

1

(a)

We will use the Hotelling T^2 Statistic because we have a two-sample multivariate hypothesis test. The formula is:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^T S_p^{-1} (\bar{x}_1 - \bar{x}_2)$$
$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

Where:

- \bar{x}_1, \bar{x}_2 are the sample mean vectors of the two groups
- n_1, n_2 are the sample sizes of each group
- S_p is the pooled sample covariance matrix

Our critical value is:

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

Where:

- p is the number of variables (dimensions)
- $F_{p, n_1 + n_2 - p - 1}$ is the F-distribution with degrees of freedom p and $n_1 + n_2 - p - 1$
 - This can be computed in R using the `qf` function

(b)

```
library(Hotelling)
```

```
## Loading required package: corpcor
```

```
# Haltica oleracea
species1 <- 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)
```

```

# Haltica carduorum
species2 <- matrix(c(
  181, 305, 184, 209,
  158, 237, 133, 188,
  184, 300, 166, 231,
  171, 273, 162, 213,
  181, 297, 163, 224,
  181, 308, 160, 223,
  177, 301, 166, 221,
  198, 308, 141, 197,
  180, 286, 146, 214,
  177, 299, 171, 192,
  176, 317, 166, 213,
  192, 312, 166, 209,
  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)

# Convert matrices to data frames
df_species1 <- as.data.frame(species1)
df_species2 <- as.data.frame(species2)
colnames(df_species1) <- c("y1", "y2", "y3", "y4")
colnames(df_species2) <- c("y1", "y2", "y3", "y4")

# Do test
alpha <- 0.05
result <- hotelling.test(df_species1, df_species2)

```

Our T^2 statistic is 133.4873 and our p-value is $7.5218e - 11$.

(c)

Since our p-value is less than α , we reject H_0 .

2

(a)

We will use the Modified Likelihood Ratio (MLR) to test whether the covariance matrix of a bivariate dataset is equal to a given matrix. The formula is:

$$u = (n - 1) [\text{tr}(S\Sigma_0^{-1}) - \log \det(S\Sigma_0^{-1}) - p]$$

Alternatively, let $\lambda_1, \dots, \lambda_p$ be the eigenvalues of $S\Sigma_0^{-1}$, then the test statistic can also be written as:

$$u = (n - 1) \sum_{i=1}^p (\lambda_i - \log \lambda_i - 1)$$

Where:

- S is the sample covariance matrix
- Σ_0 is the hypothesized covariance matrix under H_0
- n is the sample size
- p is the dimension (number of variables)

(b)

```
# Height and Weight Data
data <- matrix(c(
  69, 153,
  74, 175,
  68, 155,
  70, 135,
  72, 172,
  67, 150,
  66, 115,
  70, 137,
  76, 200,
  68, 130,
  72, 140,
  79, 265,
  74, 185,
  67, 112,
  66, 140,
  71, 150,
  74, 165,
  75, 185,
  75, 210,
  76, 220
), ncol = 2, byrow = TRUE)

# Convert matrices to data frames
df <- as.data.frame(data)
colnames(df) <- c("x", "y")

# Do calculation
Sigma0 <- matrix(c(
  20, 100,
  100, 1000
), nrow = 2)

S <- cov(df)
n <- nrow(df)

eigenvalues <- eigen(S %*% solve(Sigma0))$values
u <- (n - 1) * sum(eigenvalues - log(eigenvalues) - 1)
```

Our MLR statistic is 11.0937.

(c)

From lecture 9, u is asymptotically a Chi-squared distribution with degrees of freedom $\frac{1}{2}p(p+1)$.

So, we reject H_0 if:

$$u > \chi^2_{1-\alpha, \frac{1}{2}p(p+1)}$$

```
alpha <- 0.05
p <- ncol(df)
critical_value <- qchisq(1 - alpha, 1 / 2 * p * (p + 1))
```

Our critical value is 7.8147.

Since our test statistic is greater than our critical value, we reject the null hypothesis.

3

(a)

We are testing the hypothesis:

$$H_0 : \Sigma = \sigma^2 I \quad \text{vs} \quad H_a : \Sigma \neq \sigma^2 I$$

This is known as the sphericity test, where we test whether the population covariance matrix is a scalar multiple of the identity matrix.

The test statistic used is the Likelihood Ratio (LR):

$$\Lambda = \left[\frac{|\mathbf{S}|}{\left(\frac{\text{tr}(\mathbf{S})}{p} \right)^p} \right]^{\frac{n}{2}}$$

Where:

- S is the sample covariance matrix,
- p is the number of variables (dimensions),
- n is the sample size.

(b)

```
data <- matrix(c(
  47.8, 48.8, 49.0, 49.7,
  46.4, 47.3, 47.7, 48.4,
  46.3, 46.3, 47.8, 48.5,
  45.1, 45.3, 46.1, 47.2,
  47.6, 48.5, 48.9, 49.3,
  52.5, 53.2, 53.3, 53.7,
  51.2, 53.0, 54.3, 54.5,
  49.8, 50.0, 50.3, 52.7,
  48.1, 50.8, 52.3, 54.4,
  45.0, 47.0, 47.0, 48.3,
  51.2, 51.4, 51.6, 51.9,
  48.5, 49.2, 53.0, 55.5,
  52.1, 52.8, 53.7, 55.0,
  48.2, 49.4, 50.1, 49.8,
  49.6, 50.4, 51.2, 51.8,
  50.7, 51.7, 52.7, 53.5,
```

```

51.2, 52.0, 53.3, 54.1,
53.3, 54.6, 55.1, 55.3,
46.2, 47.1, 48.3, 51.4,
46.3, 47.6, 51.3, 51.8
), nrow = 20, byrow = TRUE)

# Convert Matrix to DataFrame
df <- as.data.frame(data)
colnames(df) <- c("Age_8", "Age_8.5", "Age_9", "Age_9.5")

# (b) Compute the test statistic
S <- cov(df) # Sample covariance matrix
p <- ncol(df) # Number of variables
n <- nrow(df) # Sample size

det_S <- det(S)
trace_S <- sum(diag(S))
lambda <- (det_S / ( (trace_S / p)^p ))^(n / 2)

```

Our LR statistic is $3.8299e - 31$.

(c)

We want to reject the null if:

$$-2\ln(\Lambda) > \chi^2_{1-\alpha, n-1}$$

```

alpha <- 0.05
chi_test_stat <- -2 * log(lambda)
critical_value <- qchisq(1 - alpha, n - 1)

```

Solving for the above equation:

$$140.0746 > 30.1435$$

Which means that we reject the null hypothesis.