Stat 642 Spring 2022 - Solutions for Assignment 8

Problem 5. (12 points)

1. $\mathbf{C}_1 = \mu_{11} - \mu_{14} - \mu_{31} + \mu_{34} = (\mu_{11} - \mu_{14}) - (\mu_{31} - \mu_{34})$

   a. The contrast, C_1 is an **interaction contrast** because it is the difference between the first and third levels of F_1 of a contrast in the levels of F_2.

   b. The contrast is **estimable** because all four μ_{ij}'s in the contrast have estimates from the data, $n_{ij} > 0$, with $\hat{\mu}_{ij} = \bar{y}_{ij}$.

2. $\mathbf{C}_2 = \mu_{11} + \mu_{21} + \mu_{31} - \mu_{13} - \mu_{23} - \mu_{33} = (\mu_{11} + \mu_{21} + \mu_{31}) - (\mu_{13} + \mu_{23} + \mu_{33}) = \mu_{.1} - \mu^*_{.3}$

   a. The contrast is a **Main Effect contrast** for F_2 consisting of the difference in the first and third levels of F_2 "averaged" over the levels of F_1 (with no data to estimate μ_{13}).

   b. The contrast is **not estimable** because $n_{13} = 0$ which makes μ_{13} non-estimable, $\bar{y}_{13}$ was not observed in the data.

3. $\mathbf{C}_5 = \mu_{11} - \mu_{14} + \mu_{21} - \mu_{24} + \mu_{31} - \mu_{34} = (\mu_{11} + \mu_{21} + \mu_{31}) - (\mu_{14} + \mu_{24} + \mu_{34}) = \mu_{.1} - \mu^*_{.4}$

   a. The contrast is a **Main Effect contrast** for F_2 consisting of the difference in the first and fourth levels of F_2 "averaged" over the levels of F_1 (with no data to estimate μ_{24}).

   b. The contrast is **not estimable** because $n_{24} = 0$ which makes μ_{24} non-estimable, $\bar{y}_{24}$ was not observed in the data..

Problem 6. (5 points)

a. Two contrasts which evaluate the Main Effect of F_1, contrasts in levels of F_1 averaged over levels of F_2:

Contrast	μ_{11}	μ_{13}	μ_{14}	μ_{22}	μ_{23}	μ_{24}	μ_{31}	μ_{33}	μ_{34}
C_1	1	1	1	0	0	0	-1	-1	-1
C_2	0	1	1	0	-2	-2	0	1	1

$C_1 = (\mu_{11} + \mu_{13} + \mu_{14}) - (\mu_{31} + \mu_{33} + \mu_{34}) = \mu^*_{1.} - \mu^*_{3.}$ is comparing the 1 and 3 levels of factor F_1 averaged over the levels of factor F_2 but with the second level of factor F_2, μ_{12} and μ_{32} missing from the averages.

$C_2 = (\mu_{13} + \mu_{14}) - 2(\mu_{23} + \mu_{24}) + (\mu_{33} + \mu_{34}) = \mu^*_{1.} - 2\mu^*_{2.} + \mu^*_{3.}$ is a contrast in the means of the three levels of F_1 but with the 1 and 2 levels of F_2 missing from the averages.

The two contrasts C_1 and C_2 are orthogonal:

$$(1)(0) + (1)(1) + (1)(1) + (0)(0) + (0)(-2) + (0)(-2) + (-1)(0) + (-1)(1) + (-1)(1) = 0$$

b. Two contrasts which evaluate the Interaction between F_1 and F_2, $F_1 \times F_2$, first comparing contrasts in levels of F_2 at two levels of F_1 and then comparing contrasts in levels of F_1 at two levels of F_2:

Contrast	μ_{11}	μ_{13}	μ_{14}	μ_{22}	μ_{23}	μ_{24}	μ_{31}	μ_{33}	μ_{34}
C_3	1	1	-2	0	0	0	-1	-1	2
C_4	0	1	-1	0	-2	2	0	1	-1

