Partial Solutions to STATISTICS 642 - EXAM II April 1, 2020

Problem I (64 points)

1. Based on the two plots and assuming the standard errors for estimating the treatment means is very small, there appears to be a three-way interaction. The Water by Day interaction for Chamber C1 has a different pattern than the Water by Day interaction for Chamber C2.

   However, examining the AOV table, the p-value for the 3-way interaction was .0733 from which it would be concluded that there is not significant evidence of a 3-way interaction. When the standard errors, $\hat{SE}(\hat{\mu}_{ijk}) = \sqrt{2.836667/4} = 0.84$ or $\hat{SE}(\hat{\mu}_{ijk}) = \sqrt{2.836667/2} = 1.19$, in the estimated means are taken into account, the evidence of a three-way interaction is not justified.

2. Do the necessary conditions for testing hypotheses and constructing confidence intervals appear to be satisfied? Justify your answers.

   Using the information on pages 18 and 22 of the SAS output:

   C_1 Normality: Shapiro-Wilks test has p-value= .3122 and the 66 plotted points in the normal probability plot are very close to straight line which confirms an excellent fit of the normal distribution to the residuals

   C_2 Equal Variance: The Brown-Forsythe's test of homogeneity is invalid because three of the treatments had only 2 observations and B-F test requires at least 3 data values. However, the plot of the 66 residuals vs the treatment means on page 18 indicates a rather equal spread in the residuals as a function of the treatment means with the exception of one of the 66 residuals. Thus, the condition of equal variances does not appear to be violated.

3. There is not significant evidence of a 3-way interaction C*W*D (p-value =.0733). However, both C*D (p-value $<$ .0001) and W*D (p-value =.0001) are significant. This implies that when making inferences about differences in the mean germination rate for the three Harvest Days, it is necessary to make the inferences about Days separately at each level of Water averaged over the two Chambers; and inferences about Days separately for each of the two Chambers averaged over the levels of Water. Because only 6 of the 15 C-W pairs of means and only 9 of the 36 D-W pairs of means are of interest, the Tukey-Kramer p-values would be too conservative for the comparisons. Thus, the unadjusted p-values for C*D and W*D means will be used with $\alpha_{pc} = .05/(6+9) = 0.0033$:

- For C = C1: G1={D=45}; G2={D=60}; G3={D=85};
- For C = C2: G1={D=45}; G2={D=60}; G3={D=85};
- For $W = 25 : G1 = \{D = 45\};\ G2 = \{D = 60\};\ G3 = \{D = 85\}$
- For $W = 50 : G1 = \{D = 45,\ D = 60\};\ G2 = \{D = 85\}$
- For $W = 75 : G1 = \{D = 45,\ D = 60\};\ G2 = \{D = 85\}$

4. $\hat{\mu}_{25,45} = 60$ with $\widehat{SE}(\hat{\mu}_{25,45}) = .7293$ with df= 66-18=48, therefore, a 95% C.I. for the mean germination rate when W=25 and D=45 is

$$\hat{\mu}_{25,45} \pm t_{.025,48}\hat{SE}(\mu_{25,45}) = 60 \pm (2.011)(.7293) = 60 \pm 1.47 = (58.53\%, 61.47\%)$$

$\widehat{SE}(\mu_{25,45}) = \sqrt{\frac{MSE}{(2)^2}\left(\frac{1}{4}+\frac{1}{2}\right)} = .7293$ or use value from SAS output page 5

Problem II (36 points) For each of the following questions, Circle the Best answer.

(1) **C** - The factor Manufacturer has Random levels and hence HSD would not be appropriate

(2) **A** - $F_1 \times F_2$ is not significant thus contrast the means of F_1 averaged over F_2

(3) **B** - Compare the levels of factor A separately for each level of Factor B using the ranks

(4) **D** - The only contrast having all $n_{ij} > 0$

(5) **C** - PROC GLM ignores all random effects except $eijk$ in calculation standard errors, use PROC MIXED

(6) **B or E** - Factor B interacts with Factor A and the three way interaction is not significant. But whether or not Factor B interacts with Factor C is not given. Therefore, to be conservative, assume that the B by C interaction is significant and select Answer B or select answer E implying that not enough information is given.

(7) **B** - With subsamples, the correct error term is $MS_{Rep(Trt)}$ with each of the treatment means averaged over (5)(7)=35 data values

(8) **B** - This is the definition of Non-estimable

(9) **D** - $t = 9,\ r = 8,\ \alpha = .01,\ \nu_1 = 8,\ \nu_2 = 63,\ \lambda = (8)(25)^2/[2*(10)^2] = 25,\ \phi = \sqrt{25/9} = 1.67,$ Power $\approx .82$

Using R, $\gamma(25) = 1 - pf(qf(.99, 8, 63), 8, 63, 25) = .8203$

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

N  = 95,   Min = 40,    Q1 = 68,   Median = 78,    Mean = 72.7,    Q3 = 84.5,    Max = 96
```