 0.06628
alternative hypothesis: true difference in means between group Af and group Apf is not equal to 0
95 percent confidence interval:
 -0.253933843  0.009489398
sample estimates:
mean in group Af mean in group Apf
      1.804444      1.926667
```

- For  $Y_1$ , the test results are:

$$t = 0.66097, \quad df = 13, \quad p\text{-value} = 0.5202,$$



Thus, we fail to reject the null hypothesis of equal means for  $Y_1$  at  $\alpha = 0.05$  (and at  $\alpha = 0.01$ ).

- For  $Y_2$ , the test gives:

$$t = -2.0047, \quad df = 13, \quad p\text{-value} = 0.0663,$$

Thus, we fail to reject the null hypothesis of equal means for  $Y_2$  at  $\alpha = 0.05$  (and at  $\alpha = 0.01$ ).

## (2) Under the Unequal Variance Assumption

```
> t_Y1 <- t.test(Y1 ~ Species, data = flyDataY, var.equal = FALSE)
> print("Two-sample t-test for Y1 ( = 0.05):")
[1] "Two-sample t-test for Y1 ( = 0.05):"
> print(t_Y1)
```

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
```

```
> t_Y2 <- t.test(Y2 ~ Species, data = flyDataY, var.equal = FALSE)
> print("Two-sample t-test for Y2 ( = 0.05):")
[1] "Two-sample t-test for Y2 ( = 0.05):"
> print(t_Y2)
```

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
```

- For  $Y_1$ , the Welch Two Sample t-test reports:

$$t = 0.7132, \quad df \approx 12.943, \quad p\text{-value} = 0.4884.$$

Thus, we fail to reject the null hypothesis of equal means for  $Y_2$  at  $\alpha = 0.05$  (and at  $\alpha = 0.01$ ).

- For  $Y_2$ , Welch's test yields:

$$t = -2.1697, \quad df \approx 12.967, \quad p\text{-value} = 0.0492.$$

Thus, we reject the null hypothesis of equal means for  $Y_2$  at  $\alpha = 0.05$  but cannot reject the null at  $\alpha = 0.01$ .

## Conclusion

- Under the equal variance assumption, both univariate tests yield  $p$ -values (0.5202 for  $Y_1$  and 0.0663 for  $Y_2$ ) above 0.05, so the hypothesis of equal species means would be **accepted** at  $\alpha = 0.05$  (and at  $\alpha = 0.01$ ).
- Under the unequal variance assumption, the  $p$ -values are 0.4884 (for  $Y_1$ ) and 0.0492 (for  $Y_2$ ). Here,  $Y_2$  is significant at the 5% level (since  $0.0492 < 0.05$ ) but not at the 1% level.

(c)

Below is an R code snippet that produces the scatterplot of  $(Y_1, Y_2)$  for the two species groups (**Af** and **Apf**). In our dataset,

$$Y_1 = \text{Antenna Length} + \text{Wing Length}, \quad Y_2 = \text{Wing Length}.$$

```
fly_Af <- subset(flyDataY, Species == "Af")
fly_Apf <- subset(flyDataY, Species == "Apf")

plot(NULL,
      xlim = range(flyDataY$Y1),
      ylim = range(flyDataY$Y2),
      xlab = "Y1 (Antenna + Wing Length)",
      ylab = "Y2 (Wing Length)",
      main = "Species Comparison: Y1 vs Y2")

points(fly_Af$Y1, fly_Af$Y2, pch = 16, col = "royalblue")
points(fly_Apf$Y1, fly_Apf$Y2, pch = 17, col = "firebrick")