$C_3 = (\mu_{11} + \mu_{13} - 2\mu_{14}) - (\mu_{31} + \mu_{33} - 2\mu_{34})$ is comparing a contrast with coefficients (1,1,-2) in the 1, 3 and 4 levels of factor F_2 at the 1 and 3 levels of factor F_1

$C_4 = (\mu_{13} - 2\mu_{23} + \mu_{33}) - (\mu_{14} - 2\mu_{24} + \mu_{34})$ is comparing a contrast (1,-2,1) in the 1, 2, and 3 levels of factor F_1 at the 3 and 4 levels of factor F_2

The two contrasts C_3 and C_4 are orthogonal:

$(1)(0) + (1)(1) + (-2)(-1) + (0)(0) + (0)(-2) + (0)(2) + (-1)(0) + (-1)(1) + (2)(-1) = 0$

Problem III. (30 points) Traffic Engineering Study:

1. Model : $y_{ijk\ell} = \mu + \alpha_i + \beta_j + I_{k(i)} + (\alpha\beta)_{ij} + (\beta I)_{jk(i)} + \gamma_\ell + (\alpha\gamma)_{i\ell} + (\beta\gamma)_{j\ell} + (\gamma I)_{\ell k(i)} + (\alpha\beta\gamma)_{ij\ell} + e_{ijk\ell},\ i = 1, 2, 3,\ j = 1, 2,\ k = 1, 2,\ \ell = 1, 2$, where
   - α_i is the fixed effect for signal type with $\alpha_3 = 0$
   - β_j is the fixed effect for level of traffic with $\beta_2 = 0$
   - γ_ℓ is the fixed effect for method of measuring with $\gamma_2 = 0$
   - $(\alpha\gamma)_{i\ell}$ is the fixed interaction effect between signal type and method of measuring with $(\alpha\gamma)_{3\ell} = 0$ and $(\alpha\gamma)_{i2} = 0$
   - $(\beta\gamma)_{j\ell}$ is the fixed interaction effect between level of traffic and method of measuring with $(\beta\gamma)_{2\ell} = 0$ and $(\beta\gamma)_{j2} = 0$
   - $(\alpha\beta\gamma)_{ij\ell}$ is the fixed interaction effect between signal type, level of traffic and method of measuring with $(\alpha\beta\gamma)_{3j\ell} = 0$, $(\alpha\beta\gamma)_{i2\ell} = 0$ and $(\alpha\beta\gamma)_{ij2} = 0$
   - $I_{k(i)}$ is the random effect due to intersection nested within signal type
   - $(\beta I)_{jk(i)}$ is the random effect due to interaction between intersection nested within signal type and traffic level
   - $(\gamma I)_{\ell k(i)}$ is the random effect due to interaction between intersection nested within signal type and method of measurement
   - $I_{k(i)},\ (\beta I)_{jk(i)},\ (\gamma I)_{\ell k(i)}$ and $e_{ijk\ell}$ are independent
   - $I_{k(i)} \sim iid\ N(0, \sigma^2_{I(S)}),\ (\beta I)_{jk(i)} \sim iid\ N(0, \sigma^2_{T*I(S)}),\ (\gamma I)_{\ell k(i)} \sim iid\ N(0, \sigma^2_{M*I(S)})$ and $e_{ijk\ell} \sim iid\ N(0, \sigma^2_e)$
   - Note that the error term $e_{ijk\ell}$ is equivalent to $I_{h(i)} \times (\beta\gamma)_{j\ell}$, that is, Error = I(S)*T*M
2. AOV Table - From SAS-MIXED output: S=Signal Type, T=Traffic level, I=Intersection, M=Method

The SAS System

The Mixed Procedure

Model Information	
Data Set	WORK.DATA8
Dependent Variable	Y
Covariance Structure	Variance Components
Estimation Method	Type 3
Residual Variance Method	Factor
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information		
Class	Levels	Values
S	3	F P S
M	2	PS PT
T	2	NR R
I	2	1 2

Dimensions	
Covariance Parameters	4
Columns in X	36
Columns in Z	30
Subjects	1
Max Obs per Subject	24

