STATISTICS 642 - ASSIGNMENT 5

DUE DATE: 8am Central, THURSDAY, March 10, 2022

Name (**Typed**) ____

Email Address (**Typed**) ____

Make this cover sheet the first page of your Solutions.

STATISTICS 642 - ASSIGNMENT #5

- Read Handouts 5
- Supplemental Reading: Chapter 5 in the Design & ANOVA book
- Hand in the following problems by 8am Central, THURSDAY, March 10, 2022

1. **(35 points)** An accelerated (temperature stressed) life test was performed on HD televisions to assist the manufacturer in determining an appropriate warranty. Six TV's were tested at each of four temperatures: $40°C, 45°C, 55°C, 70°C$. The time to failure in hours are given in the following table for the 24 television sets.

Temperature During Test	Hours to Failure					
$40°C$	1953	2135	2471	4727	6134	6314
$45°C$	1190	1286	1550	2125	2557	2845
$55°C$	651	817	848	1038	1361	1543
$70°C$	511	651	651	652	688	729

   a. Do the conditions of normality, equal variance, and independence of the data values hold for the data from the four temperatures? Justify your answers with plots, tests, and logical reasoning.

   b. Determine a reasonable transformation of the data using the slope of the regression line based on $\log(S_i)$ vs $\log(\bar{y}_{i.})$.

   c. Use the Box-Cox Technique for selecting a transformation of the data. Is the transformation from Box-Cox procedure consistent with your transformation from part b.?

   d. Using the transformation from part c., is the transformed data appropriate for conducting an AOV?

   e. Perform an AOV on both the original data and the transformed data. Compare the results from the two analyses.

   f. Use Tukey's HSD to group the four temperatures relative to the mean time to failure.

   g. Test for a trend in the time to failures as a function of Temperature? Because the Temperatures were unequally spaced, the following contrast coefficients were obtained from R. The coefficients for the three contrasts, Linear, Quadratic, and Cubic are given below:

```
con = contr.poly(4,scores=c(40,45,55,70))

       Linear    Quadratic    Cubic
40   -0.5455447  0.5128226 -0.43519414
45   -0.3273268 -0.1709409  0.78334945
55    0.1091089 -0.7407437 -0.43519414
70    0.7637626  0.3988620  0.08703883

con_std = matrix(0,4,3)
m = c(rep(0,3))
for (i in 1:3) {
m[i] =min(abs(con[,i]))
con_std[,i] = con[,i]/m[i]
}

con_std
     [,1]       [,2] [,3]
[1,]   -5  3.000000   -5
[2,]   -3 -1.000000    9
[3,]    1 -4.333333   -5
[4,]    7  2.333333    1

con_std[,2] = 3*con_std[,2]
con_std
      Lin  Quad  Cubic
[1,]   -5    9   -5
[2,]   -3   -3    9
[3,]    1  -13   -5
[4,]    7    7    1
```

2. **(15 points)** For the data time to failure data in Problem 1.,

   a. Use a rank based test to compare the average time to failure for the four temperatures.

   b. Use a rank based multiple comparison procedure to group the four temperatures relative to the average time to failure.

   c. Compare your results to your analysis on the untransformed data.

3. **(30 points)** An entomologist counted the number of eggs laid by female moths on successive days in three strains of tobacco budworm (USDA, Field, Resistant) from each of 15 matings. The entomologist was interested in evaluating whether the average number of eggs was different for the three strains. The number of eggs laid on the third day after the mating for each female is given in the following table:

Strain	Number of Eggs per Moth														
USDA	448	906	28	277	634	48	369	137	29	522	319	242	261	566	734
FIELD	211	276	415	787	18	118	1	151	0	253	61	0	275	0	153
RESIST	0	9	143	1	26	127	161	294	0	348	0	14	21	0	218

   a. The entomologist suspects that the data is from Poisson distributions. Based on the data do Poisson distributions appear to be reasonable distributions for the egg data?

   b. Using PROC GENMOD in SAS, perform an analysis using a model having a Poisson distributions for the three egg count distributions. Make sure to check for variance inflation.

   c. The 15 female moths of each strain were placed in individual cages on a laboratory bench in a north to south arrangement. The three strains were reasonably separated thus eliminating any potential influence on egg production between moths of different strains. The data displayed in the above table is in the same arrangement. Does there appear to be a spatial correlation in the egg counts?

4. **(20 points, 4 each)** Answer each of the following questions using at most **<u>20 WORDS</u>**.

   a. In a CRD, a residual analysis indicated a strong positive correlation in the residuals. Your assistant tells the project director that this is great because the actual power of the AOV F-test will now be greater than the power of the AOV F-test when the residuals are independent. Is your assistant's statement true? If yes, then explain why we do not design experiments to have positively correlated residuals. If your assistant is not correct, then what is the actual effect of positive correlation on the power of the AOV F-test.

   b. In a CRD with a single factor having 7 levels, the experimenter plotted the seven treatment means and standard deviations $(\bar{y}_{i.}, S_i)$ on a graph. The following model was fitted to the seven plotted points $ln(S_i) = \beta_o + \beta_1 ln(\bar{y}_{i.})$. The fitted model yielded $\hat{\beta}_o = 2.8, \quad \hat{\beta}_1 = 1.5, \quad R^2 = 0.92$, and a p-value=.0023 for testing $H_o : \beta_1 = 0$. Which of the three conditions in AOV appear to be violated? What transformation would you suggest for the data to moderate the violation of this condition?

   c. Refer to the previous question. The researcher was interested in determining a group of treatments having the largest mean response. She applied Hsu's procedure to the transformed data using the **best is largest** definition. The value of MSE was very small so a one term Taylor series approximation is very accurate and the resulting transformed data appears to satisfy the three AOV conditions. What is a serious complication in directly applying Hsu's procedure to the transformed data in order to determine the group of treatments having greatest mean? State how the problem could be easily overcome.

   d. What are the necessary conditions on the distributions of the data in a CRD in order for the Kruskal-Wallis statistic to be valid in testing for treatment differences? How are these conditions different from the conditions required for applying the $F = MS_{TRT}/MSE$ test statistic?

   e. The ANOVA from a completely randomized design with r=10 reps of 5 treatments yielded MS_{Error} = 9. There were N = 50 observations in the experiment with the residuals ranging from -6.5 to 8.7. Would you consider any of the residuals to be outliers? Justify your answer.