STAT 500 Homework 5

Vu Thi-Hong-Ha; NetID: 851924086

September 25, 2020

1 Question 1

(a)

(i)

Denote:

Treatment means: $\mu_1, \mu_2, \mu_3, \mu_4$ (r = 4)

Treatment variances: $\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$

Treatment standard deviations: $\sigma_1, \sigma_2, \sigma_3, \sigma_4$

Samples: Y_{ij}, where $i = 1, 2, 3, 4$ denotes the treatment groups and $j = 1, 2, 3$ denotes the observations ($n_i = 3$).

$$\boldsymbol{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{41} \\ Y_{42} \\ Y_{43} \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \end{pmatrix}$$

(ii)

$\boldsymbol{\hat{\beta} = (X^T X)^{-1} X^T Y}$

$$= \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{3} Y_{1j} \\ \sum_{j=2}^{3} Y_{2j} \\ \sum_{j=3}^{3} Y_{3j} \\ \sum_{j=3}^{3} Y_{4j} \end{pmatrix} = \begin{pmatrix} \bar{Y_{1.}} \\ \bar{Y_{2.}} \\ \bar{Y_{3.}} \\ \bar{Y_{4.}} \end{pmatrix}$$

(b)

(i)

$$\boldsymbol{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{41} \\ Y_{42} \\ Y_{43} \end{pmatrix} = \begin{pmatrix} \mu+\alpha_1+\epsilon_{11} \\ \mu+\alpha_1+\epsilon_{12} \\ \mu+\alpha_1+\epsilon_{13} \\ \mu+\alpha_2+\epsilon_{21} \\ \mu+\alpha_2+\epsilon_{22} \\ \mu+\alpha_2+\epsilon_{23} \\ \mu+\alpha_3+\epsilon_{31} \\ \mu+\alpha_3+\epsilon_{32} \\ \mu+\alpha_3+\epsilon_{33} \\ \mu+\alpha_4+\epsilon_{41} \\ \mu+\alpha_4+\epsilon_{42} \\ \mu+\alpha_4+\epsilon_{43} \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \end{pmatrix}$$

(ii) Constraint $\sum_{i=1}^{r=4} \alpha_i = 0$, so we have $\alpha_4 = -\alpha_1 - \alpha_2 - \alpha_3$. Our model becomes

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{41} \\ Y_{42} \\ Y_{43} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T Y} = \begin{pmatrix} \frac{\bar{Y_{1.}} + \bar{Y_{2.}} + \bar{Y_{3.}} + \bar{Y_{4.}}}{4} \\ \bar{Y_{1.}} - \frac{\bar{Y_{1.}} + \bar{Y_{2.}} + \bar{Y_{3.}} + \bar{Y_{4.}}}{4} \\ \bar{Y_{2.}} - \frac{\bar{Y_{1.}} + \bar{Y_{2.}} + \bar{Y_{3.}} + \bar{Y_{4.}}}{4} \\ \bar{Y_{3.}} - \frac{\bar{Y_{1.}} + \bar{Y_{2.}} + \bar{Y_{3.}} + \bar{Y_{4.}}}{4} \end{pmatrix} = \begin{pmatrix} \hat{\mu} \\ \hat{\alpha_1} \\ \hat{\alpha_2} \\ \hat{\alpha_3} \end{pmatrix}$$

2 Question 2

(a) Pooled estimate of variance $= \dfrac{\sum_{i=1}^{r=9}(n_i - 1)S_i^2}{\sum_{i=1}^{r=9} n_i - r} =$

$$\frac{126(0.4979^2) + 43(0.4235^2) + 23(0.3955^2) + 40(0.3181^2) + 17(0.3111^2) + 15(0.4649^2) + 10(0.2963^2) + 6(0.3242^2) + 5(0.5842^2)}{127 + 44 + 24 + 41 + 18 + 16 + 11 + 7 + 6 - 9}$$

$= 0.1919$

(b)

$$\bar{Y}_{..} = \frac{\sum_{i=1}^{r=9} n_i\bar{Y}_{i.}}{\sum_{i=1}^{r=9} n_i} = \frac{127(7.347)+44(7.369)+24(7.428)+41(7.487)+18(7.563)+16(7.568)+11(8.214)+7(8.272)+6(8.297)}{127+44+24+41+18+16+11+7+6}$$

$= 7.4755$

$SS_{model} = \sum_{i=1}^{r=9} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 = 127(7.347-7.4755)^2+44(7.369-7.4755)^2+24(7.428-7.4755)^2+41(7.487-7.4755)^2+$
$18(7.563-7.4755)^2+16(7.568-7.4755)^2+11(8.214-7.4755)^2+7(8.272-7.4755)^2+6(8.297-7.4755)^2 = 17.4197$

