2nd Exam 25-02-2022

G I

a) $\hat{Y} = -5.71892 - 0.02031\,\text{Area} + 0.31498\,\text{Elevation} - 0.07528\,\text{Adjacent}$

$\hat{Y}_{Baltra} = -5.71892 - 0.02031 \times 25.09 + 0.31498 \times 346 - 0.07528 \times 1.84 =$
$= 102.6161$

$\text{residual}_{Baltra} = Y_{Baltra} - \hat{Y}_{Baltra} = 58 - 102.6161 = -44.6161$

b) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ vs $H_1: \exists' \beta_i \neq 0,\ i = 1,2,3$; $p = 4,\ n = 30$

Assumption : $\varepsilon_i \underset{iid.}{\sim} N(0, \sigma^2) \Rightarrow Y_i \underset{indep}{\sim} N(\mu_i, \sigma^2)$

under H_0 we have the test statistic:

$F_0 = \dfrac{MSR}{MSE} \sim F_{(3,26)}$, rej H_0 if $F_0 > F^{-1}_{F_{(3,26)}}(0.99) = 4.64$

output : $\sqrt{MSE} = 61.01 \Rightarrow MSE = 3722.22$

$SSE = MSE \times 26 =$
$= 96777.72$

$r^2 = 1 - \dfrac{SSE}{SST} \Leftrightarrow SST = \dfrac{SSE}{1 - r^2} \Leftrightarrow SST = \dfrac{96777.72}{1 - 0.746} =$
$= 381014.7$

$SSR = SST - SSE = 381014.7 - 96777.72$
$= 284237$

$f_0 = \dfrac{\frac{284237}{3}}{3722.22} = \dfrac{94745.67}{3722.22} \simeq 25.45 \in C.R \Rightarrow$ rej H_0, $\Rightarrow \exists$ linear Association between the E(Species) and the 3 predictors

c) $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

Pivotal quantity: $T = \dfrac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 c_{2,2}}} \sim t_{(26)}$

under H_0, we have the test statistic: $T_0 = \dfrac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 c_{2,2}}} \sim t_{(26)}$

obs. value : $t_0 = \dfrac{-0.02031}{0.02181} = -0.931$
output

P-value $= 2P(T_0 \geq 0.931) = 2\left(1 - F_{t_{(26)}}(0.931)\right) \Rightarrow 0.3 < \text{P-value} < 0.4$
$\Rightarrow$ rej H_0 $\forall \alpha \geq 0.4$ and not rej H_0 $\forall \alpha \leq 0.3$, ie
not rej H_0 $\forall \alpha$ usual. So, Area is not statistically helpful.

d) $CI_{95\%}(\mu_{Y|Baltra}) = ?$

Pivotal quantity : $T = \dfrac{\hat{\mu}_{Y|Baltra} - \mu_{Y|Baltra}}{\sqrt{\hat{\sigma}^2 \underset{\sim}{x}^T_{Baltra} \underset{\sim}{C}\, \underset{\sim}{x}_{Baltra}}} \sim t_{(26)}$

choosing the symmetric C.I, we have:

$$P(-a \le T \le a) = 0.95 \qquad a = F^{-1}_{t(26)}(0.975) = 2.056$$

$$C.I._{95\%}(\mu_{Y|Baltra}) = \left[\hat{\mu}_{Y|Baltra} \pm 2.056\sqrt{\hat{\sigma}^2 \underset{\sim}{x}^T_{Baltra} \underset{\sim}{C} \underset{\sim}{x}_{Baltra}}\right]$$

$$\hat{\mu}_{Y|Baltra} = 102.6161$$

$$\sqrt{\hat{\sigma}^2 \underset{\sim}{x}^T_{Baltra} \underset{\sim}{C} \underset{\sim}{x}_{Baltra}} = \sqrt{3722.22 \times 0.0468} = 13.19848$$

$$CI_{95\%}(\mu_{Y|Baltra}) = [102.6161 \pm 2.056 \times 13.19848] = [75.480 ; 129.752]$$

2. $a = 3 ; N = 83$

$SSE = 190 \quad ; MSTR = 12.50 \implies SSTR = 12.50 \times 2 = 25$

$SST = SSTR + SSE = 190 + 25 = 215$

source	SS	df	MS	F
treat.	25	2	12.5	5.263
Error	190	80	2.375	
total	215	82		

$H_0: \mu_1 = \mu_2 = \mu_3$ vs $H_1: \mu_i \neq \mu_j \ \exists (i,j)$

under H_0, $F_0 = \dfrac{MSTR}{MSE} \sim F_{(2,80)}$

rej H_0 if $F_0 > F^{-1}_{F(2,80)}(0.95) = 3.11$

As $f_0 = 5.263 \in \mathcal{R}$, for $\alpha = 0.05$, rej $H_0 \implies$ the mean of the response variable is not the same for all treatments.

