STA 6208 – Homework 3

Part 1: Using weighted Least Squares, test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ for the Miniscus data. The weights should be $w_{ij} = 1/S_i^2$ (the sample variance for treatment $i$). Use the 3rd response variable (stiffness).

Note that you will be conducting a Complete and Reduced model of weighted errors:

$$SSE_w(R) = \sum_{i=1}^{t} \sum_{j=1}^{r} w_{ij} \left( Y_{ij} - \mu \right)^2$$

$$SSE_w(C) = \sum_{i=1}^{t} \sum_{j=1}^{r} w_{ij} \left( Y_{ij} - \mu_i \right)^2$$

- Derive the WLS means under the Reduced and Complete models.
- Obtain the F-test using the aov function with the weight option
- Reproduce the F-test based on the matrix form of the model.

Part 2: Analyze the sugar cane pest data (response = weight of juice) for a 1-Way ANOVA.

- Use Bartlett’s and Levene’s tests for equality of variance.
- Estimate the power relationship between the standard deviation and the mean, and suggest a transformation based on it. Re-fit the ANOVA based on the transformed data.
- Use Welch’s test and the Kruskal-Wallis test, and compare results with the F-test.

Dataset: pests_cane.dat


Description: Results of experiments on the effects of four types of sugarcane pests on weight of 25 canes (lbs) and the weight of juice for the 25 canes (lbs). Experimental units=25 grouped canes, 6 replicates per treatment.

Treatments:
1=Healthy (Control)
2=Top-shoot Borer
3=Stem Borer
4=Top-shoot & Stem Borer
5=Root Borer
6=Termites
7=Stem&Root Borer
8=Root Borer & Termites
9=Top-Shoot & Root Borer
10=Top-Shoot & Stem & Root Borer
11=Top-Shoot Borer & Termites

Variables/Columns
Treatment 7-8
Weight of Canes (lbs) 10-16
Weight of Juice (lbs) 18-24
Part 3: Analyze the Laundry Shrinkage Data with 5 factors (all fixed). Include all Main Effects, 2-Way, and 3-Way Interactions. Write out the model, and give the Analysis of Variance and test for all relevant effects. Analyze the response length shrinkage percent.

Obtain residuals for your “best” model and plot them against their fitted values. Any sign of non-constant error variance? Are they approximately normally distributed?

Dataset: laundry1.dat


Description: Length and Width Shrinkage and Skewness Measurements for a 5 factor experiment varying structure, presence of wrinkle resistant finish, cycle number, rinse cycle softener, and drying method.

Variables/columns
- Structure 8 /* 1=Poplin, 2=Sheeting */
- Wrinkle resistant finish 16 /* 0=No, 1=Yes */
- Cycle Number 24
- Rinse Cycle Softener 32 /* 0=No, 1=Yes */
- Drying Method 40 /* 1=Tumble Dry, 0=Line Dry */
- Specimen Number 47-48 /* Presumably */
- Length Shrinkage Percent 54-56
- Width Shrinkage Percent 62-64
- Skewness Percent 70-72

Part 4

A study was conducted to observe the effects of germination time (days) and germination temperature of sorghum malt. The response is α-Amylase development.

- Fit a 2-Way ANOVA with additive effects
- Conduct Tukey’s One degree of freedom test for non-additivity

<table>
<thead>
<tr>
<th>Germination time (days)</th>
<th>Sample ID</th>
<th>28°C germinated sorghum malt</th>
<th>30°C germinated sorghum malt</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Laboratory malt</td>
<td>42</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>Laboratory malt</td>
<td>55</td>
<td>59</td>
</tr>
<tr>
<td>5</td>
<td>Laboratory malt</td>
<td>57</td>
<td>63</td>
</tr>
<tr>
<td>6</td>
<td>Laboratory malt</td>
<td>37</td>
<td>49</td>
</tr>
</tbody>
</table>

Part 5
QA.1. A hotel is interested in studying the effects of washing machines and detergents on whiteness of bed sheets. The hotel has 4 washing machines and 3 brands of detergent. They randomly assign n=4 sheets for each combination of machine and detergent (each sheet is only observed for one combination of machine and detergent). After washing, the sheets are measured for whiteness (high scores are better). The model fit is:

\[ y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \quad e_{ijk} \sim NID(0, \sigma^2) \quad \sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0 \]