$SS_{error} = \sum_{i=1}^{r=9} (n_i-1)S_i^2 = 126(0.4979^2)+43(0.4235^2)+23(0.3955^2)+40(0.3181^2)+17(0.3111^2)+15(0.4649^2)+$
$10(0.2963^2)+6(0.3242^2)+5(0.5842^2) = 54.6956$

ANOVA table:

Source of variation	Degrees of freedom	Sums of squares	Mean square	F
Model	$r-1=8$	$SS_{model}=17.4197$	$\frac{SS_{model}}{r-1}=2.1775$	$\frac{MS_{model}}{MS_{error}}=11.3471$
Error	$N-r=285$	$SS_{error}=54.6956$	$\frac{SS_{error}}{N-r}=0.1919$	
Total	$N-1=293$	$SS_{total}=72.1153$		

(c)

$F = 11.3471$. The F statistic has a central F-distribution with 8 numerator and 285 denominator degrees of freedom.

$P-value = P(F_{8,285} > F) < 0.0001$

(d) At a significance level of 5%, we reject the null that means of the natural logarithm of cavity entrance areas are the same for all nine species.

(e)

Firstly, we have $\bar{Y}_{..} = \frac{\sum_{i=1}^{r} n_i\bar{Y}_{i.}}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} n_i\bar{Y}_{i.}}{N}$. Hence, $\sum_{i=1}^{r} n_i\bar{Y}_{i.} = N\bar{Y}_{..}$

$$\sum_{i=1}^{r} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum_{i=1}^{r} n_i[\bar{Y}_{i.}^{\,2} - 2\bar{Y}_{i.}\bar{Y}_{..} + \bar{Y}_{..}^{\,2}] = \sum_{i=1}^{r} n_i\bar{Y}_{i.}^{\,2} - 2\bar{Y}_{..}\sum_{i=1}^{r} n_i\bar{Y}_{i.} + \bar{Y}_{..}^{\,2}\sum_{i=1}^{r} n_i$$

$$= \sum_{i=1}^{r} n_i\bar{Y}_{i.}^{\,2} - 2N\bar{Y}_{..} + N\bar{Y}_{..} = \sum_{i=1}^{r} n_i\bar{Y}_{i.}^{\,2} - N\bar{Y}_{..}$$

3 Question 3

(a)

$N = 12$; $r = 4$; $n_i = 3$ for $i = 1, 2, 3, 4$

$$\bar{Y}_{..} = \frac{\sum_{i=1}^{r} n_i\bar{Y}_{i.}}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} n_i\bar{Y}_{i.}}{N} = \frac{22*3+25*3+29*3+32*3}{12} = 27$$

Source of variation	Degrees of freedom	Sums of squares	Mean square
Model	$r-1=3$	$SS_{model}=174$	$\frac{SS_{model}}{r-1}=58$
Error	$N-r=8$	$SS_{error}=224$	$\frac{SS_{error}}{N-r}=28$
Total	$N-1=11$	$SS_{total}=398$	

(b)

H_0 : Four group means are equal.

H_a : At least one group mean is different from the rest.

Test statistic: $F = \dfrac{MS_{model}}{MS_{error}} = \dfrac{58}{28} = 2.07$

The test statistic follows central F distribution with 3 numerator and 8 denominator degrees of freedom.

$p - value = P(F_{3,8} > F) = 0.18268$

At a significance level of 5%, we fail to reject the null that means of the four treatment groups are the same.

(c)

$$\boldsymbol{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{41} \\ Y_{42} \\ Y_{43} \end{pmatrix} = \begin{pmatrix} \mu+\alpha_1+\epsilon_{11} \\ \mu+\alpha_1+\epsilon_{12} \\ \mu+\alpha_1+\epsilon_{13} \\ \mu+\alpha_2+\epsilon_{21} \\ \mu+\alpha_2+\epsilon_{22} \\ \mu+\alpha_2+\epsilon_{23} \\ \mu+\alpha_3+\epsilon_{31} \\ \mu+\alpha_3+\epsilon_{32} \\ \mu+\alpha_3+\epsilon_{33} \\ \mu+\alpha_4+\epsilon_{41} \\ \mu+\alpha_4+\epsilon_{42} \\ \mu+\alpha_4+\epsilon_{43} \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \end{pmatrix}$$

Using constraint $\alpha_4 = 0$

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

(d) We set the last group parameter to zero, which means group 4 is considered the baseline group. Now $E(Y_{4j}) = \mu + \alpha_4 = \mu$. Hence, μ represents the mean of group 4 under this constraint. α_2 now is the distance of group 2 from group 4.