Number of Observations	
Number of Observations Read	24
Number of Observations Used	24
Number of Observations Not Used	0

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
S	2	3143.023333	1571.511667	Var(Residual) + 2 Var(M*I(S)) + 2 Var(T*I(S)) + 4 Var(I(S)) + Q(S,S*T,S*M,S*M*T)	MS(I(S))	3	2.30	0.2484
T	1	236.881667	236.881667	Var(Residual) + 2 Var(T*I(S)) + Q(T,S*T,M*T,S*M*T)	MS(T*I(S))	3	7.37	0.0728
S*T	2	275.773333	137.886667	Var(Residual) + 2 Var(T*I(S)) + Q(S*T,S*M*T)	MS(T*I(S))	3	4.29	0.1318
M	1	96.000000	96.000000	Var(Residual) + 2 Var(M*I(S)) + Q(M,S*M,M*T,S*M*T)	MS(M*I(S))	3	3.00	0.1817
S*M	2	51.430000	25.715000	Var(Residual) + 2 Var(M*I(S)) + Q(S*M,S*M*T)	MS(M*I(S))	3	0.80	0.5255
M*T	1	31.740000	31.740000	Var(Residual) + Q(M*T,S*M*T)	MS(Residual)	3	7.61	0.0702
S*M*T	2	11.970000	5.985000	Var(Residual) + Q(S*M*T)	MS(Residual)	3	1.44	0.3652
I(S)	3	2053.050000	684.350000	Var(Residual) + 2 Var(M*I(S)) + 2 Var(T*I(S)) + 4 Var(I(S))	MS(T*I(S)) + MS(M*I(S)) - MS(Residual)	5.2014	11.41	0.0101
T*I(S)	3	96.370000	32.123333	Var(Residual) + 2 Var(T*I(S))	MS(Residual)	3	7.71	0.0638
M*I(S)	3	96.015000	32.005000	Var(Residual) + 2 Var(M*I(S))	MS(Residual)	3	7.68	0.0641
Residual	3	12.505000	4.168333	Var(Residual)	.	.	.	.

Covariance Parameter Estimates							
Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
I(S)	156.10	140.00	1.11	0.2649	0.05	-118.30	430.49
T*I(S)	13.9775	13.2242	1.06	0.2905	0.05	-11.9415	39.8965
M*I(S)	13.9183	13.1763	1.06	0.2908	0.05	-11.9068	39.7435
Residual	4.1683	3.4034	1.22	0.1103	0.05	1.3377	57.9484

3. AOV ($S-3, T-2, I=2, M-2$)

SV	Q_S	Q_T	$\sigma^2_{I(S)}$	Q_{S*T}	$\sigma^2_{T*I(S)}$	Q_M	Q_{S*M}	Q_{T*M}	$\sigma^2_{M*I(S)}$	Q_{T*S*M}	σ^2_e
S	8	0	4	0	2	0	0	0	2	0	1
T	0	12	0	0	2	0	0	0	0	0	1
I(S)	0	0	0	0	2	0	0	0	2	0	1
S*T	0	0	0	4	2	0	0	0	0	0	1
T*I(S)	0	0	0	0	2	0	0	0	0	0	1
M	0	0	0	0	0	12	0	0	2	0	1
S*M	0	0	0	0	0	0	4	0	2	0	1
T*M	0	0	0	0	0	0	0	6	0	0	1
M*I(S)	0	0	0	0	0	0	0	0	2	0	1
S*T*M	0	0	0	0	0	0	0	0	0	2	1
Error	0	0	0	0	0	0	0	0	0	0	1