<table>
<thead>
<tr>
<th>Washer \ Detergent</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>35</td>
<td>65</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>45</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>50</td>
<td>70</td>
<td>50</td>
</tr>
<tr>
<td>Mean</td>
<td>25</td>
<td>40</td>
<td>55</td>
<td>40</td>
</tr>
</tbody>
</table>

pA.1.a. Give least squares estimates of the following parameters:
\[ \alpha_1: \quad \beta_1: \]
\[ (\alpha\beta)_{11}: \quad \sigma^2: \]

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>F(0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detergent</td>
<td></td>
<td>36000</td>
<td>#N/A</td>
<td>#N/A</td>
<td></td>
</tr>
<tr>
<td>M*D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>36000</td>
<td>#N/A</td>
<td>#N/A</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>57000</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
</tr>
</tbody>
</table>

pA.1.b. Complete the ANOVA table.

pA.1.c. Is there a significant interaction between Machine and Detergent on whiteness scores?  Yes / No

pA.1.d. Is there a significant main effect for Machines?  Yes / No

pA.1.e. Is there a significant main effect for Detergents?  Yes / No

QA.2. For the 2-way ANOVA with fixed effects and interaction, considering the following two models:

1) \[ y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk} \quad i = 1,...,a; \quad j = 1,...,b; \quad k = 1,...,r \quad \sum_{i} \alpha_i = \sum_{j} \beta_j = \sum_{i} \alpha\beta_{ij} = \sum_{j} \alpha\beta_{ij} = 0 \]

2) \[ y_{ijk} = \mu + e_{ijk} \]

pA.2.a. Write out the Sums of Squares and Degrees of Freedom for each model:

Model 1:  \[ SSA = \quad df_A = \]
\[ SSB = \quad df_B = \]
\[ SSAB = \quad df_{AB} = \]

Model 2:  \[ SSt sts = \quad df_{st ts} = \]
pA.2.b. (Algebraically) Show that SSA + SSB + SSAB = SSTrts and that df_A + df_B + df_AB = df_Tri.

Hint: The expansion for SSAB is very much like the expansion for SSA.

QA.6. An experiment was conducted to compare \( a = 3 \) theories for the apparent modulus of elasticity (\( Y \)) of \( b = 3 \) apple varieties. The 3 theories were: Hooke’s, Hertz’s, and Boussineq’s; the 3 apple varieties were: Golden Delicious, Red Delicious, and Granny Smith. The researchers determined the elasticity for \( n = 15 \) based on each combination of theory and variety. For the purposes of this experiment, each factor is fixed.

Model: \( Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \sim NID(0, \sigma^2) \) \( \sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0 \)

\[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{i \cdot \cdot})^2 = 17.095 \]

\[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{\cdot \cdot \cdot})^2 = 113.119 \]

Complete the following Analysis of Variance Table, and test for interaction effects and main effects.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>F(.95)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theory*Variety</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

QA.3 Consider a 2-factor, fixed effects, interaction model with \( a = b = n = 2 \).

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \sim NID(0, \sigma^2) \] \( \sum_{i=1}^{2} \alpha_i = \sum_{j=1}^{2} \beta_j = \sum_{i=1}^{2} (\alpha\beta)_{ij} = \sum_{j=1}^{2} (\alpha\beta)_{ij} = 0 \)

pA.3.a. Given the following data, obtain MS_{Err}.
### QA.3

#### b. Use the matrix form of the model to obtain: the least squares estimate of the parameter vector $\hat{\beta}^*$, its estimated variance-covariance matrix, standard errors, and t-tests for all of the parameters.

