STATISTICS 642 - FINAL EXAM - May 5, 2022 - SOLUTIONS

Problem I. (30 points) For the following experiment, provide the requested information:

A study was designed to evaluate the impact of chemical plant discharges into Michigan lakes on the uptake of carcinogens by game fish. The four most widely caught species (S1, S2, S3, S4) of fish were selected for this study. The biologist decided to assess the fish's uptake of carcinogens by feeding mice portions of fish known to have been contaminated with the carcinogens. The biologist were also interested in assessing three different methods of determining the amount of carcinogen in the bladder of mice. To control for the variation in bladder density from mouse to mouse, the biologists randomly selected four mice from each of five litters of mice for use in the study. Within each litter, the four mice were randomly assigned to the four species of fish, one mouse to each species. The mice were fed a diet of fish for six weeks and then were sacrificed. The bladder of each mouse was divided into three sections and the sections randomly assigned to one of the three assessment techniques (A1, A2, A3). The amount of carcinogen in the bladder sections (ppb) was recorded.

1. D
   - **RCBD with Litters as blocks and a Split-Plot Treatment Assignment**
2. D
   - **Species of Fish crossed with Assessment Technique (4x3, with Blocks being the Litters)**
3. B,C
   - **Species- 4 Fixed levels; Assessment Method- 3 Fixed levels**
4. A
   - **Litters- 5 Random levelss**
5. B
   - **EU for Species is Mouse; EU for Assessment Method is portion of mouse's bladder which is also the MU**
6. A
   - **The response is the amount of carcinogen in a section of a mouse's bladder. There are no covariates, subsampling, or repeated measures.**
7. A
   - $\tau_a = 0$, All students have the credit for this question
8. B
9. B
7. D There are two out of three factors are random, all three two way and one three way interactions additional to error are random.

Problem II (31 points.) An industrial engineer designs a study which will hopefully improve the speed of the assembly operation of inserting electronic components on circuit boards. She has decided on three types of assembly fixtures (F1, F2, F3) and two types of workplace layouts (L1, L2) for investigation. Operators perform the assembly. Because the two workplace layouts are in two different locations, the engineer decides to randomly select four operators at each of the two workplace layouts for a total of 8 operators in the study. The times (minutes) needed for each operator to assembly two circuit boards for each of the three Fixture types are given below.

1. Determine the degrees of freedom (DF) and the expected mean squares (EMS) for each of the sources of variance in the following AOV table. Note: The **Mean Squares (MS)** are provided in the AOV table not the Sum of Squares. The notation F=Fixture, L=Assembly Layout, O=Operator is used in the AOV table.

SOURCE	DF	MS	EMS
F	2	41.40	$\sigma_e^2 + 2\sigma^2_{F*O(L)} + 16Q_F$
L	1	4.09	$\sigma_e^2 + 2\sigma^2_{F*O(L)} + 6\sigma^2_{O(L)} + 24Q_L$
O(L)	6	11.99	$\sigma_e^2 + 2\sigma^2_{F*O(L)} + 6\sigma^2_{O(L)}$
F*L	2	9.52	$\sigma_e^2 + 2\sigma^2_{F*O(L)} + 8Q_{F*L}$
F*O(L)	12	5.49	$\sigma_e^2 + 2\sigma^2_{F*O(L)}$
ERROR	24	2.33	σ_e^2
TOTAL	47		

2. At the $\alpha = 0.05$ level, are there significant effects of Fixture and Assembly Layout on the time to complete a circuit board? Justify your answer.

- $H_o : Q_{F*L} = 0$ vs $H_a : Q_{F*L} \neq 0$: $F = \frac{MS_{F*L}}{MS_{F*O(L)}} = \frac{9.52}{5.49} = 1.73 < 3.89 = F_{.05,2,12} \Rightarrow$

  There is not significant evidence of an interaction between Fixture and Layout.

- $H_o : Q_F = 0$ vs $H_a : Q_F \neq 0$: $F = \frac{MS_F}{MS_{F*O(L)}} = \frac{41.40}{5.49} = 7.54 > 3.89 = F_{.05,2,12} \Rightarrow$

  There is significant evidence of a difference in the mean responses across the three Fixture Types.

- $H_o : Q_L = 0$ vs $H_a : Q_L \neq 0$: $F = \frac{MS_L}{MS_{O(L)}} = \frac{4.08}{11.99} = 0.34 < 5.99 = F_{.05,1,6} \Rightarrow$

  There is not significant evidence of a difference in the mean responses across the two Layouts.

