STA 6208 – Homework 3

Part 1: Using weighted Least Squares, test H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_\bullet$ 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:

$$SSE_w(R) = \sum_{i=1}^{t}\sum_{j=1}^{r} w_{ij}\left(Y_{ij} - \hat{\mu}_\bullet^{W}\right)^2 \qquad SSE_w(C) = \sum_{i=1}^{t}\sum_{j=1}^{r} w_{ij}\left(Y_{ij} - \hat{\mu}_i^{W}\right)^2$$

- 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

Source: P.C. Mahalanobis and S.S. Bose (1934). "A Statistical Note on
the Effect of Pests on the Yield of Sugarcane and the Quality of Cane-Juice,"
Sankhya: The Indian Journal of Statistics, Vol.1, #4, pp.399-406.

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
Treatmet 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

```
Source: L. Higgins, S.C. Anand, D.A. Holmes, M.E. Hall, and K. Underly (2003).
"Effects of Various Home Laundering on the Dimensional Stability, Wrinkling,
and Other Properties of Plain Woven Cotton Fabrics: Part II: Effect of Rinse
Cycle Softener and Drying Method and of Tumble Sheet Softener and Tumble
Drying Time," Textile Research Journal, Vol. 73, pp. 407-420.

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 I. α-Amylase (U/g) development of sorghum malt germinated at 28°C and 30°C.

Germination time (days)	Sample ID	28°C germinated sorghum malt	30°C germinated sorghum malt
3	Laboratory malt	42	57
4	Laboratory malt	55	59
5	Laboratory malt	57	63
6	Laboratory malt	37	49

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:

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

Washer\Detergent	1	2	3	Mean
1	25	30	35	30
2	20	35	65	40
3	25	45	50	40
4	30	50	70	50
Mean	25	40	55	40

pA.1.a. Give least squares estimates of the following parameters:

α_1: β_1:

$(\alpha\beta)_{11}$: σ^2:

Source	df	SS	MS	F	F(0.05)
Machine					
Detergent					
M*D					
Error		36000		#N/A	#N/A
Total		57000	#N/A	#N/A	#N/A

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) $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \quad i = 1,\ldots,a;\ j = 1,\ldots,b; k = 1,\ldots,r \qquad \sum_i \alpha_i = \sum_j \beta_j = \sum_i \alpha\beta_{ij} = \sum_j \alpha\beta_{ij} = 0$

2) $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$

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

Model 1: SSA = df_A =

SSB = df_B

SSAB = df_{AB} =

Model 2: SSTrts = df_{Trts} =

pA.2.b. (Algebraically) Show that SSA + SSB + SSAB = SSTrts and that $df_A + df_B + df_{AB} = df_{Trts}$

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.

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

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \overline{Y}_{ij\bullet}\right)^2 = 17.095 \qquad \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \overline{Y}_{\bullet\bullet\bullet}\right)^2 = 113.119$$

Cell Means	GoldenDelicious	RedDelicious	GrannySmith	Row Mean
Hooke	2.68	3.46	4.23	3.457
Hertz	2.44	3.06	3.84	3.113
Boussinesq	1.53	1.89	2.36	1.927
Column Mean	2.217	2.803	3.477	2.832

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

Source	df	SS	MS	F	F(.95)	P-value
Theory						>0.05 or <0.05
Variety						>0.05 or <0.05
Theory*Variety						>0.05 or <0.05
Error				#N/A	#N/A	#N/A
Total			#N/A	#N/A	#N/A	#N/A

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

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

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

Y111	Y112	Y121	Y122	Y211	Y212	Y221	Y222
22	18	31	29	8	12	37	43

SS_{Err} = ______________________________ MS_{Err} = ________________________

pA.3.b. Use the matrix form of the model to obtain: the least squares estimate of the parameter vector $\boldsymbol{\beta}^*$, its estimated variance-covariance matrix, standard errors, and t-tests for all of the parameters.

$$\boldsymbol{\beta}^* = \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ (\alpha\beta)_{11} \end{bmatrix} \qquad \mathbf{X} = \qquad \mathbf{X'X} =$$

$$(\mathbf{X'X})^{-1} = \qquad \mathbf{X'Y} =$$

$$\hat{\boldsymbol{\beta}}^* = \qquad \hat{V}\left\{\hat{\boldsymbol{\beta}}^*\right\} =$$

Parameter	Estimate	Std Error	t	t(.975)
μ				
$\alpha 1$				
$\beta 1$				
$(\alpha\beta)11$				

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 combinations of factors A and B. The error sum of squares is SS_{Err} = 720.

pA.4.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: _______________________________

pA.4.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).

Tukey's HSD: ___________________________________ Bonferroni's MSD: _______________________________

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:

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

pA.5.a. Give the form of the (full rank) **X** matrix and β 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'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: $E\{Y_{ijk}\} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ Model 2: $E\{Y_{ijk}\} = \mu + \alpha_i + \beta_j$ Model 3: $E\{Y_{ijk}\} = \mu + \beta_j + (\alpha\beta)_{ij}$ Model 4: $E\{Y_{ijk}\} = \mu + \alpha_i + (\alpha\beta)_{ij}$

$SSErr_1 = 884.0$ $SSErr_2 = 937.0$ $SSErr_3 = 941.5$ $SSErr_4 = 978.3$

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: $\overline{Y}_{11\bullet} = 216.94$ $\overline{Y}_{12\bullet} = 90.26$ $\overline{Y}_{21\bullet} = 106.12$ $\overline{Y}_{22\bullet} = 88.06$. Treating the cell sample sizes as $n_{ij} = 18$ as an approximation, use Bonferroni's method to compare all pairs of treatment means