G II.

1. a) $a: \begin{cases} \underset{\sim}{\gamma}_2^T \underset{\sim}{\gamma}_2 = 1 \iff (-0.5683)^2 + (-0.0682)^2 + a^2 = 1 \iff a^2 \cong 1 - 0.3276 \iff \\ \qquad a = \pm\sqrt{1-0.3276} \iff a \cong \pm 0.82 \\ \underset{\sim}{\gamma}_2^T \underset{\sim}{\gamma}_1 = 0 \iff a = 0.82 \end{cases}$

b: total variance $= \sum_{i=1}^{3} \hat{\lambda}_i = 3 \iff b = 3 - 2.0157 - 0.8306 = 0.1537$

b) $\hat{Y}_1 = -0.5806 Z_1 - 0.6429 Z_2 - 0.4583 Z_3 \rightarrow$ global measure $z_i \uparrow \implies \hat{Y}_1 \downarrow$

$\hat{Y}_2 = -0.5683 Z_1 - 0.0682 Z_2 + 0.82 Z_3 \rightarrow$ contrast between z_1 and z_2

c) Total variance $= 3 = \sum_{i=1}^{3} \hat{\lambda}_i$

% total variance explained by $\hat{Y}_i$ = $\begin{cases} \dfrac{Var(\hat{Y}_1)}{\text{total variance}} \times 100\% = \dfrac{\hat{\lambda}_1}{\sum_{i=1} \lambda_i} \times 100\% = \dfrac{2.0157}{3} \times 100\% \cong 67.19\%, \ i=1 \\ (0.8306/3) \times 100\% \simeq 27.69\%, \ i=2 \\ (0.0512/3) \times 100\% \simeq 5.12\%, \ i=3 \end{cases}$

taking the rule % total variance explained by $\sum_i \hat{Y}_i \geq 80\%$ we should take 2 PC's

$$cor(\hat{Y}_1, Z_j) = \hat{\gamma}_{1j}\sqrt{\hat{\lambda}_1} = \begin{cases} -0.5806\sqrt{2.0157} \cong -0.824, & j=1 \\ -0.6729\sqrt{2.0157} \cong -0.955, & j=2 \\ -0.4583\sqrt{2.0157} \cong -0.651, & j=3 \end{cases}$$

$\hat{Y}_1$ is negative correlated with z_j, $j = 1,2,3$

As $z_j \uparrow$, $\hat{Y}_1 \downarrow$

$$Z_j = \frac{X_j - \bar{X}_j}{S_j}$$

2. a)

$$\underset{\sim}{J} = \begin{array}{c|ccccc} & o_1 & o_2 & o_3 & o_4 & o_5 \\ \hline o_1 & 1 & & & & \\ o_2 & 2/4 & 1 & & & \\ o_3 & 1/3 & 2/4 & 1 & & \\ o_4 & 1/4 & 3/4 & 2/3 & 1 & \\ o_5 & 0 & 1/4 & 1/2 & 1/3 & 1 \end{array}$$

$$j_{1,2} = \frac{a}{a+b+c} = 2/4$$

$$\begin{array}{cc|cc} & & \multicolumn{2}{c}{o_2} \\ & & 1 & 0 \\ \hline o_1 & 1 & 2 & 0 \\ & 0 & 2 & 0 \end{array}$$

$a = 2, b = 0$
$c = 2$

$$\underset{\sim}{D}^{(1)} = \begin{array}{c|ccccc} & o_1 & o_2 & o_3 & o_4 & o_5 \\ \hline o_1 & 0 & & & & \\ o_2 & 2/4 & 0 & & & \\ o_3 & 2/3 & 2/4 & 0 & & \\ o_4 & 3/4 & \boxed{1/4} & 1/3 & 0 & \\ o_5 & 1 & 3/4 & 1/2 & 2/3 & 0 \end{array}$$

b)

