1. total Sample Variance $= TR[S_x]$, Generalized Sample Variance $= |S_x|$

① $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $TR[S_x] = 3$, $|S_x| = 1$ ② $\begin{bmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{bmatrix}$, $TR[S_x] = 3$, $|S_x| = 1 - \frac{1}{8} - \frac{1}{8} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4} = 0$.

2. Show that $|S| = |R| \prod_{i=1}^{n} S_{ii}$

※ for question 4.5.6. I write a R markdown file to demonstrate it.

$r_{xy} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$，且 $S_{xx} = Var(X)$.

若 $P=2$, $r = \begin{bmatrix} 1 & r_{xy} \\ r_{xy} & 1 \end{bmatrix}$, $S = \begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$

$Var(X)Var(Y) - [Cov(X,Y)]^2 = \left(1 - \frac{[Cov(X,Y)]^2}{Var(X)Var(Y)}\right) \times Var(X)Var(Y)$.

$= Var(X)Var(Y) - [Cov(X,Y)]^2$ #.

Let $R = \begin{bmatrix} \frac{1}{\sqrt{S_{11}}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{S_{pp}}} \end{bmatrix} S_x \begin{bmatrix} \frac{1}{\sqrt{S_{11}}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{S_{pp}}} \end{bmatrix}$，$\because \det(ABC) = \det A \cdot \det B \cdot \det C$，

$\therefore |R| = \left(\frac{1}{\sqrt{S_{11} \cdot S_{22} \cdots S_{pp}}}\right)^2 |S|$. $\Rightarrow |S| = \left(\cancel{\frac{1}{S_{11} \cdot S_{22} \cdots S_{pp}}}\right) |S| \cancel{S_{11} S_{22} \cdots S_{pp}}$. $\Rightarrow |S| = \left(\cancel{\frac{S_{11} \cdot S_{22} \cdots S_{pp}}{S_{11} \cdot S_{22} \cdots S_{pp}}}\right) |S|$

3. ① Sample mean of $Y_1 = \bar{X}_1 + \bar{X}_2 = 0.766 + 0.508 + 0.438 + 0.161 = 1.875$ #

Sample variance of $Y_1 = Var(\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4) = \sum_{i=1}^{4} S_{ii} + 2[Cov(X_1,X_2) + Cov(X_1,X_3) + Cov(X_1,X_4) + Cov(X_2,X_3) + Cov(X_2,X_4) + Cov(X_3,X_4)]$

$= 0.856 + 0.568 + 0.171 + 0.043 + 2[0.635 + 0.173 + 0.128 + 0.096 + 0.067 + 0.039] = 3.914$ #

② Sample mean of $Y_2 = \bar{X}_1 - \bar{X}_2 = 0.766 - 0.508 = 0.258$ #

Sample variance of $Y_2 = Var(X_1) + Var(X_2) - 2Cov(X_1, X_2) = 0.856 + 0.568 - 0.635 \times 2 = 0.154$ #.

③ Covariance $= E(Y_1Y_2) - E(Y_1)E(Y_2)$, $E(Y_1Y_2) = E[(X_1 + X_2 + X_3 + X_4)(X_1 - X_2)] = E[X_1^2 + X_1X_3 + X_1X_4 - X_2^2 - X_2X_3 - X_2X_4]$

$\doteq 0.587 + 0.336 + 0.123 - 0.258 - 0.223 - 0.082 \doteq 0.483$.

$0.483 - 0.258 \times 1.875 \doteq -0.00075 \doteq 0$ #.

$\frac{168}{4}$ 42.

$E(X) = 6$, $E(X^2) = [4 + 64 + 36 + 64] = 42$. $42 - 36 = 6$.

4. (I demonstrate this in R, too).

