STAT 642 - SOLUTIONS - FINAL EXAMINATION - May 5, 2020

Problem I. (25 points) The EPA designed a study to investigate the impact of excess nitrogen and air pollution on plant growth in wetland areas. The researchers selected four levels of nitrogen (N_1-N_4), three types of particulates (P_1-P_3) found in air pollution, and two exposure times (E_1, E_2). The experiment was conducted in an artificial setting within a greenhouse. Eight trays were used in the study with each tray holding three artificial wetlands. All of the artificial wetlands receive a standard set of seeds to start growth. Four of the trays are placed on a table located at the north end of the greenhouse and the other four trays are placed on a table located at the south end of the greenhouse. Separately for each table, the four trays are randomly assigned to the four levels of nitrogen. Within each tray, the three wetlands are randomly assigned to the three types of particulates. Each wetland is then split in half with one half randomly assigned to be low exposure time, E1; the other half to the high exposure time, E2. At the end of 8 months, the researcher determined the amount of biomass in each wetland. The weight (kg) of biomass are given here with notation: Tray (TR), Nitrogen (N), Particulate (P), and Exposure (E).

For the experiment, provide the requested information:

1. Type of Randomization: CRD, RCBD, LSD, Split-Plot, Crossover, etc.

- RCBD (Tables are blocks) with a Split-Split Plot treatment assignment

2. Type of Treatment Structure: single factor, crossed, nested, fractional, etc.

- 4x3x2 crossed treatments with a split plot treatment assignment
- Whole Plot Factor - Nitrogen crossed with Split-Plot Factor - Particulate crossed with Split-Split-Plot Factor - Exposure

3. Identify each of the Factors as being Fixed or Random:

- Factors with Fixed Levels: Nitrogen, Particulate, Exposure; Tables have random levels

4. Describe the Experimental Units and Measurement Units:

- $EU_{WholePlot}$ = Tray; $EU_{SplitPlot}$ = Wetland; $EU_{SplitSplitPlot}$ = Half-Wetland; MU = Half-Wetland

5. Describe the Measurement Process: Response Variable, Covariates, SubSampling, Repeated Measures

- Response is the biomass in half-wetland; no covariates, no subsampling, no repeated measures

Problem II (30 points.) A structural engineer is studying the strength of aluminum alloy purchased from the three largest vendors. Each vendor submits the alloy in standard-sized bars, either 1, 2, or 3 inches. The processing of different sizes of bar stock from a common ingot involves different forging techniques, and this factor may be important. Furthermore, the bar stock is forged from ingots made in different batches. Each vendor submits two test specimens of each size bar stock from 3 batches. The strength of each bar is determined and is reported in the following table. The three vendors are the only vendors under consideration, the batches are randomly selected from the Vendors production output.

1. Complete the following AOV table. Note: The Mean Square (MS) are provided in the table not the Sum of Squares and we will use the notation S=Bar Size, V=Vendor, B=Batch.

SOURCE	DF	MS	EMS
S	2	0.1263	$\sigma_e^2 + 2\sigma^2_{S*B(V)} + 18Q_S$
V	2	0.4424	$\sigma_e^2 + 2\sigma^2_{S*B(V)} + 6\sigma^2_{B(V)} + 18Q_V$
S*V	4	0.0594	$\sigma_e^2 + 2\sigma^2_{S*B(V)} + 6Q_{S*V}$
B(V)	6	1.6702	$\sigma_e^2 + 2\sigma^2_{S*B(V)} + 6\sigma^2_{B(V)}$
S*B(V)	12	0.0919	$\sigma_e^2 + 2\sigma^2_{S*B(V)}$
ERROR	27	0.0404	σ_e^2
TOTAL	53		

2. At the $\alpha = 0.05$ level, evaluate the effect of Vendor and Bar Size on the strength of the aluminum alloy.

- $H_o : Q_{S*V} = 0$ vs $H_a : Q_{S*V} \neq 0$ $F = \frac{MS_{S*V}}{MS_{S*B(V)}} = \frac{.0594}{.0919} = 0.65 < 3.26 = F_{.05,4,12}$ $\Rightarrow$ There is not significant evidence of an interaction between Vendor and BarSize, p-value=.638.
- $H_o : Q_S = 0$ vs $H_a : Q_S \neq 0$ $F = \frac{MS_S}{MS_{S*B(V)}} = \frac{.1263}{.0919} = 1.37 < 3.89 = F_{.05,2,12}$ $\Rightarrow$ There is not significant evidence of a difference in the mean responses across the three Bar Sizes, p-value=.291.
- $H_o : Q_V = 0$ vs $H_a : Q_V \neq 0$ $F = \frac{MS_V}{MS_{B(V)}} = \frac{.4424}{1.6702} = 0.26 < 5.14 = F_{.05,2,6}$ $\Rightarrow$ There is not significant evidence of a difference in the mean responses across the three Vendors, p-value=.779.