Source	df	EMS
S	2	$\sigma^2_e + 2\sigma^2_{M*I(S)} + 2\sigma^2_{T*I(S)} + 4\sigma^2_{I(S)} + 8Q_S$
T	1	$\sigma^2_e + 2\sigma^2_{T*I(S)} + 12Q_T$
I(S)	3	$\sigma^2_e + 2\sigma^2_{M*I(S)} + 2\sigma^2_{T*I(S)} + 4\sigma^2_{I(S)}$
S*T	2	$\sigma^2_e + 2\sigma^2_{T*I(S)} + 4Q_{S*T}$
T*I(S)	3	$\sigma^2_e + 2\sigma^2_{T*I(S)}$
M	1	$\sigma^2_e + 2\sigma^2_{M*I(S)} + 12Q_M$
S*M	2	$\sigma^2_e + 2\sigma^2_{M*I(S)} + 4Q_{S*M}$
T*M	1	$\sigma^2_e + 6Q_{T*M}$
M*I(S)	3	$\sigma^2_e + 2\sigma^2_{M*I(S)}$
S*T*M	2	$\sigma^2_e + 2Q_{S*T*M}$
Error	3	σ^2_e
Total	23	

4. From AOV table above, we conclude there is significant evidence of an effect due to intersections nested within signal type but all other effects are not significant.

5. $\hat{\sigma}^2_{I(S)} = 156.10\ (83.0\%)$; $\hat{\sigma}^2_{T*I(S)} = 13.98\ (7.4\%)$; $\hat{\sigma}^2_{M*I(S)} = 13.92\ (7.4\%)$; $\hat{\sigma}^2_e = 4.17\ (2.2\%)$,

Problem IV: (30 points) Factor A has 4 randomly selected levels, Factor B has 5 fixed levels, Factor C has 3 randomly selected levels at each level of Factor B (C is nested within B), and there are 6 EU's at each of the t=60 treatments:

1.

Source	DF	**MS**	**Expected Mean Squares**
A	3	24.5	$\sigma^2_e + 6\sigma^2_{A*C(B)} + 18\sigma^2_{A*B} + 90\sigma^2_A$
B	4	19.7	$\sigma^2_e + 6\sigma^2_{A*C(B)} + 18\sigma^2_{A*B} + 24\sigma^2_{C(B)} + 72Q_B$
$A \times B$	12	8.9	$\sigma^2_e + 6\sigma^2_{A*C(B)} + 18\sigma^2_{A*B}$
$C(B)$	10	7.5	$\sigma^2_e + 6\sigma^2_{A*C(B)} + 24\sigma^2_{C(B)}$
$A \times C(B)$	30	6.8	$\sigma^2_e + 6\sigma^2_{A*C(B)}$
Error	300	5.8	σ^2_e

2. Test for a significant AB interaction ($\alpha = 0.05$). Note that the AOV table is providing the MS, not SS for each source of variation.

   Test $H_o : \sigma^2_{AB} = 0$ vs $H_1 : \sigma^2_{AB} > 0$. Using the expected MS, we have the following test statistic:

   $F = \frac{MS_{AB}}{MS_{A*C(B)}} = \frac{8.9}{6.8} = 1.309 < 2.09 = F_{.05,12,30}$ and $p - value = 1 - pf(1.309, 12, 30) = .2644 > .05 \quad \Rightarrow$

   There is not significant evidence that $\sigma_{AB} > 0$

3. Test for a significant B main effect ($\alpha = 0.05$): $H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ vs H_1 : difference in μ_i's or equivalently, test $H_o : Q_B = 0$ vs $H_1 : Q_B \neq 0$ Examining the EMS' for Factor B with $Q_B = 0$, there is no other EMS which matches it. Thus, we must create a linear combination of several MS's:

   Let $M = MS_{AB} + MS_{C(B)} - MS_{AC(B)} = 9.6$. When $Q_B = 0, E[M] = E[MS_B]$, thus the appropriate test statistic is

   $F = \frac{MS_B}{M} = \frac{19.7}{9.6} = 2.052$ with $p - value = pf(2.052, 4, 6.6942) = .1952 > \alpha = .05 \quad \Rightarrow$ There is not significant evidence that $Q_B \neq 0$, that is, there is not significant evidence of a difference in the 5 treatment means associated with the levels of Factor B.