$$
\beta^* = 
\begin{bmatrix}
\mu \\
\alpha_i \\
\beta_j \\
(\alpha\beta)_{ij}
\end{bmatrix}
$$

$$
X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\
 x_{21} & x_{22} & \cdots & x_{2n} \\
 \vdots & \vdots & \ddots & \vdots \\
 x_{n1} & x_{n2} & \cdots & x_{nn} \\
\end{bmatrix}
$$

$$
X'X = \begin{bmatrix} 
\sum x_{11}^2 & \sum x_{12}x_{11} & \cdots & \sum x_{1n}x_{11} \\
\sum x_{21}x_{11} & \sum x_{22}^2 & \cdots & \sum x_{2n}x_{21} \\
\vdots & \vdots & \ddots & \vdots \\
\sum x_{n1}x_{11} & \sum x_{n2}x_{21} & \cdots & \sum x_{nn}^2 \\
\end{bmatrix}
$$

$$
(XX)^{-1} =
$$

$$
X'Y = \begin{bmatrix} 
\sum y_{11}x_{11} \\
\sum y_{12}x_{12} \\
\vdots \\
\sum y_{nn}x_{nn} \\
\end{bmatrix}
$$

$$
V \hat{\beta} = 
\begin{bmatrix} 
\sigma^2 & 0 & \cdots & 0 \\
0 & \sigma^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma^2 \\
\end{bmatrix}
$$

### QA.4

A 2-Way fixed effects model is fit, with factor A at 4 levels, and factor B at 3 levels. There are 3 replicates for each combination of factors A and B. The error sum of squares is $SS_{Err} = 720$.

#### a. Suppose that the interaction is not significant. What will be Tukey’s and Bonferroni’s minimum significant differences be for comparing the means for factor A (averaged across levels of factor B).

Tukey’s HSD: _____________________________  Bonferroni’s MSD: _____________________________

#### b. Suppose that the interaction is significant. What will be Tukey’s and Bonferroni’s minimum significant differences be for comparing the means for factor A (within a particular level of factor B).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std Error</th>
<th>t</th>
<th>t(.975)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\alpha\beta)_{11}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
pA.4.c. How large would the interaction sum of squares, SS_{AB} need to be to reject $H_0$: No interaction between factors A&B?

QA.5. A study was conducted to test for effects on willingness to pay during online auctions. There were 2 factors, each with 2 levels (both fixed): Urgency (Present (i=1) /Absent (i=2)) and Contrast (3 of 6 items “Featured” (High, j=1)/all 6 items “Featured” (Low, j=2)). There were 6 watches, and the response was the amount the participant was willing to pay for the watch. Although the researchers started with 80 subjects, 9 were eliminated due to incomplete information, so $N = 71$. The model fit is:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + \epsilon_{ijk} \quad i = 1, 2; j = 1, 2; k = 1, \ldots, n_{ij} \quad \alpha_1 + \alpha_2 = \beta_1 + \beta_2 = \sum_{i=1}^{2}(\alpha \beta)_{ij} = \sum_{j=1}^{2}(\alpha \beta)_{ij} = 0 \quad i, j = 1, 2$$

pA.5.a. Give the form of the (full rank) $X$ matrix and $\beta$ vector for this model. Note that although $X$ has 71 rows, there are only 4 “blocks” of distinct levels, each with a particular numbers of subjects.

pA.5.b. Obtain $X^T X$ as functions of the cell sample sizes (Just give the values on or above the main diagonal).

pA.5.c. The authors fit the following 4 models, with approximate Error sums of squares (divided by 1000 for ease of calculation):

Model 1: $E\{Y_{ik}\} = \mu + \alpha_i + \beta_j$  
Model 2: $E\{Y_{ik}\} = \mu + \alpha_i + \beta_j$  
Model 3: $E\{Y_{ik}\} = \mu + \beta_j + (\alpha \beta)_{ij}$  
Model 4: $E\{Y_{ik}\} = \mu + \alpha_i + (\alpha \beta)_{ij}$

SSE_{r1} = 884.0  
SSE_{r2} = 937.0  
SSE_{r3} = 941.5  
SSE_{r4} = 978.3

Test: $H_0^{AB}: (\alpha \beta)_{11} = (\alpha \beta)_{12} = (\alpha \beta)_{21} = (\alpha \beta)_{22} = 0$

Test Statistic: ______________ Rejection Region: ______________ Do you conclude the Urgency effect “depends” on Contrast?  Y / N

pA.5.d. The sample means for the 4 treatments are: $\bar{Y}_{11i} = 216.94 \quad \bar{Y}_{12i} = 90.26 \quad \bar{Y}_{21i} = 106.12 \quad \bar{Y}_{22i} = 88.06$. Treating the cell sample sizes as $n_{ij} = 18$ as an approximation, use Bonferroni’s method to compare all pairs of treatment means.