$Var(X_1) = \{2,8,6,8\} \Rightarrow \frac{1}{n-1}\sum(X-E(X))^2 = 8$

$Var(X_2) = \{12,9,9,10\} \Rightarrow \frac{1}{4-1}\sum(X-E(X))^2 = 2.$

$Cov(X_1,X_2) = -\frac{10}{3}.$

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

$X = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$, $\bar{x} = [6, 10]$, $\mu_0 = [7, 11]$, $(\bar{x}-\mu_0) = (-1,-1)$. $n = 4$, $S = \begin{bmatrix} 8 & -\frac{10}{3} \\ -\frac{10}{3} & 2 \end{bmatrix} \Rightarrow S^{-1} = \frac{9}{44}\begin{bmatrix} 2 & \frac{10}{3} \\ \frac{10}{3} & 8 \end{bmatrix}$

$T^2 = \frac{9}{11}[-1,-1]\begin{bmatrix} 2 & \frac{10}{3} \\ \frac{10}{3} & 8 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \frac{9}{11}[-\frac{16}{3}, -\frac{34}{3}]\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \frac{9}{11} \times \frac{50}{9} = \frac{50}{11}.$

(b). $T^2 \sim T^2_{2,3} = \frac{3 \cdot \cancel{2}}{\cancel{2}} F_{2,2} = 3 \cdot F_{2,2}.$

Let's $\alpha = 95\% \Rightarrow T^2 \sim T^2_{95\%,2,3} = 3 \cdot F_{95\%,2,2} = 3 \times 19 = 57.$

(c). $\because \frac{50}{11} < 57 \therefore$ We can't reject $H_0 = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} = \begin{pmatrix} \mu_{01} \\ \mu_{0p} \end{pmatrix}.$

I demonstrate in R.

5. $i = 1$ (femal) $n\# = 24$, sample mean vector $= \bar{y}_1 : \begin{bmatrix} 136.04 \\ 102.58 \\ 52.42 \end{bmatrix}$, $\bar{y}_2 : \begin{bmatrix} 113.38 \\ 88.29 \\ 40.71 \end{bmatrix}$ ← L (length) ← W (width) ← H (height)

by using MANOVA, we reject H_0. because $Pr(>F)$ is $1.768e-09 \ll 0.001$.

From the output above, it can be seen that the variables are highly significantly different among turtles.

I demonstrate in R.

6. by using MANOVA, we can't reject H_0 and say there are not a significant species effect and nutrients effect one the spectral reflectance.

by taking two-way ANOVA twice, we get the same insights. The results of the two way ANOVA and MANOVA are consistent with each other.

→ Nutrients effect: Not significant in either the MANOVA or the two-way ANOVAs for each 560 nm and 720 nm.

Species effect: "

there was no significant interaction effect in either of the two-way ANOVA.

Untitled

柯宥圻

2023-03-25

HW4

QUESTION 4

I calculate the T2 value by hand on the second page, and I just use the package to confirm whether I calculate right.

```
#Question 4
data <- read.csv(file="C:/Users/user/Desktop/HW4-4.csv")
data
```

```
##   x1 x2
## 1  2 12
## 2  8  9
## 3  6  9
## 4  8 10
```

```
# install.packages("ICSNP")
library(ICSNP)
```

```
## Warning: 套件 'ICSNP' 是用 R 版本 4.2.3 來建造的
```

```
## 載入需要的套件：mvtnorm
```

```
## 載入需要的套件：ICS
```

```
## Warning: 套件 'ICS' 是用 R 版本 4.2.3 來建造的
```

```
HotellingsT2(data, mu=c(7,11))
```

```
##
##  Hotelling's one sample T2-test
##
## data:  data
## T.2 = 4.5455, df1 = 2, df2 = 2, p-value = 0.1803
## alternative hypothesis: true location is not equal to c(7,11)
```

since the P value is 0.1803, we can’t reject H0 and say that true location is not equal to c(7,11).

