

1

(a)

We will use Box M test.

The adjusted statistic is:

$$U_1 = -2(1 - c_1) \ln M$$

M statistic:

$$M = \frac{|S_1|^{\frac{n_1-1}{2}} |S_2|^{\frac{n_2-1}{2}}}{|S_p|^{\frac{n_1-1}{2} + \frac{n_2-1}{2}}}$$

Pooled Covariance Matrix:

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

Correction factor:

$$c_1 = \left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1} - \frac{1}{n_1 + n_2 - 2} \right) \frac{(2p^2 + 3p - 1)}{6(p + 1)(k - 1)}$$

Where:

- n_1, n_2 are the sample sizes
- k is number of groups
- p is number of variables

We reject H_0 if:

$$u_1 > \chi_{1-\alpha, \frac{1}{2}p(p+1)}^2$$

(b)

```
haltica_oleracea <- matrix(c(
  189, 245, 137, 163,
  192, 260, 132, 217,
  217, 276, 141, 192,
  221, 299, 142, 213,
  171, 239, 128, 158,
  192, 262, 147, 173,
  213, 278, 136, 201,
  192, 255, 128, 185,
  170, 244, 128, 192,
  201, 276, 146, 186,
  195, 242, 128, 192,
  205, 263, 147, 192,
  180, 252, 121, 167,
  192, 283, 138, 183,
  200, 294, 138, 188,
  192, 277, 150, 177,
  200, 287, 136, 173,
  181, 255, 146, 183,
  192, 287, 141, 198
), ncol = 4, byrow = TRUE)
colnames(haltica_oleracea) <- c("y1", "y2", "y3", "y4")

haltica_carduorum <- matrix(c(
```

```

181, 305, 184, 209,
158, 237, 133, 188,
192, 300, 166, 231,
171, 273, 162, 213,
181, 297, 163, 224,
181, 308, 160, 222,
177, 301, 166, 221,
198, 308, 141, 197,
180, 286, 146, 214,
177, 299, 171, 192,
176, 317, 166, 213,
192, 312, 166, 205,
176, 285, 141, 200,
169, 287, 162, 214,
164, 265, 147, 192,
181, 308, 157, 204,
192, 276, 154, 209,
181, 278, 149, 235,
175, 271, 140, 192,
197, 303, 170, 205
), ncol = 4, byrow = TRUE)
colnames(haltica_carduorum) <- c("y1", "y2", "y3", "y4")

# Sample sizes
n1 <- nrow(haltica_oleracea)
n2 <- nrow(haltica_carduorum)

# Number of groups
k <- 2

# Number of variables (predictors)
p <- ncol(haltica_oleracea)

# Sample covariance matrices
S1 <- cov(haltica_oleracea)
S2 <- cov(haltica_carduorum)

# Pooled covariance matrix
Sp <- ((n1 - 1) * S1 + (n2 - 1) * S2) / (n1 + n2 - 2)

# Determinants
det_S1 <- det(S1)
det_S2 <- det(S2)
det_Sp <- det(Sp)

# Compute M
M <- (det_S1^((n1 - 1)/2) * det_S2^((n2 - 1)/2)) / (det_Sp^((n1 + n2 - 2)/2))

# Correction factor c1
c1 <- (
  ((1 / (n1 - 1)) + (1 / (n2 - 1)) - (1 / (n1 + n2 - 2)))
  * (2 * p^2 + 3 * p - 1) / (6 * (p + 1) * (k - 1))
)

```

```
# Adjusted test statistic U1
U1 <- -2 * (1 - c1) * log(M)
```

The statistic is $u_1 = 7.9531$.

(c)

```
alpha <- 0.05
df <- 0.5 * p * (p + 1)
critical_value <- qchisq(1 - alpha, df)

# Hypothesis test
reject_H0 <- U1 > critical_value
```

Our critical value is 18.307.

We will accept the null hypothesis.

2

(a)

We will use Hotelling's generalized T-squared statistic.

The statistic is:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{C}^T [\mathbf{C} \mathbf{S}_p \mathbf{C}^T]^{-1} \mathbf{C} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

Pooled Covariance Matrix:

$$\mathbf{S}_p = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2}$$

Contrast Matrix:

$$\mathbf{C} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Mean Vectors:

$$\bar{\mathbf{x}}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{x}_{1i}$$

$$\bar{\mathbf{x}}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{x}_{2i}$$

Where:

- n_1, n_2 are the sample sizes.

We reject H_0 if:

$$T^2 > T_{1-\alpha, p-1, n_1+n_2-2}^2$$

$$T^2 > \frac{(n_1 + n_2 - 2)(p - 1)}{n_1 + n_2 - p} F_{1-\alpha, p-1, n_1+n_2-p}$$

(b)

```

# Sample means
xbar1 <- colMeans(haltica_oleracea)
xbar2 <- colMeans(haltica_carduorum)

# Contrast matrix
C <- matrix(c(
  -1, 1, 0, 0,
  0, -1, 1, 0,
  0, 0, -1, 1
), nrow = 3, byrow = TRUE)

# Difference in means
d <- xbar1 - xbar2

# Compute T^2
T2 <- (
  (n1 * n2) / (n1 + n2) * t(d) %*% t(C) %*% solve(C %*% Sp %*% t(C)) %*% C %*% d
)
T2 <- as.numeric(T2) # convert to scalar

```

Our test statistic is $T^2 = 132.3532$.