$h_1 = 1/4 \Rightarrow C_1 = \{o_2, o_4\}$

Average Linkage: $D_{A,B} = \frac{1}{n_A \times n_B} \sum_{i \in A} \sum_{j \in B} d_{ij}$

$n_A = \#A$; $n_B = \#B$

$$\underset{\sim}{D}^{(2)} = \begin{array}{c|cccc} & C_1 & O_1 & O_3 & O_5 \\ \hline C_1 & 0 & & & \\ O_1 & \frac{5}{8} & 0 & & \\ O_3 & \boxed{\frac{5}{12}} & 2/3 & 0 & \\ O_5 & \frac{17}{24} & 1 & 1/2 & 0 \end{array}$$

$$d_{C_1,O_1} = \frac{d_{1,2} + d_{1,4}}{2} = \frac{2/4 + \frac{3}{4}}{2} = \frac{5}{8}$$

$$d_{C_1,O_3} = \frac{d_{2,3} + d_{3,4}}{2} = \frac{2/4 + \frac{1}{3}}{2} = \frac{5}{12}$$

$$d_{C_1,O_5} = \frac{d_{2,5} + d_{4,5}}{5} = \frac{3/4 + 2/3}{2} = \frac{17}{24}$$

$h_2 = \frac{5}{12} \Rightarrow C_2 = \{C_1, O_3\}$

$$d_{C_2,O_1} = \frac{d_{1,2} + d_{1,3} + d_{1,4}}{3} = \frac{2/4 + 2/3 + 3/4}{3} = \frac{23}{36}$$

$$d_{C_2,O_5} = \frac{d_{2,5} + d_{3,5} + d_{4,5}}{3} = \frac{\frac{3}{4} + \frac{1}{2} + \frac{2}{3}}{3} = \frac{23}{36}$$

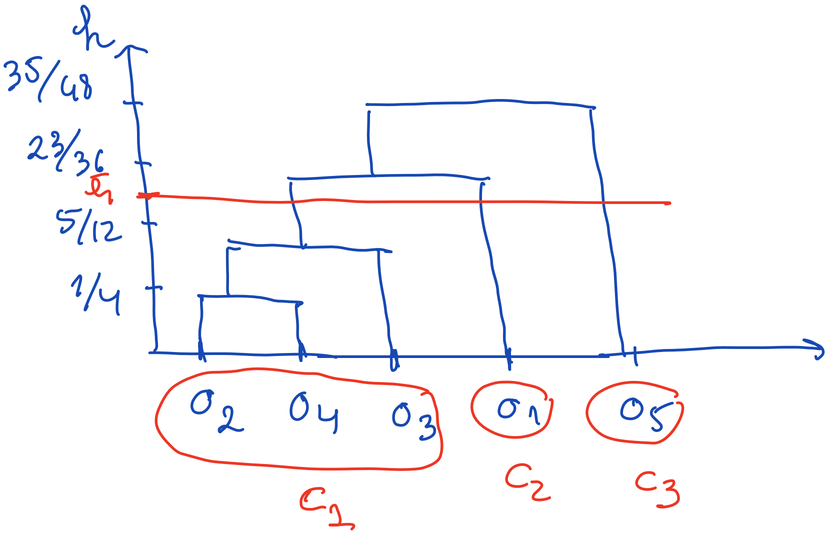
$$\underset{\sim}{D}^{(3)} = \begin{array}{c|ccc} & C_2 & O_1 & O_5 \\ \hline C_2 & 0 & & \\ O_1 & \frac{23}{3} & 0 & \\ O_5 & \frac{23}{36} & 3/4 & 0 \end{array}$$

$h_3 = \frac{23}{36} \Rightarrow C_3 = \{C_2, O_1\}$ ou O_5

$$\underset{\sim}{D}^{(4)} = \begin{array}{c|cc} & C_3 & O_5 \\ \hline C_3 & 0 & \\ O_5 & \frac{35}{48} & \end{array}$$

$$d_{C_3,O_5} = \frac{d_{1,5} + d_{2,5} + d_{3,5} + d_{4,5}}{4} =$$

$$= \frac{1 + 3/4 + 1/2 + 2/3}{4} = \frac{35}{48}$$



$\bar{h} \simeq 0.51$