legend("topleft",
      legend = c("Af", "Apf"),
      pch = c(16, 17),
      col = c("royalblue", "firebrick"),
      title = "Species")
```

An example of such a plot is shown in Figure ??.

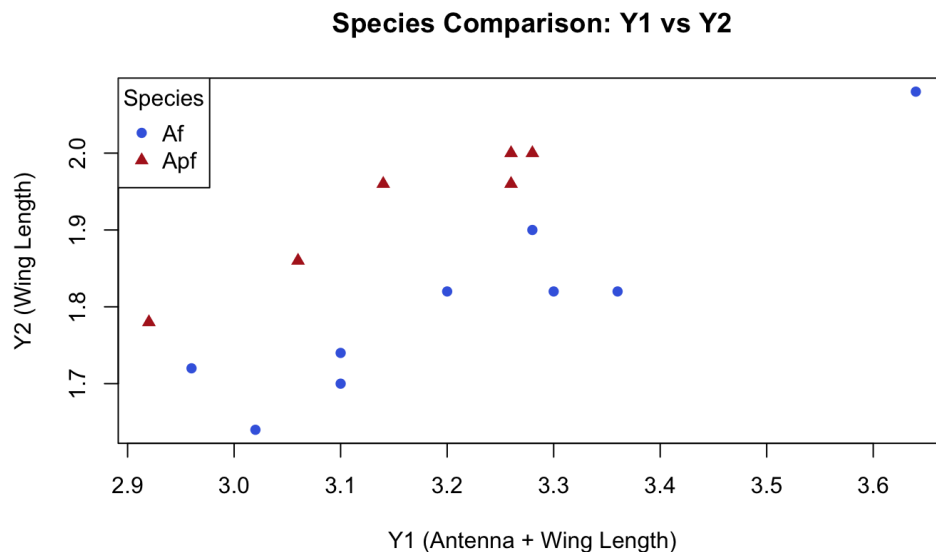


Figure 3: Scatterplot of  $(Y_1, Y_2)$  for the two species, with different symbols and colors for each group.

**Why can (a) and (b) yield different conclusions?** The scatterplot of  $Y_1$  versus  $Y_2$  for the two species shows that the data points from the two species (Af and Apf) have some degree of overlap along each individual axis. When we perform separate univariate  $t$ -tests on  $Y_1$  and  $Y_2$ , the tests examine only the marginal differences in the means for each variable; in our case, the  $t$ -test for  $Y_1$  yields a  $p$ -value of approximately 0.5202 and for  $Y_2$  a  $p$ -value of about 0.0663. Individually, these tests do not strongly reject the null hypothesis at the 5% significance level.

However, Hotelling's  $T^2$  test (as in part (a)) simultaneously considers the joint distribution of  $(Y_1, Y_2)$  and explicitly accounts for the correlation between these two variables. In the scatterplot, the direction of greatest separation between the species may not align with the  $Y_1$  or  $Y_2$  axis. Instead, it may lie along an oblique direction where the combined effect of  $Y_1$  and  $Y_2$  differences is more pronounced. The multivariate test leverages this joint variation and detects a statistically significant difference in the mean vectors, even though the marginal differences appear weak.

More specifically,

- **Part (a)** refers to the multivariate Hotelling's  $T^2$  test on  $(Y_1, Y_2)$  simultaneously. This test accounts for the potential correlation between  $Y_1$  and  $Y_2$  and produces a single test of whether the mean vector  $(\mu_{Y_1}, \mu_{Y_2})$  differs between species.
- **Part (b)** performs two univariate two-sample  $t$ -tests—one on  $Y_1$  and another on  $Y_2$ . These tests do not account for the correlation between  $Y_1$  and  $Y_2$ , and test the hypotheses

$$H_0 : \mu_{Y_1}^{(\text{Af})} = \mu_{Y_1}^{(\text{Apf})} \quad \text{and} \quad H_0 : \mu_{Y_2}^{(\text{Af})} = \mu_{Y_2}^{(\text{Apf})}$$

separately.

Because of this difference in methodology, the separate  $t$ -tests can fail to detect a joint effect if, for instance,  $\mu_{Y_1}$  and  $\mu_{Y_2}$  do not differ much individually (leading to relatively high  $p$ -values for each univariate test), but the combination of those small differences in a correlated setting may be significant when tested jointly by Hotelling's  $T^2$ .

Conversely, it is also possible that one variable alone may show significance while the multivariate test, considering correlation and the joint variation, might not.

In this particular dataset, the univariate tests gave:

$$p\text{-value for } Y_1 \approx 0.5202, \quad p\text{-value for } Y_2 \approx 0.0663.$$

So at the 5% level, neither test rejects the null hypothesis individually, whereas the multivariate test might reject when considering  $(Y_1, Y_2)$  together if there is enough joint separation of the species means. This discrepancy arises because the multivariate test uses the covariance structure of  $(Y_1, Y_2)$ , leveraging correlation between variables to detect differences in the mean vectors that do not appear significant in either variable alone.

(d)

(i) and (ii)

Using the following code to produce the plot we need:

