STATISTICS 642 - ASSIGNMENT 6 - Summer 2015

DUE DATE: NOON, FRIDAY, **July 17, 2015**

Name (**Typed**) ____

Email Address (**Typed**) ____

HOMEWORK #6: DUE Noon, FRIDAY, **July 17, 2015**

Read Handouts 8 and 9 and Chapter 7 in Textbook

Hand in only Problem 1, Problem 2, and Problem 3. Work the other problems but do not submit them for grading. I will provide solutions for all the problems.

Problem 1. (35 points) A traffic engineering study was designed to evaluate the effects of three types of traffic signals: *pre-timed* signals, *semi-actuated* signals, and *fully actuated* signals; on traffic delay at intersections. Also, two methods of measuring traffic delays: *point-sample* and *path-trace* were used to estimate stopped time per vehicle at an intersection. Two intersections were randomly assigned to each of the three types of traffic signals. Data was collected at each of the six intersections during a rush hour period and a nonrush hour period and using both methods of measuring traffic delays. The measured traffic delays (in seconds) per vehicle are given here:

		Point-Sample		Path-Trace	
Signal	Intersection	Rush Traffic	NonRush Traffic	Rush Traffic	NonRush Traffic
Pretimed	1	61.7	57.4	53.1	36.5
	2	35.8	18.5	35.5	15.9
Semi-actuated	3	20.0	24.6	17.0	21.0
	4	2.7	3.1	1.5	1.1
Fully Actuated	5	35.7	26.8	35.4	20.7
	6	24.3	25.9	27.5	23.3

Use the following SAS code to assist you in answering the following questions:

```
ODS HTML;
ODS GRAPHICS ON;
OPTIONS LS=90 PS=55 nocenter nodate;
DATA ;
INPUT S $ I $ M $ T $ Y @@;
LABEL S='SIGNAL' I='INTERSECTION' M='MEASUREDEVICE' T='TRAFFIC';
CARDS;

P 1 PS R  61.7    P 1 PS NR 57.4    P 1 PT R  53.1    P 1 PT NR 36.5
P 2 PS R  35.8    P 2 PS NR 18.5    P 2 PT R  35.5    P 2 PT NR 15.9
S 3 PS R  20.0    S 3 PS NR 24.6    S 3 PT R  17.0    S 3 PT NR 21.0
S 4 PS R   2.7    S 4 PS NR  3.1    S 4 PT R   1.5    S 4 PT NR  1.1
F 5 PS R  35.7    F 5 PS NR 26.8    F 5 PT R  35.4    F 5 PT NR 20.7
F 6 PS R  24.3    F 6 PS NR 25.9    F 6 PT R  27.5    F 6 PT NR 23.3

PROC GLM;
CLASS S M T I ;
MODEL Y = S T I(S) S*T T*I(S) M S*M T*M M*I(S) S*T*M;
RANDOM I(S) T*I(S) M*I(S)/TEST;
RUN;

PROC MIXED CL ALPHA=.05 COVTEST ;
CLASS S M T I ;
MODEL Y = S T  S*T  M S*M T*M  S*T*M;
RANDOM I(S) T*I(S) M*I(S);
LSMEANS S|M|T/ADJUST=TUKEY;
RUN;
ods graphics off;
ods html close;
```

a. Write a model for this study.

b. Provide an AOV table for this study.

c. Provide the Expected Mean Squares for all sources of Variation in the AOV table.

d. What can you conclude at the $\alpha = .05$ levels about the effect of Type of Traffic Signal, Measuring Method, and Level of Traffic on the average traffic delay?

e. Provide estimates of all the variance components and their proportions of the total variance in traffic delay measurements.

Problem 2. (41 points) An experiment was conducted with three Factors: A at 4 random levels, B at 5 fixed levels, and C at 3 random levels nested within factor B. There were 6 experimental units randomly assigned to each of the treatments. The following model was fit to the 360 responses obtained in the experiment:

$$y_{ijkl} = \mu + a_i + \beta_j + c_{k(j)} + (a\beta)_{ij} + (ac)_{ik(j)} + e_{ijkl},$$

with $i = 1,2,3,4;\quad j = 1,2,3,4,5;\quad k = 1,2,3;\quad l = 1,2,3,4,5,6;$

where μ and β_j are population parameters with $\beta_5 = 0$; and a_i, $c_{k(j)}$, $(a\beta)_{ij}$, $(ac)_{ik(j)}$ and e_{ijkl} are independent rv's with $N(0,\sigma^2_A)$, $N(0,\sigma^2_{C(B)})$, $N(0,\sigma^2_{AB})$, $N(0,\sigma^2_{AC(B)})$, and $N(0,\sigma^2_e)$ distributions, respectively.