Question 5

we want to know the relationship between man’s turtle and woman’s turtle. Therefore, we have three dimensions. We take MANOVA to see if the two populations are equal.

```
turtle <- read.csv(file="C:/Users/user/Desktop/HW4-5.csv")
turtle
```

```
##    sex length width height
## 1    1     98    81     38
## 2    1    103    84     38
## 3    1    103    86     42
## 4    1    105    86     44
## 5    1    109    88     50
## 6    1    123    92     46
## 7    1    123    95     51
## 8    1    133    99     51
## 9    1    133   102     51
## 10   1    133   102     48
## 11   1    134   100     49
## 12   1    136   102     51
## 13   1    138    98     51
## 14   1    138    99     51
## 15   1    141   105     53
## 16   1    147   108     57
## 17   1    149   107     55
## 18   1    153   107     56
## 19   1    155   115     63
## 20   1    155   117     60
## 21   1    158   115     62
## 22   1    159   118     63
## 23   1    162   124     61
## 24   1    177   132     67
## 25   2     93    74     37
## 26   2     94    78     35
## 27   2     96    80     35
## 28   2    101    84     39
## 29   2    102    85     38
## 30   2    103    81     37
## 31   2    104    83     39
## 32   2    106    83     39
## 33   2    107    82     38
## 34   2    112    89     40
## 35   2    113    88     40
## 36   2    114    86     40
## 37   2    116    90     43
## 38   2    117    90     41
## 39   2    117    91     41
## 40   2    119    93     41
## 41   2    120    89     40
## 42   2    120    93     44
## 43   2    121    95     42
## 44   2    125    93     45
## 45   2    127    96     45
## 46   2    128    95     45
## 47   2    131    95     46
## 48   2    135   106     47
```

```
# install.packages("dplyr")
set.seed(1234)
dplyr::sample_n(turtle, 48)
```

```
##    sex length width height
## 1    2    101    84     39
## 2    1    147   108     57
## 3    1    159   118     63
## 4    2    116    90     43
## 5    2    125    93     45
## 6    1    133   102     51
## 7    1    109    88     50
## 8    2    117    90     41
## 9    2    131    95     46
## 10   1    105    86     44
## 11   2    112    89     40
## 12   2    128    95     45
## 13   2     94    78     35
## 14   1    123    92     46
## 15   1    141   105     53
## 16   1    138    99     51
## 17   1    155   117     60
## 18   2    107    82     38
## 19   2    103    81     37
## 20   1    177   132     67
## 21   2    117    91     41
## 22   2    135   106     47
## 23   1    158   115     62
## 24   1    133    99     51
## 25   2    106    83     39
## 26   1    103    86     42
## 27   2     96    80     35
## 28   2    120    93     44
## 29   1    103    84     38
## 30   2    120    89     40
## 31   2     93    74     37
## 32   2    119    93     41
## 33   1    136   102     51
## 34   1    162   124     61
## 35   1    123    95     51
## 36   2    121    95     42
## 37   1    155   115     63
## 38   2    113    88     40
## 39   2    127    96     45
## 40   1    153   107     56
## 41   1    138    98     51
## 42   2    102    85     38
## 43   1    149   107     55
## 44   2    104    83     39
## 45   1     98    81     38
## 46   1    134   100     49
## 47   2    114    86     40
## 48   1    133   102     48
```

```
res.man <- manova(cbind(length, width, height) ~ sex, data = turtle)
summary(res.man)
```

```
##           Df  Pillai approx F num Df den Df    Pr(>F)
## sex        1 0.62562   24.509      3     44 1.768e-09 ***
## Residuals 46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By using MANOVA, it can be seen that the P value is quite small, and the variables are not all the same. We use the summary.aov() to see which differ.

```
summary.aov(res.man)
```

```
##  Response length :
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## sex          1  6165.3  6165.3  20.889 3.656e-05 ***
## Residuals   46 13576.6   295.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response width :
##             Df Sum Sq Mean Sq F value    Pr(>F)
## sex          1 2451.0 2451.02  22.104 2.376e-05 ***
## Residuals   46 5100.8  110.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response height :
##             Df Sum Sq Mean Sq F value    Pr(>F)
## sex          1 1645.0 1645.02  46.006 1.931e-08 ***
## Residuals   46 1644.8   35.76
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

as the code shown, the three variables are all highly significantly different among those drawn turtles.