```
fly$Y1 <- fly$Ant.Length + fly$Wing.Length
fly$Y2 <- fly$Wing.Length
flyDataY <- data.frame(Y1 = fly$Y1, Y2 = fly$Y2, Species = fly$Species)

fly_Af <- subset(flyDataY, Species == "Af")
fly_Apf <- subset(flyDataY, Species == "Apf")
```

```

mean_Af <- colMeans(fly_Af[, c("Y1", "Y2")])
mean_Apf <- colMeans(fly_Apf[, c("Y1", "Y2")])

# Part(i)
center <- mean_Af - mean_Apf

n1 <- nrow(fly_Af)
n2 <- nrow(fly_Apf)
p <- 2

cov_Af <- cov(fly_Af[, c("Y1", "Y2")])
cov_Apf <- cov(fly_Apf[, c("Y1", "Y2")])
Spool <- ((n1 - 1)*cov_Af + (n2 - 1)*cov_Apf) / (n1 + n2 - 2)

S_diff <- Spool * (1/n1 + 1/n2)

alpha <- 0.98
F_crit <- qf(alpha, p, n1 + n2 - p - 1)

T2_crit <- ( (n1 + n2 - 2)*p / (n1 + n2 - p - 1) ) * F_crit
radius <- sqrt(T2_crit)

plot(NA,
      xlab = "Mean Difference in Y1",
      ylab = "Mean Difference in Y2",
      main = "98% Confidence Ellipse for Mean Difference",
      xlim = center[1] + c(-2, 2) * radius * sqrt(S_diff[1,1]),
      ylim = center[2] + c(-2, 2) * radius * sqrt(S_diff[2,2]))

ellipse(center = center,
        shape = S_diff,
        radius = radius,
        col = "red",
        lwd = 2,
        add = TRUE)

# Part(ii)
alpha_rect <- 0.01
df_t <- n1 + n2 - 2

var_Y1 <- Spool[1,1]
var_Y2 <- Spool[2,2]
se_Y1 <- sqrt(var_Y1*(1/n1 + 1/n2))
se_Y2 <- sqrt(var_Y2*(1/n1 + 1/n2))

t_crit <- qt(1 - alpha_rect/2, df_t)

x1 <- center[1] - t_crit * se_Y1
x2 <- center[1] + t_crit * se_Y1
y1 <- center[2] - t_crit * se_Y2
y2 <- center[2] + t_crit * se_Y2

rect(xleft = x1, ybottom = y1, xright = x2, ytop = y2,
     border = "blue", lwd = 1.5)

```