   The df for the F-test are obtained from the Satterthwaite approximation are obtained as follows:

   $df_M = \frac{(9.6)^2}{\frac{(8.9)^2}{12} + \frac{(7.5)^2}{10} + \frac{(6.8)^2}{30}} = 6.6942$

4. Compute the variance of the difference in treatment means for levels 1 and 2 of Factor B:

   $y_{ijkl} = \mu + a_i + \beta_j + c_{k(j)} + (a\beta)_{ij} + (ac)_{ik(j)} + e_{ijkl} \Rightarrow$

   $\bar{y}_{.1..} = \mu + \bar{a}_{.} + \beta_1 + \bar{c}_{.(1)} + (\bar{a\beta})_{.1} + (\bar{ac})_{..(1)} + \bar{e}_{.1..}$

   $\bar{y}_{.2..} = \mu + \bar{a}_{.} + \beta_2 + \bar{c}_{.(2)} + (\bar{a\beta})_{.2} + (\bar{ac})_{..(2)} + \bar{e}_{.2..} \Rightarrow$

   $Var[\bar{y}_{.1..} - \bar{y}_{.2..}] = Var(\bar{c}_{.(1)} - \bar{c}_{.(2)}) + Var((\bar{a\beta})_{.1} - (\bar{a\beta})_{.2}) + Var((\bar{ac})_{..(1)} - (\bar{ac})_{..(2)}) + Var(\bar{e}_{.1..} - \bar{e}_{.2..})$

   Therefore, we have

   $$\begin{aligned} Var[\bar{y}_{.1..} - \bar{y}_{.2..}] &= \frac{2\sigma^2_{C(B)}}{3} + \frac{2\sigma^2_{AB}}{4} + \frac{2\sigma^2_{AC(B)}}{12} + \frac{2\sigma^2_e}{72} \\ &= 2\left[\frac{24\sigma^2_{C(B)} + 18\sigma^2_{AB} + 6\sigma^2_{AC(B)} + \sigma^2_e}{72}\right] \\ &= \frac{2[EMS_{AB} + EMS_{C(B)} - EMS_{AC(B)}]}{72} \end{aligned}$$

   Provide an estimate of this variance and the degrees of freedom of the estimate.

   $\widehat{Var}[\bar{y}_{.1..} - \bar{y}_{.2..}] = \frac{2[MS_{AB} + MS_{C(B)} - MS_{AC(B)}]}{72} = \frac{2[8.9+7.5-6.8]}{72} = \frac{2[9.6]}{72} = 0.2667.$

   Using the Sattherwaite approximation: $df_M \approx \frac{(8.9+7.5-6.8)^2}{\frac{(8.9)^2}{12} + \frac{(7.5)^2}{10} + \frac{(6.8)^2}{30}} = 6.6942.$

5. Compute the value of Tukey-Kramer HSD with $\alpha = .05$ that would be used to determine which pairs of means across the levels of Factor B are different:

$$HSD = q_{(.05,t,\nu)}\sqrt{\frac{1}{2}\widehat{Var}(\hat{\mu}_1 - \hat{\mu}_2)} = q_{(.05,5,6.6942)}\sqrt{\frac{1}{2}\frac{2[MS_{AB}+MS_{C(B)}-MS_{AC(B)}]}{72}} = 5.1257\sqrt{\frac{9.6}{72}} = 1.87$$

where $q_{(.05,5,6.6942)} = qtukey(.95, 5, 6.6942) = 5.1257$ using the r-function qtukey.

Problem V. (23 points)

1. Factor A = Patients, have random effects, Factor B = Runs, with random effects and nested within Patients, and e-Tubes (error term);

   model $y_{ijk} = \mu + a_i + b_{j(i)} + e_{ijk}$; $a = 5, b = 4, r = 2$:

DF	SV	σ^2_A	$\sigma^2_{B(A)}$	σ^2_e	EMS
4	A	8	2	1	$\sigma^2_e + 2\sigma^2_{b(a)} + 8\sigma^2_a$
15	B(A)	0	2	1	$\sigma^2_e + 2\sigma^2_{b(a)}$
20	e(A,B)	0	0	1	σ^2_e