4 Question 4

The GLM Procedure

Dependent Variable: velocity

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	94514.0000	23628.5000	4.29	0.0031
Error	95	523510.0000	5510.6316		
Corrected Total	99	618024.0000			

Test statistic F = 4.29, p-value = 0.0031. At a significance level of 5%, we reject the null of equal means for the five trials.

5 Question 5

(a) Functions of contrast are linear combinations of the population means with summation of coefficients equals to 0. Therefore, (i) and (iii) are functions of contrast as both of them are linear functions and sum of the coefficients are 0 ($1+3-4=0$ and $\frac{1}{3}+\frac{1}{3}+\frac{1}{3}-\frac{1}{2}-\frac{1}{2}=0$).

(b) $\gamma_1 = \mu_1 + 3\mu_2 - 4\mu_3 + 0\mu + 4 + 0\mu_5$ and $\gamma_2 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3 - \frac{1}{2}\mu_4 - \frac{1}{2}\mu_5$. We have:

$$\frac{1 * \frac{1}{3}}{n_1} + \frac{3 * \frac{1}{3}}{n_2} + \frac{-4 * \frac{1}{3}}{n_3} + \frac{0 * \frac{-1}{2}}{n_4} + \frac{0 * \frac{-1}{2}}{n_5}.$$

Then, the two functions of contrast will become orthogonal if $n_2n_3 + 3n_1n_3 - 4n_1n_2 = 0$.

(c)

trial=1

Analysis Variable : velocity				
N	Mean	Std Dev	Minimum	Maximum
20	909.0000000	104.9260391	650.0000000	1070.00

trial=2

Analysis Variable : velocity				
N	Mean	Std Dev	Minimum	Maximum
20	856.0000000	61.1641450	760.0000000	960.0000000

trial=3

Analysis Variable : velocity				
N	Mean	Std Dev	Minimum	Maximum
20	845.0000000	79.1068564	620.0000000	970.0000000

trial=4

Analysis Variable : velocity				
N	Mean	Std Dev	Minimum	Maximum
20	820.5000000	60.0416522	720.0000000	920.0000000

trial=5

Analysis Variable : velocity				
N	Mean	Std Dev	Minimum	Maximum
20	831.5000000	54.2193401	740.0000000	950.0000000

For $\gamma_1 = \mu_1 + 3\mu_2 - 4\mu_3 + 0\mu + 4 + 0\mu_5$:

$\hat{\gamma} = \sum_{i=1} c_i \bar{Y}_{i.} = 1 * 909 + 3 * 856 - 4 * 845 = 97$

Assuming σ^2 is not known, and we have $MS_{error} = 5510.6316$.

Then $S_{\hat{\gamma}} = \sqrt{5510.6316 * (\frac{1^2}{20} + \frac{3^2}{20} + \frac{(-4)^2}{20})} = 84.6394$
Also, $N - r = 95$, $1 - \alpha/2 = 0.975$ if we choose $\alpha = 0.05$. Then, $t_{95,0.975} = 1.98525$.
100(1-α)% confidence intervals: $\hat{\gamma} \pm t_{N-r,1-\alpha/2} S_{\hat{\gamma}} = (97 - 1.98525 * 84.6394; 97 + 1.98525 * 84.6394) = (-71.03037; 265.0304)$

For $\gamma_2 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3 - \frac{1}{2}\mu_4 - \frac{1}{2}\mu_5$:
$\hat{\gamma} = \sum_{i=1} c_i \bar{Y}_i = \frac{1}{3} * 909 + \frac{1}{3} * 856 + \frac{1}{3} * 845 - \frac{1}{2} * 820.5 - \frac{1}{2} * 831.5 = 44$
Assuming σ^2 is not known, and we have $MS_{error} = 5510.6316$.
Then $S_{\hat{\gamma}} = \sqrt{5510.6316 * (\frac{\frac{1}{9}}{20} + \frac{\frac{1}{9}}{20} + \frac{\frac{1}{9}}{20} + \frac{\frac{1}{4}}{20} + \frac{\frac{1}{4}}{20})} = 15.1529$
Also, $N - r = 95$, $1 - \alpha/2 = 0.975$ if we choose $\alpha = 0.05$. Then, $t_{95,0.975} = 1.98525$.
100(1-α)% confidence intervals: $\hat{\gamma} \pm t_{N-r,1-\alpha/2} S_{\hat{\gamma}} = (44 - 1.98525 * 15.1529; 44 + 1.98525 * 15.1529) = (13.91771; 74.08229)$