(c)

```

# Degrees of freedom
p <- ncol(haltica_oleracea)
r <- nrow(C)
v <- n1 + n2 - 2

# Corresponding F-statistic
F_stat <- (v - r + 1) * T2 / (v * r)

# Critical value
alpha <- 0.05
F_crit <- qf(1 - alpha, df1 = r, df2 = v - r + 1)

# Hypothesis test
reject_H0 <- F_stat > F_crit

```

The corresponding F-statistic is $F = 41.733$.

The critical F-value is 2.8742.

We will reject the null hypothesis.

3

(a)

```

mydata <- t(matrix(c(
  # Method 1
  5.4, 5.2, 6.1, 4.8, 5.0, 5.7, 6.0, 4.0, 5.7, 5.6, 5.8, 5.3,
  6.0, 6.2, 5.9, 5.0, 5.7, 6.1, 6.0, 5.0, 5.4, 5.2, 6.1, 5.9,
  6.3, 6.0, 6.0, 4.9, 5.0, 6.0, 5.8, 4.0, 4.9, 5.4, 5.2, 5.8,

```

```

6.7, 5.8, 7.0, 5.0, 6.5, 6.6, 6.0, 5.0, 5.0, 5.8, 6.4, 6.0,

# Method 2
5.0, 4.8, 3.9, 4.0, 5.6, 6.0, 5.2, 5.3, 5.9, 6.1, 6.2, 5.1,
5.3, 4.9, 4.0, 5.1, 5.4, 5.5, 4.8, 5.1, 6.1, 6.0, 5.7, 4.9,
5.3, 4.2, 4.4, 4.8, 5.1, 5.7, 5.4, 5.8, 5.7, 6.1, 5.9, 5.3,
6.5, 5.6, 5.0, 5.8, 6.2, 6.0, 6.0, 6.4, 6.0, 6.2, 6.0, 4.8,

# Method 3
4.8, 5.4, 4.9, 5.7, 4.2, 6.0, 5.1, 4.8, 5.3, 4.6, 4.5, 4.4,
5.0, 5.0, 5.1, 5.2, 4.6, 5.3, 5.2, 4.6, 5.4, 4.4, 4.0, 4.2,
6.5, 6.0, 5.9, 6.4, 5.3, 5.8, 6.2, 5.7, 6.8, 5.7, 5.0, 5.6,
7.0, 6.4, 6.5, 6.4, 6.3, 6.4, 6.5, 5.7, 6.6, 5.6, 5.9, 5.5
), nrow = 12, ncol = 12))

k <- 4
n <- dim(mydata)[1]
p <- dim(mydata)[2]/k

mydata1 <- mydata[,1:3]
mydata2 <- mydata[,4:6]
mydata3 <- mydata[,7:9]
mydata4 <- mydata[,10:12]

ybar1 <- apply(mydata1,2,mean)
ybar2 <- apply(mydata2,2,mean)
ybar3 <- apply(mydata3,2,mean)
ybar4 <- apply(mydata4,2,mean)

ybar <- (ybar1+ybar2+ybar3+ybar4)/k

H <- n*((ybar1-ybar)%*%t(ybar1-ybar)+
        (ybar2-ybar)%*%t(ybar2-ybar)+
        (ybar3-ybar)%*%t(ybar3-ybar)+
        (ybar4-ybar)%*%t(ybar4-ybar))

E1 <- matrix(0,p,p)
for(j in 1:n){
  E1 <- E1+(mydata1[j,]-ybar1)%*%t(mydata1[j,]-ybar1)
}
E2 <- matrix(0,p,p)
for(j in 1:n){
  E2 <- E2+(mydata2[j,]-ybar2)%*%t(mydata2[j,]-ybar2)
}
E3 <- matrix(0,p,p)
for(j in 1:n){
  E3 <- E3+(mydata3[j,]-ybar3)%*%t(mydata3[j,]-ybar3)
}
E4 <- matrix(0,p,p)
for(j in 1:n){
  E4 <- E4+(mydata4[j,]-ybar4)%*%t(mydata4[j,]-ybar4)
}

```

```

E <- E1+E2+E3+E4

eig <- eigen(solve(E)%*%H)

v_h <- k-1
v_e <- k*(n-1)
w <- v_e+v_h-(p+v_h+1)/2
t <- sqrt((p^2*v_h^2-4)/(p^2+v_h^2-5))
s <- min(v_h,p)
df1 <- p*v_h
df2 <- w*t-(p*v_h-2)/2

# Wilk's Lambda
Lambda <- prod(1/(1+eig$values[1:s])) #prod(1/(1+eig$values)) different
#Lambda <- det(E)/det(E+H)
## F approximation
F_ <- (1-Lambda^(1/t))/Lambda^(1/t)*df2/df1

# Hypothesis test
F_crit <- qf(0.95,df1,df2)
reject_H0 <- F_ > F_crit

```

Our Wilk's Lambda value is $\Lambda = 0.5295$.

Our F approximation is $F = 3.3954$.

Our F critical value is 1.9726.

We will reject the null hypothesis.

(b)

```

theta <- eig$values[1]/(1+eig$values[1])
d <- max(p,v_h)
F_ <- (v_e-d-1)*eig$values[1]/d # upper bound

# Hypothesis test
F_crit <- qf(0.95,d,v_e-d-1)
reject_H0 <- F_ > F_crit

```

Our Roy's largest root value is $\theta = 0.3678$.

Our F approximation is $F = 7.7572$.

Our F critical value is 2.8387.

We will reject the null hypothesis.