Source	DF	**MS**	Expected Mean Squares .
A		24.5	
B		19.7	
$A \times B$		8.9	
$C(B)$		7.5	
$A \times C(B)$		6.8	
Error		5.8	

a. Complete the above AOV table for the experiment by filling in the degrees of freedom and Expected Mean Squares.

b. Test for a significant AB interaction ($\alpha = 0.05$). Note that the AOV table is providing the MS, not SS for each source of variation.

c. Test for a significant B main effect ($\alpha = 0.05$).

d. Compute the variance of the difference in treatment means for levels 1 and 2 of Factor B: $\bar{y}_{.1..} - \bar{y}_{.2..}$. Provide an estimate of this variance and the degrees of freedom of the estimate.

e. Compute the value of Tukey's HSD with $\alpha = .05$ that would be used to determine which pairs of means across the levels of Factor B are different.

Problem 3. (24 points) An experiment was run to evaluate the effect of two Factors on the mean of a response variable. There are three levels of factor F_1 and four levels of factor F_2. Initially, three reps of each of the 12 treatments were to be run. However, it was found that certain levels of F_1 were incompatible with certain levels of factor F_2. Thus, the experiment was unbalanced with only 8 of 12 treatments being observed in the experiment. The **mean** responses for the 8 treatments are given here:

MEANS

	F_2			
F_1	1	2	3	4
1	$\bar{y}_{11}$	*	*	$\bar{y}_{14}$
2	$\bar{y}_{21}$	$\bar{y}_{22}$	$\bar{y}_{23}$	*
3	$\bar{y}_{31}$	*	$\bar{y}_{33}$	$\bar{y}_{34}$

Suppose each mean in the above table is based on the results of 3 reps. For each of the following contrasts, determine whether the contrast is testing a **Main Effect** or **Interaction Effect** or **Neither**. Then determine whether or not the contrast is **Estimable** based on the observed data.

a. $\mathbf{C}_1 = \mu_{11} - \mu_{14} - \mu_{31} + \mu_{34}$

i. **Effect** - **Select One of the following**:

Main Effect - F_1 **Main Effect -** F_2 **Interaction Effect -** $F_1 * F_2$ **Neither**

ii. **Estimable** (**Circle One**): **Yes** **No**

b. $\mathbf{C}_2 = \mu_{11} + \mu_{21} + \mu_{31} - \mu_{13} - \mu_{23} - \mu_{33}$

i. **Effect** - **Select One of the following**:

Main Effect - F_1 **Main Effect -** F_2 **Interaction Effect -** $F_1 * F_2$ **Neither**

ii. **Estimable** (**Circle One**): **Yes** **No**

c. $\mathbf{C}_5 = \mu_{11} - \mu_{14} + \mu_{21} - \mu_{24} + \mu_{31} - \mu_{34}$

i. **Effect** - **Select One of the following**:

Main Effect - F_1 **Main Effect -** F_2 **Interaction Effect -** $F_1 * F_2$ **Neither**

ii. **Estimable** (**Circle One**): **Yes** **No**

Answer the following questions but Do Not Submit them for grading:

Problem 4. The yield of a chemical process is under study. The most important variables are thought to be the pressure and the temperature in the reactor. Three levels of each of the two variables are selected for experimentation. A factorial experiment is conducted with two reps/treatment. During the experiment both reps of two treatments were not obtained. The yields from the remaining treatments are given in the following table:

	Pressure		
Temperature	Low	Medium	High
150	90.4	90.7	
	90.2	90.6	
200	90.1		89.9
	90.3		90.1
250	90.5	90.8	92.4
	90.7	90.9	92.1

a. Using contrasts in a Cell Means model, test the following hypotheses with F_1 = Temperature and F_2 = Pressure

  i. $H_o : (\mu_{11} + \mu_{12}) = (\mu_{31} + \mu_{32})$ AND $(\mu_{21} + \mu_{23}) = (\mu_{31} + \mu_{33})$