2. A is random, B is random and nested within A; C is random and nested within B; D is random and nested within C; $a = 4, b = 3, c = 2, d = 3$:

   Model : $y_{ijk\ell} = \mu + a_i + b_{j(i)} + c_{k(i,j)} + d_{\ell(ijk)},\ i = 1,2,3,4\ j = 1,2,3\ k = 1,2,\ \ell = 1,2,3,$

DF	SV	σ^2_A	$\sigma^2_{B(A)}$	$\sigma^2_{C(AB)}$	σ^2_e	EMS
3	A	18	6	3	1	$\sigma^2_d + 3\sigma^2_{C(AB)} + 6\sigma^2_{B(A)} + 18\sigma^2_A$
8	B(A)	0	6	3	1	$\sigma^2_d + 3\sigma^2_{C(AB)} + 6\sigma^2_{B(A)}$
12	C(A,B)	0	0	3	1	$\sigma^2_d + 3\sigma^2_{C(AB)}$
48	D(A,B,C)	0	0	0	1	σ^2_d

3. A,B,D are fixed, C is random nested within A and B and A,B, and D are crossed:

   The model is given by

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + c_{k(i,j)} + \delta_l + (\alpha\delta)_{il} + (\beta\delta)_{jl} + (\alpha\beta\delta)_{ijl} + (c\delta)_{lk(i,j)} + e_{m(i,j,k,l)}$$

$$\text{with}\ \ i = 1,2,3;\ \ j = 1,2;\ \ k = 1,2,3,4,5,6;\ \ l = 1,2,3,4,5;\ \ m = 1,2,3,4,5,6$$

DF	SV	Q_A	Q_B	Q_{A*B}	$\sigma^2_{C(A,B)}$	Q_D	Q_{A*D}	Q_{B*D}	Q_{A*B*D}	$\sigma^2_{D*C(A,B)}$	σ^2_e
2	A	360	0	0	30	0	0	0	0	6	1
1	B	0	540	0	30	0	0	0	0	6	1
2	A*B	0	0	180	30	0	0	0	0	6	1
30	C(A,B)	0	0	0	30	0	0	0	0	6	1
4	D	0	0	0	0	216	0	0	0	6	1
8	A*D	0	0	0	0	0	72	0	0	6	1
4	B*D	0	0	0	0	0	0	108	0	6	1
8	A*B*D	0	0	0	0	0	0	0	36	6	1
120	C(A,B)*D	0	0	0	0	0	0	0	0	6	1
900	e(A,B,C,D)	0	0	0	0	0	0	0	0	0	1

$$
\begin{aligned}
E(MS_A) &= \sigma_e^2 + 6\sigma^2_{D*C(A,B)} + 30\sigma^2_{C(A,B)} + 360Q_A \\
E(MS_B) &= \sigma_e^2 + 6\sigma^2_{D*C(A,B)} + 30\sigma^2_{C(A,B)} + 540Q_B \\
E(MS_{A*B}) &= \sigma_e^2 + 6\sigma^2_{D*C(A,B)} + 30\sigma^2_{C(A,B)} + 180Q_{A*B} \\
E(MS_{C(A,B)}) &= \sigma_e^2 + 6\sigma^2_{D*C(A,B)} + 30\sigma^2_{C(A,B)} \\
E(MS_D) &= \sigma_e^2 + 6\sigma^2_{D*C(A,B)} + 216Q_D \\
E(MS_{A*D}) &= \sigma_e^2 + 6\sigma^2_{D*C(A,B)} + 72Q_{A*D} \\
E(MS_{B*D}) &= \sigma_e^2 + 6\sigma^2_{D*C(A,B)} + 108Q_{B*D} \\
E(MS_{A*B*D}) &= \sigma_e^2 + 6\sigma^2_{D*C(A,B)} + 36Q_{A*B*D} \\
E(MS_{C*D(A,B)}) &= \sigma_e^2 + 6\sigma^2_{D*C(A,B)} \\
E(MSE) &= \sigma_e^2
\end{aligned}
$$