Question 6-a

we want to study whether the spectral reflectance at wavalength 560nm and 720nm has the species effect and nutrient effect on the alpha 5%. We first do the MANOVA test to see whether the data has these effect.

```
CM_mean <- data.frame(x1 = c(10.35, 13.41, 7.78, 10.4, 17.78, 10.4),
                      x2 = c(25.93, 38.63, 25.15, 24.25, 41.45, 29.2),
                      species = factor(c(1, 2, 3, 1, 2, 3)),
                      nutrient = factor(c(1,1,1,2,2,2)))

# MANOVA for the species effect
species_manova <- manova(cbind(x1, x2) ~ species, data = CM_mean)
summary(species_manova)
```

```
##           Df Pillai approx F num Df den Df  Pr(>F)
## species    2 1.4034   3.5281      4      6 0.08244 .
## Residuals  3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# MANOVA for the nutrient effect
nutrient_manova <- manova(cbind(x1, x2) ~ nutrient, data = CM_mean)
summary(nutrient_manova)
```

```
##           Df  Pillai approx F num Df den Df Pr(>F)
## nutrient   1 0.38864  0.95353      2      3  0.478
## Residuals  4
```

Based on the MANOVA results, we can conclude that there is no significant species effect on the spectral reflectance, but no significant nutrient effect.

Question 6-b

Based on the MANOVA results, we fail to reject the null hypothesis for the species effect, but we can still perform the two-way ANOVAs for both the 560CM and the 720CM observations to check for any significant effects of the nutrients and their interaction.

```
# two-way ANOVA for 560CM
anov560 <- aov(CM_mean$x1 ~ CM_mean$species + CM_mean$nutrient + CM_mean$species:CM_mean$nutrient)
summary(anov560)
```

```
##                                  Df Sum Sq Mean Sq
## CM_mean$species                   2  47.48  23.738
## CM_mean$nutrient                  1   8.26   8.260
## CM_mean$species:CM_mean$nutrient  2   4.72   2.361
```

```
# two-way ANOVA for 720CM
anov720 <- aov(CM_mean$x2 ~ CM_mean$species + CM_mean$nutrient + CM_mean$species:CM_mean$nutrient)
summary(anov720)
```

```
##                                  Df Sum Sq Mean Sq
## CM_mean$species                   2 262.24  131.12
## CM_mean$nutrient                  1   4.49    4.49
## CM_mean$species:CM_mean$nutrient  2   9.10    4.55
```

The output of each ANOVA includes the Sum of Squares (SS), the Degrees of Freedom (DF), the Mean Square (MS), the F statistic, and the p-value for each main effect and interaction.

Here is the expected output:

```
                                 Df Sum Sq Mean Sq F value  Pr(>F)
CM_mean$species                   2  47.48  23.738  1.0357 0.42410
CM_mean$nutrient                  1   8.26   8.260  0.3596 0.57989
CM_mean$species:CM_mean$nutrient  2   4.72   2.361  0.1026 0.90992
Residuals                         3  73.17  24.389

                                 Df Sum Sq Mean Sq F value  Pr(>F)
CM_mean$species                   2 262.24 131.120  2.0359 0.22498
CM_mean$nutrient                  1   4.49   4.490  0.0698 0.80434
CM_mean$species:CM_mean$nutrient  2   9.10   4.550  0.0707 0.80155
Residuals                         3 158.13  52.710
```

The results of the MANOVA test and the two-way ANOVAs are consistent with each other. Specifically, the species effect was not significant in either the MANOVA or the two-way ANOVAs for either the 560CM or the 720CM observations. The nutrient effect was not significant in either the MANOVA or the two-way ANOVAs for either the 560CM or the 720CM observations. There was also no significant interaction effect in either of the two-way ANOVAs.