  Versus

  $H_1 : (\mu_{11} + \mu_{12}) \neq (\mu_{31} + \mu_{32})$ AND/OR $(\mu_{21} + \mu_{23}) \neq (\mu_{31} + \mu_{33})$

  ii. $H_o : (\mu_{11} + \mu_{31}) = (\mu_{12} + \mu_{32})$ AND $(\mu_{21} + \mu_{31}) = (\mu_{23} + \mu_{33})$

  Versus

  $H_1 : (\mu_{11} + \mu_{31}) \neq (\mu_{12} + \mu_{32})$ AND/OR $(\mu_{21} + \mu_{31}) \neq (\mu_{23} + \mu_{33})$

b. Interpret both of these sets of hypotheses in terms of the mean yield of the chemical process.

c. Compare your results to the output for Type IV Sums of Squares from running PROC GLM in SAS using an effects model ignoring the fact that some of the treatments were not observed in the experiment.

Problem 5. An experiment was run to evaluate the effect of two Factors on the mean of a response variable. There are three levels of factor F_1 and four levels of factor F_2. Initially, four reps of each of the 12 treatments were to be run. However, it was found that certain levels of F_1 were incompatible with certain levels of factor F_2. Thus, the experiment was unbalanced. The following **mean** responses were obtained for 9 of the 12 treatments are given below:

MEANS

	F_2			
F_1	1	2	3	4
1	$\bar{y}_{11}$	*	$\bar{y}_{13}$	$\bar{y}_{14}$
2	*	$\bar{y}_{22}$	$\bar{y}_{23}$	$\bar{y}_{24}$
3	$\bar{y}_{31}$	*	$\bar{y}_{33}$	$\bar{y}_{34}$

Suppose each mean in the above table is based on the results of 4 reps.

a. Write **TWO** contrasts in the 9 treatments which were observed in the experiment which would evaluate *main effect type* effects for factor F_1. Select contrasts which involve the maximum number of treatment means. Are your contrasts orthogonal? Justify your answer.

b. Write **TWO** contrasts in the 9 treatments which were observed in the experiment which would evaluate *interaction type* effects for factors F_1 and F_2. Select contrasts which involve the maximum number of treatment means. Are your contrasts orthogonal? Justify your answer.

Problem 6. For each of the following experiments provide an AOV table with Source of Variation, DF, and Expected Mean Squares.

a. Cholesterol was measured in the serum samples of five randomly selected patients from a large pool of patients. Two independent replicate tubes were prepared for each patient for each of four runs on a spectrophotometer. The objective of the study was to determine whether the relative cholesterol measurements for patients were consistent from run to run in the clinic. The data are mg/dl of cholesterol in the the replicate samples from each patient on each run.

b. An animal scientist conducted an experiment to study the effect of water quality on feedlot performance of steer calves. Four water quality treatments were used for the experiment. The feeding trial consisted of the four water quality treatments with two replicate pens of animals for each level of water treatment in a CRD. The trial was conducted on two consecutive summers. Thus we have a 4×2 factorial treatment structure with four water treatment levels and two summers. The data was the average daily weight gains for the 16 pens of steers.

c. An experiment was run with four factors A, B, C, and D with B nested within A, C nested within B, and D nested within C. All four factors have randomly selected levels thus producing the model:

$$y_{ijkl} = \mu + a_i + b_{j(i)} + c_{k(i,j)} + d_{l(i,j,k)} \text{ with } \; i = 1,2,3,4; \quad j = 1,2,3; \quad k = 1,2; \quad l = 1,2,3$$

d. An experiment was conducted with four factors, A with 3 fixed levels; B with 2 fixed levels; C nested within A and B with 6 random levels at each of the 6 levels of A and B; and D with 5 fixed levels. The experiment was a CRD with 6 replications. The model is given by

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + c_{k(i,j)} + \delta_l + (\alpha\delta)_{il} + (\beta\delta)_{jl} + (\alpha\beta\delta)_{ijl} + (cd)_{lk(i,j)} + e_{m(i,j,k,l)}$$

$$\text{with } \; i = 1,2,3; \quad j = 1,2; \quad k = 1,2,3,4,5,6; \quad l = 1,2,3,4,5; \quad m = 1,2,3,4,5,6$$