```

abline(h = 0, v = 0, col = "gray50", lty = 3)
points(center[1], center[2], pch = 21, bg = "gold", cex = 1.2)
legend("topleft",
      legend = c("Hotelling's 98% Ellipse", "Marginal 99% CIs"),
      col = c("red", "blue"), lty = c(2, 1), lwd = 1.5)
grid()

```

Output:

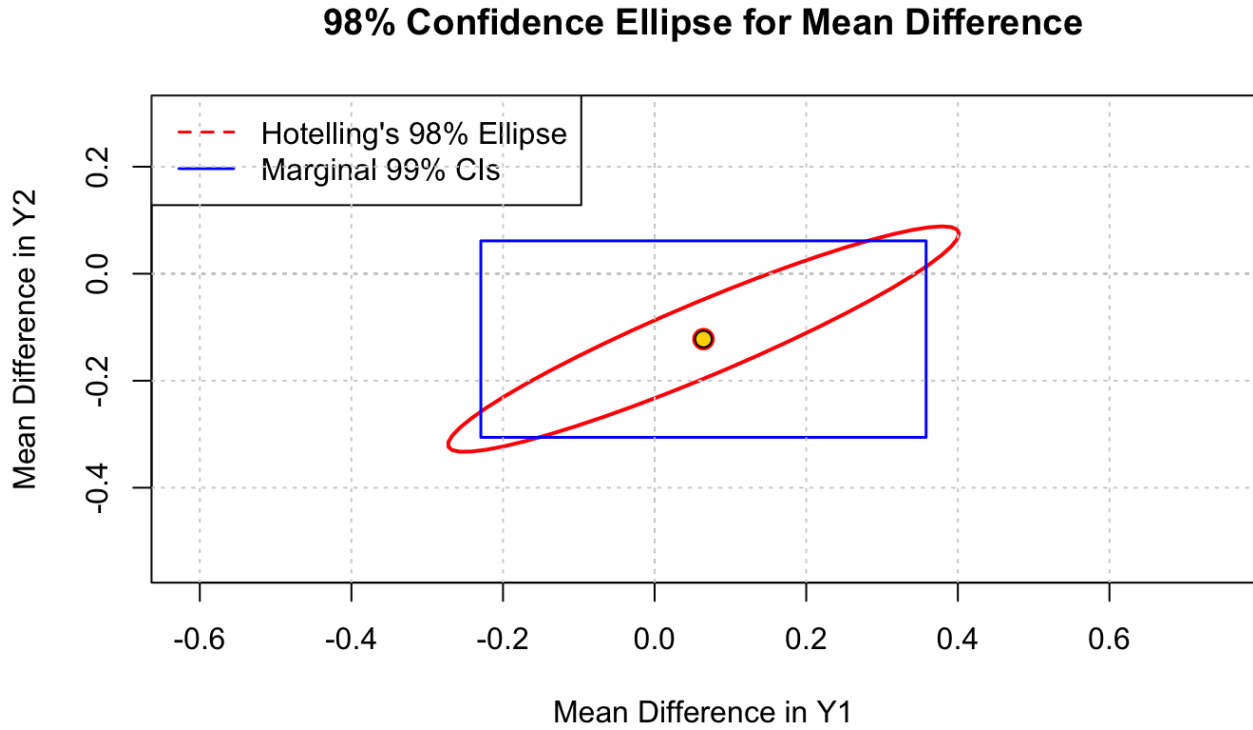


Figure 4: Confidence region

(iii)

Let  $\theta_1$  and  $\theta_2$  denote the true species mean differences for  $Y_1$  and  $Y_2$ , respectively.

Our goal is to construct simultaneous confidence intervals for  $\theta_1$  and  $\theta_2$  with an overall confidence level of  $1 - \alpha = 0.98$ , where  $\alpha = 0.02$ .

Suppose we construct individual (marginal) confidence intervals for  $\theta_1$  and  $\theta_2$  each at level

$$1 - \frac{\alpha}{2} = 1 - 0.01 = 0.99.$$

That is, define

$$\mathcal{I}_1 = \left[ \hat{\theta}_1 - t_{0.01, \nu} \text{SE}(\hat{\theta}_1), \hat{\theta}_1 + t_{0.01, \nu} \text{SE}(\hat{\theta}_1) \right],$$

and

$$\mathcal{I}_2 = \left[ \hat{\theta}_2 - t_{0.01, \nu} \text{SE}(\hat{\theta}_2), \hat{\theta}_2 + t_{0.01, \nu} \text{SE}(\hat{\theta}_2) \right],$$

where  $\nu$  represents the appropriate degrees of freedom, and  $\hat{\theta}_1$  and  $\hat{\theta}_2$  are the sample estimates (the components of the difference of the two sample mean vectors).

By the Bonferroni inequality, the probability that both intervals simultaneously contain their respective true parameters is at least

$$P(\theta_1 \in \mathcal{I}_1 \text{ and } \theta_2 \in \mathcal{I}_2) \geq 1 - [P(\theta_1 \notin \mathcal{I}_1) + P(\theta_2 \notin \mathcal{I}_2)].$$

Since each marginal interval has a failure probability at most  $\alpha/2 = 0.01$ , we obtain

$$P(\theta_1 \in \mathcal{I}_1 \text{ and } \theta_2 \in \mathcal{I}_2) \geq 1 - (0.01 + 0.01) = 0.98.$$

Thus, the rectangle defined by  $\mathcal{I}_1 \times \mathcal{I}_2$  is a 98% simultaneous confidence region for the mean difference vector  $(\theta_1, \theta_2)$ , which is a 98% confidence region by the Bonferroni method.

(iv)

In the figure provided, we see two confidence regions for the difference in mean vectors  $(Y_1, Y_2)$  between the two species:

- A **red dashed ellipse**, constructed via Hotelling's  $T^2$  at the 98% level.
- A **blue rectangle**, representing the product of the two marginal 99% confidence intervals for  $Y_1$  and  $Y_2$ , thus yielding a Bonferroni 98% confidence region.

By inspecting the plot:

$(0, 0)$  is not contained in the red ellipse, but it is contained within the blue rectangle.