3. The following model was fit to the data where $y_{ijk\ell}$ is the time needed to assemble

   circuit board ℓ by Operator k using Fixture i in Layout j :

$$y_{ijk\ell} = \mu \ + \ \tau_i \ + \ \gamma_j \ + \ b_{k(j)} \ + \ (\tau\gamma)_{ij} \ + \ d_{ik(j)} \ + \ e_{ijk\ell}$$

Using the numeric values of the MS's and EMS's given in the AOV table, compute a 95% confidence interval for the mean time to assemble a circuit board using Fixture 3:

- Estimate the standard error of the estimated mean time to assemble a circuit board using Layout 2:

$$Var[\bar{y}_{.2..}] = Var[\bar{b}_{.(2)} + \bar{d}_{..(2)} + \bar{e}_{.2..}] = \frac{\sigma^2_{O(L)}}{4} + \frac{\sigma^2_{F*O(L)}}{12} + \frac{\sigma^2_e}{24} = \frac{6\sigma^2_{O(L)} + 2\sigma^2_{F*O(L)} + \sigma^2_e}{24} = \frac{EMS_{O(L)}}{24} \Rightarrow$$

$$\widehat{SE}\left(\hat{\mu}_{.2.}\right) = \sqrt{\frac{MS_{O(L)}}{24}} = \sqrt{\frac{11.99}{24}} = 0.7068 \text{ with df} = 6$$

A 95% C.I. on $\bar{\mu}_{.2.}$ is given by $\hat{\mu}_{.2.} \pm t_{.025,6}\widehat{SE}\left(\hat{\mu}_{.2.}\right) = 26.38 \pm (2.447)(.7068) = 26.38 \pm 1.73 = (24.65,\ 28.11)$

Problem III. (12 points) A simulation model of an inventory system involved 7 factors: A, B, C, D, E, F, G, each having two levels. The evaluation of the model used a single replication of a 2^{7-3} fractional factorial design. The generators $ABDG = -1$, $ACDF = +1$, and $BCDE = -1$ were used to select the treatments to be included in the study.

1. How many treatments would be observed in this experiment?

- $t = 2^{7-3} = 2^4 = 16$

2. Would the following treatment $(A, B, C, D, E, F, G) = (H, L, H, H, L, H, H) = (+, -, +, +-, +, +)$ be used in the experiment? Justify your answer.

- ABDG = (+)(-)(+)(+) = - ; ACDF = (+)(+)(+)(+) = + ; BCDE = (-)(+)(+)(-) = + $\neq -1$

  Therefore the treatment $(A, B, C, D, E, F, G) = (+, -, +, +-, +, +)$ will not be used in the experiment.

3. What is the resolution of this design? Justify your answer.

- The implicit generators are

  ABDG*ACDF = BCFG; ABDG*BCDE = ACEG; ACDF*BCDE = ABEF;

  ABDG*ACDF*BCDE = DEFG

  The minimum length of the three generators and four implicit generators is 4, thus, the Resolution is IV.

4. Which of the interactions would need to be 0 in order to be able to estimate Factor A? Justify your answer.

- The main effect of factor A is confounded with all the members of its alias set. Membership in the set is obtained by multiplying the generators and implicit generators by A yielding
- I = ABDG = ACDF = BCDE = BCFG = ACEG = ABEF = DEFG
- A = BDG = CDF = ABCDE = ABCFG = CEG = BEF = ADEFG
- All of the interactions BDG, CDF, ABCDE, ABCFG, CEG, BEF, ADEFG would have to be negligible in order to be able to estimate the main effect of factor A.

Problem IV. (27 points) **Circle** one of (A, B, C, D, E) corresponding to the **BEST** answer.

(1.) **B** - If the groups of levels of factor A obtained from the multiple comparison procedure differ across the levels of factor B, then there is strong evidence of an interaction.

(2.) **C** - B is nested within A and B(A) does not interact with C

(3.) **A** -

- $L = -3\mu_{11} - \mu_{21} + \mu_{31} + 3\mu_{41} - 3\mu_{12} - \mu_{22} + \mu_{32} + 3\mu_{42} = -3\mu_{1.} - \mu_{2.} + \mu_{3.} + 3\mu_{4.}$

(4.) **D** - Positive correlation inflates the probability of Type I error over its nominal value

(5.) **B** -definition of non-estimable

(6.) **C** - The second level of Factor F1 appears with the third level of Factor F2 but not with the second level of Factor F2. Thus, including μ_{23} would be a biased comparison of the two levels of Factor F2. choice has the more appropriate contrast

(7.) **A** - There is essentially an interaction between the 5 treatments and covariate. Thus, inferences about differences in the 5 treatments must be made at specific values of the covariate.

(8.) **D** - Correlation is potentially present in the multiple measurements on the same EU. Thus, the random term involving both ℓ and k would represent the multiple measurements on the ℓ EU during time period k, that is, the $e_{ijk\ell}$ terms would be correlated.

(9.) **C**

- $RE = K + (1-K)H$ with $K = \frac{r(t-1)}{rt-1} = \frac{10(4-1)}{10*4-1} = \frac{30}{39}$, $H = \frac{MS_{Block}}{MSE} = \frac{36}{12} = 3$

$$RE = \frac{30}{39} + (1 - \frac{30}{39})3 = 1.46 \;\Rightarrow\; n_{CRD} = (RE) * (n_{RCBD}) = (1.46)(10) = 14.6 \approx 15$$

Summary of FINAL EXAM SCORES

Minimum= 47 ; Q(.25)= 48.2 ; Mean= 79.07 ; Median = 82 ; Q(.75)= 87 ; Maximum= 98