3. Using the numeric values of the MS's given in the AOV table and your EMS's, find the values of the following quantities where $y_{ijk\ell}$ is the strength of bar ℓ of Size j made using material from Batch k of Vendor i's aluminum:

$$y_{ijk\ell} = \mu + \tau_i + a_{k(i)} + \gamma_j + (\tau\gamma)_{ij} + b_{jk(i)} + e_{ijkl} \text{ with } i = 1,2,3;\ j = 1,2,3;\ k = 1,2,3;\ \ell = 1,2$$

Estimate the standard error of the estimated difference in the mean strength of the bars produced by Vendors I and II:

- $Var(\hat{\mu}_{1.} - \hat{\mu}_{2.}) = Var[\bar{y}_{1...} - \bar{y}_{2...}] = Var[(\bar{a}_{.(1)} - \bar{a}_{.(2)}) + (\bar{b}_{..(1)} - \bar{b}_{..(2)}) + (\bar{e}_{1...} - \bar{e}_{2...})]$

  $= \frac{2\sigma^2_{B(V)}}{3} + \frac{2\sigma^2_{S*B(V)}}{9} + \frac{2\sigma^2_e}{18} = \frac{2}{18}(6\sigma^2_{B(V)} + 2\sigma^2_{S\times B(V)} + \sigma^2_e) = \frac{1}{9}[EMS_{B(V)}] \quad \Rightarrow$

  $\widehat{SE}\left(\hat{\hat{\mu}}_{1..} - \hat{\hat{\mu}}_{2..}\right) = \sqrt{\frac{1.6702}{9}} = .4307$

Problem III. (9 points) The study design was a fractional factorial with 16 runs using the generators

$I_1 = ABDE = +$ and $I_2 = ACE = +$ to generate the 16 treatments to be used in the study.

1. For each of the following treatments, check YES if the treatment will appear in the experiment, otherwise check NO.

   i. $(A, B, C, D, E, F) = (+, -, +, +, -, +) \Rightarrow ABDE = +$ and $ACE = - \Rightarrow$ No
   ii. $(A, B, C, D, E, F) = (+, -, +, -, +, +) \Rightarrow ABDE = +$ and $ACE = + \Rightarrow$ Yes

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

- (ABDE)(ACE)=BCD therefore the length of the shortest generator and implicit generator is 3. Thus, Resolution = III

3. What effects which must be assumed to be negligible in order that the data from the experiment will provide an estimate of the interaction between Factors C and F.

- I = ABDE = ACE = BCD. Therefore, the following effects are confounded with CF:

  (CF)(ABDE) = ABCDEF; (CF)(ACE)= AEF (CF)(BCD) = BDF.

  Thus, in order to estimate CF we would need to be assured that the three effects ABCDEF, AEF, and BDF are negligible.

Problem IV. (36 pts)

(1). C. normal distribution of the residuals does not appear to be violated

(2). B. Tukey's HSD applied to the levels of F_1 averaged over all combinations of (F_2, F_3)

- Because F_1 does not interact with F_3 and F_2 is random

(3). E. could be made using Tukey's HSD on the adjusted treatment means at specified values of the covariate.

- The size of the difference in the Trt means depends on the value of the covariate

(4). C. the calculation of $\widehat{SE}\,(\hat{\mu}_i)$ is incorrect, SAS only considers $\hat{\sigma}^2_e$ and not $\hat{\sigma}^2_{PLOT(DOSE)}$ in the calculation.

(5). B. Run the Friedman test even though the residuals are very right skewed.

- Because the ordering of the Trt means will be reversed in the transformed scale

(6). E. none of the above

- There are 120 populations, one for each combination of the 12 treatments and 10 blocks

(7). A. $b = 10$ and $k = 9$

- All four designs have the same number of reps per treatments but Design A has the most EU's per block

(8). D. the 5 measurements on each EU are not independent.

(9). C. $L = -3\mu_{11} - \mu_{12} + \mu_{13} + 3\mu_{14} + 3\mu_{21} + \mu_{22} - \mu_{23} - 3\mu_{24} = (-3\mu_{11} - 1\mu_{12} + \mu_{13} + 3\mu_{14}) - (-3\mu_{21} - \mu_{22} + \mu_{23} + 3\mu_{24})$

- From Table XI, the coefficients are (-3,-1,1,3), and we want the difference in the contrasts for the 2 levels of F_1:

(10). B. b_i's are independent, $d_{j\ell}$'s are independent, and $e_{ijk\ell}$'s are correlated

- Because the block effects are independent, the measurements from 2 different EU's are independent, but the measurements from the same EU are correlated

(11). A. the probability of a Type I will be higher than expected under no correlation

- Because positive correlation causes the estimated standard error of the LS means to be underestimated and hence the probability of rejecting H_o is inflated which results in larger values for the power but also an increase in probability of Type I errors.

(12). D. 0.70$<$ Power ≤ 0.90

- $t = (4)(2) = 8,\ r = 5\ \Rightarrow\ \nu_1 = t - 1 = 7,\ \nu_2 = t(r-1) = 32$ and $D = 6,\ \sigma_e = 1.86,\ \Rightarrow$
  $\lambda = \frac{rD^2}{2\sigma^2_e} = 26.014,\ \Rightarrow \phi = \sqrt{\frac{\lambda}{t}} = 1.8.$
  Using Table IX on page 610 with $\alpha = .01$, the vertical line through $\phi = 1.8$ intersects the $\nu_2 = 30$ line at .78 so the Power at $\nu_2 = 32$ should be approximately .80.

  Using R: Power = 1-pf(qf(.99, 7, 32),7,32,26.014) = .796

```
Summary of Scores on FINAL EXAM - STAT 642 Spring 2020

N  = 89,   Min = 38,    Q1 = 75,   Median = 84,    Mean = 78.0,    Q3 = 89,    Max = 100
```