## Comparison of the Confidence Regions

- **Hotelling's Ellipse:** This region is based on Hotelling's  $T^2$  statistic and takes into account the joint variability and correlation between  $Y_1$  and  $Y_2$ . In our example, because  $Y_1$  and  $Y_2$  are correlated, the ellipse better captures the actual shape of the joint uncertainty. It is generally smaller in area (and thus more precise) than the corresponding Bonferroni rectangle. Notably, in the plotted ellipse, the point  $(0, 0)$  is not included, suggesting that the difference in species mean vectors is statistically significant.
- **Bonferroni Rectangle:** This region is constructed by taking the product of the two marginal 99% confidence intervals for  $Y_1$  and  $Y_2$ . Although simple to compute and interpret, this rectangle ignores the correlation between  $Y_1$  and  $Y_2$  and is typically more conservative (i.e., has a larger area). In this case, the rectangle covers  $(0, 0)$ , indicating less evidence against the null hypothesis of equal means when considering the two variables separately.

- **Which Region is Better?**

In this scene, the Hotelling ellipse is more appropriate for a joint confidence region because it incorporates the correlation between  $Y_1$  and  $Y_2$  and tends to have a smaller area, thereby providing a more precise statement about the species mean difference.

The Bonferroni rectangle, while easier to interpret as it comes from independent marginal tests, is more conservative and overestimates the area needed to achieve the same overall confidence level (98%) when the variables are correlated.

## Question 5

(a)

Let

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}$$

be an  $n \times p$  data matrix of  $n$  observations, where the  $j$ -th observation is  $x_j \in R^p$ . The sample mean is given by

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j.$$

The sample covariance matrix is defined as

$$S = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T.$$

First, note that

$$\begin{aligned} H &= I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T, \\ HX &= X - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T X. \end{aligned}$$

let  $\mathbf{1}_n$  denote the  $n \times 1$  column vector whose entries are all 1, i.e.,

$$\mathbf{1}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Then its transpose is the  $1 \times n$  row vector

$$\mathbf{1}_n^T = (1 \quad 1 \quad \cdots \quad 1).$$

The product  $\mathbf{1}_n^T X$  is a  $1 \times p$  row vector. Its  $k$ th entry is computed as follows:

$$(\mathbf{1}_n^T X)_k = \sum_{j=1}^n (\mathbf{1}_n^T)_j x_{jk} = \sum_{j=1}^n 1 \cdot x_{jk} = \sum_{j=1}^n x_{jk}.$$

Thus,

$$\mathbf{1}_n^T X = \left( \sum_{j=1}^n x_{j1}, \sum_{j=1}^n x_{j2}, \dots, \sum_{j=1}^n x_{jp} \right).$$

Now consider the product

$$\frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T X.$$

Here,  $\frac{1}{n} \mathbf{1}_n$  is an  $n \times 1$  column vector where every entry is  $\frac{1}{n}$ .

When we multiply an  $n \times 1$  column vector by a  $1 \times p$  row vector, we obtain an  $n \times p$  matrix. In particular, the  $j$ th row of the resulting matrix is given by

$$\left( \frac{1}{n} \right)_n \left( \sum_{k=1}^n x_{k1}, \sum_{k=1}^n x_{k2}, \dots, \sum_{k=1}^n x_{kp} \right) = \left( \frac{1}{n} \sum_{k=1}^n x_{k1}, \frac{1}{n} \sum_{k=1}^n x_{k2}, \dots, \frac{1}{n} \sum_{k=1}^n x_{kp} \right).$$

Note that the  $1 \times p$  row vector

$$\bar{x}^T = \left( \frac{1}{n} \sum_{j=1}^n x_{j1}, \frac{1}{n} \sum_{j=1}^n x_{j2}, \dots, \frac{1}{n} \sum_{j=1}^n x_{jp} \right)$$

defines the sample mean, we have

$$\frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T X = \begin{pmatrix} \bar{x}^T \\ \bar{x}^T \\ \vdots \\ \bar{x}^T \end{pmatrix},$$

which is the  $n \times p$  matrix having  $\bar{x}^T$  as every row.

Therefore,

$$H X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix} - \begin{pmatrix} \bar{x}^T \\ \bar{x}^T \\ \vdots \\ \bar{x}^T \end{pmatrix} = \begin{pmatrix} (x_1 - \bar{x})^T \\ (x_2 - \bar{x})^T \\ \vdots \\ (x_n - \bar{x})^T \end{pmatrix}.$$

For the matrix  $H$ :

$$H = I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T,$$

note that the identity matrix  $I_n$  is symmetric, i.e.,

$$I_n^T = I_n.$$

and,

$$\left( \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right)^T = \frac{1}{n} (\mathbf{1}_n \mathbf{1}_n^T)^T = \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T,$$

Thus, both  $I_n$  and  $\frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$  are symmetric. Therefore,

$$H^T = \left( I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right)^T = I_n^T - \left( \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right)^T = I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T = H.$$

Hence,  $H$  is symmetric.

And,

$$H^2 = \left( I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right)^2.$$

$$H^2 = I_n^2 - 2 \cdot \frac{1}{n} I_n \mathbf{1}_n \mathbf{1}_n^T + \frac{1}{n^2} \mathbf{1}_n \mathbf{1}_n^T \mathbf{1}_n \mathbf{1}_n^T.$$

Since  $I_n^2 = I_n$  and  $I_n \mathbf{1}_n = \mathbf{1}_n$ , this simplifies to:

$$H^2 = I_n - \frac{2}{n} \mathbf{1}_n \mathbf{1}_n^T + \frac{1}{n^2} \mathbf{1}_n (\mathbf{1}_n^T \mathbf{1}_n) \mathbf{1}_n^T.$$

and,

$$\mathbf{1}_n^T \mathbf{1}_n = n.$$

Thus,

$$\begin{aligned} \frac{1}{n^2} \mathbf{1}_n (n) \mathbf{1}_n^T &= \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T. \\ H^2 &= I_n - \frac{2}{n} \mathbf{1}_n \mathbf{1}_n^T + \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T = I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T. \end{aligned}$$

Hence,

$$H^2 = H.$$



we have

$$H X = \begin{pmatrix} (x_1 - \bar{x})^T \\ (x_2 - \bar{x})^T \\ \vdots \\ (x_n - \bar{x})^T \end{pmatrix}.$$

Taking the transpose of  $H X$ , we obtain:

$$(H X)^T = (x_1 - \bar{x} \quad x_2 - \bar{x} \quad \cdots \quad x_n - \bar{x}).$$

Because  $H$  is symmetric ( $H = H^T$ ) and idempotent ( $H^2 = H$ ), we have

$$X^T H X = X^T H H X = (H X)^T H X$$

Now,  $(H X)^T (H X)$  is a  $p \times p$  matrix computed as:

$$(H X)^T (H X) = \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T.$$

This follows because the  $j$ th column of  $(H X)^T$  is  $x_j - \bar{x}$ , so the  $(u, v)$ th entry of the product is

$$[(H X)^T (H X)]_{uv} = \sum_{j=1}^n (x_j - \bar{x})_u (x_j - \bar{x})_v,$$

which is the  $(u, v)$ th entry of  $\sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T$ .

Thus, we have that

$$X^T H X = \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T.$$

$$S = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T = \frac{1}{n-1} X^T H X.$$

**(b)**

$$\text{Cov}(\mathbf{W}) = E[(\mathbf{W} - E[\mathbf{W}])(\mathbf{W} - E[\mathbf{W}])^T].$$

Since  $E[\mathbf{W}] = A E[\mathbf{Y}] + \mathbf{c}$ :

$$\mathbf{W} - E[\mathbf{W}] = (A\mathbf{Y} + \mathbf{c} - A E[\mathbf{Y}] - \mathbf{c}) = A(\mathbf{Y} - E[\mathbf{Y}]).$$

$$(\mathbf{W} - E[\mathbf{W}])^T = [A(\mathbf{Y} - E[\mathbf{Y}])]^T = (\mathbf{Y} - E[\mathbf{Y}])^T A^T$$

Since  $A$  is a fixed, scalar matrix, it can be factored out of expectation:

$$\text{Cov}(\mathbf{W}) = E[A(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{Y} - E[\mathbf{Y}])^T A^T] = A E[(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{Y} - E[\mathbf{Y}])^T] A^T = A \text{Cov}(\mathbf{Y}) A^T.$$

**(c)**

Since  $a^T$  and  $b^T$  are fixed vectors:

$$E[a^T Y] = a^T E[Y],$$

$$E[b^T W] = b^T E[W].$$

Thus,

$$a^T Y - E[a^T Y] = a^T (Y - E[Y]),$$

$$b^T W - E[b^T W] = b^T (W - E[W]).$$

By definition,

$$\text{Cov}(a^T Y, b^T W) = E \left[ \left( a^T (Y - E[Y]) \right) \left( b^T (W - E[W]) \right) \right].$$

Since fixed vectors can be factored out:

$$\text{Cov}(a^T Y, b^T W) = a^T E \left[ (Y - E[Y])(W - E[W])^T \right] b.$$

By the definition of covariance,

$$E \left[ (Y - E[Y])(W - E[W])^T \right] = \text{Cov}(Y, W).$$

Thus,

$$\text{Cov}(a^T Y, b^T W) = a^T \text{Cov}(Y, W) b.$$

Since  $\text{Cov}(W, Y) = \text{Cov}(Y, W)^T$  and the scalar  $\text{Cov}(a^T Y, b^T W)$  is equal to its own transpose, we also have

$$\text{Cov}(a^T Y, b^T W) = \left( a^T \text{Cov}(Y, W) b \right)^T = b^T \text{Cov}(Y, W)^T a = b^T \text{Cov}(W, Y) a.$$

Thus, we have shown:

$$\text{Cov}(a^T Y, b^T W) = a^T \text{Cov}(Y, W) b = b^T \text{Cov}(W, Y) a.$$