SB

STA 820 Final Exam – Thursday, Dec. 5
SHOW ALL WORK

Name: Key

1. (20 points) Fifteen guinea pigs were given a growth inhibiting substance and body weight measurements (in grams) were recorded at the ends of weeks 1,3,4,5,6, and 7. At the beginning of week 1 vitamin E therapy was started, the guinea pigs being randomly divided into three groups of five to receive zero, low, or high doses of vitamin E. The primary research question was whether the growth patterns differed between the groups. The data from this experiment are as follows:

		Week					
Group	Animal	1	3	4	5	6	7
Control	1	455	460	510	504	436	466
Control	2	467	565	610	596	542	587
Control	3	445	530	580	597	582	619
Control	4	485	542	594	583	611	612
Control	5	480	500	550	528	562	576
Low	6	514	560	565	524	552	597
Low	7	440	480	536	484	567	569
Low	8	495	570	569	585	576	677
Low	9	520	590	610	637	671	702
Low	10	503	555	591	605	649	675
High	11	496	560	622	622	632	670
High	12	498	540	589	557	568	609
High	13	478	510	568	555	576	605
High	14	545	565	580	601	633	649
High	15	472	498	540	524	532	583

Attached are print-outs of a SAS program "final1.sas" and the associated SAS output "final1.lst" for these data. Refer to these print-outs.

a. (4 points) Identify the design used here. Be specific.

One-way layout (CRD) w/ repeated measures

b. (4 points) Propose an appropriate model for the analysis of these data.

$y_{ijk} = \mu + \tau_i + \zeta_{ij} + \beta_k + (\tau\beta)_{ik} + e_{ijk}$

μ: overall mean; τ_i: i^{th} group effect; ζ_{ij}: w/error; β_k: k^{th} week effect; $(\tau\beta)_{ik}$: interaction group*week; e_{ijk}: s/error

c. (4 points) Test the null hypothesis that the mean body weights were the same in the three groups. State your conclusions.

$F = \frac{9274.0333}{8786.1833} = 1.056$

Can't reject

d. (4 points) Test contrasts to determine whether or not vitamin E has an effect and whether or not the dosage (Low vs. High) of vitamin E matters. State your conclusions.

$F_1 = \frac{17940.05}{8786.1833} = 2.042$ Neither significant (Bonferroni)

$F_2 = \frac{608.01667}{8786.1833} = 0.069$

e. (4 points) Test the null hypothesis that the mean body weights were the same across weeks (ignoring group). State your conclusion.

$F = 52.55$ Change over week (growth)

2. (16 points) To compare the effects of two fertilizers on the yield of tomatoes a gardener conducted an experiment on the tomato plants within one row of his garden. For each of the eleven plants within the row, the gardener flipped a coin to determine which fertilizer to apply to that plant. The resulting design and results (yield, in pounds of tomatoes) are given below:

	Tomato Plant										
	1	2	3	4	5	6	7	8	9	10	11
Fertilizer	A	A	B	B	A	B	B	B	A	A	B
Yield	29.9	11.4	26.6	23.7	25.3	28.5	14.2	17.9	16.5	21.1	24.3

Sums and Means:

Fertilizer	Sum	Mean
A	104.2	20.84
B	135.2	22.53
Total	239.4	21.76

$\sum_i \sum_j y_{ij}^2 = 5575.56$

a. (3 points) Identify the design used here. Be specific.

One-way layout

b. (3 points) State an appropriate model for the analysis of these data.

$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

$i = 1, \ldots, 2$ fertilizers

$j = 1, \ldots, n_i$ tomato plants

μ: overall mean; τ_i: i^{th} fert. effect; ε_{ij}: error

c. (6 points) Give the ANOVA table for these data.

$SS_T = \sum\sum y_{ij}^2 - \frac{y_{..}^2}{N} = 5575.56 - \frac{239.4^2}{11} = 365.35$

$SS_{Trt} = \sum \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} = \left(\frac{104.2^2}{5} + \frac{135.2^2}{6}\right) - \frac{239.4^2}{11} =$

7.82

Source	SS	df	MS	F
Fertilizer	7.82	1	7.82	0.197
Error	357.53	9	39.73	
Total	365.35	10		

d. (4 points) Based on the ANOVA Table that you computed in part (c), test the hypothesis that the mean yields are the same for the two fertilizers. Suggest an alternative (equivalent) method for testing this null hypothesis.

$F = 0.197$ do not reject

2 simple t-test.

3. (14 points) These data are from a clinical trial on 29 patients of a new drug for the treatment of enuresis (bed-wetting). The patients were given the drug for 14 days and a placebo for a separate 14 days, the order of administration being chosen randomly for each patient. For each patient, the two 14 day treatment periods were separated by a ten day period in which no treatments were applied. The data show the number of dry nights out of 14.

	Period					Period			
	1		2			1		2	
Patient	Trt.	Resp.	Trt.	Resp.	Patient	Trt.	Resp.	Trt.	Resp.
1	A	8	B	5	16	A	13	B	12
2	B	12	A	11	17	B	2	A	4
3	A	14	B	10	18	A	10	B	2
4	A	8	B	0	19	A	7	B	5
5	B	6	A	8	20	B	8	A	13
6	A	9	B	7	21	A	13	B	13
7	A	11	B	7	22	A	8	B	10
8	B	13	A	9	23	B	9	A	7
9	A	3	B	5	24	A	7	B	7
10	B	8	A	8	25	A	9	B	0
11	A	6	B	0	26	B	7	A	10
12	B	8	A	9	27	A	10	B	6
13	A	0	B	0	28	A	2	B	2
14	B	4	A	8	29	B	7	A	6
15	B	8	A	14					

a. (3 points) Identify the design used here. Be specific.

Crossover

b. (5 points) Explain why the ten day period between the two treatment periods is an important feature of this experiment.

washout remove carry-over effects

c. (6 points) Complete the following ANOVA Table for these data:

Source of Variation	Sum of Squares	d.f.	Mean Squares	F
Treatment	56.69	1	56.69	10.66
Individuals	603.66	28		
Time Periods	17.66	1		
Error	143.65	27	5.32	
Total	821.66	57		

4. (15 points) A split-plot experiment was conducted to determine the effects of cooking method and bean variety on the tenderness of beans. A bean field was divided into a four-by-four grid of 16 plots. Because it was believed that there would be systematic variability among the rows and among the columns, both rows and columns were used as blocking factors. The four varieties were planted in such a way that each variety occurred exactly once in each row and once in each column. The bean harvest within each plot was divided into three samples and each of the three samples for each block was randomly assigned to one of three cooking methods. The tenderness of the cooked beans was then rated by an expert taster for each of the 48 resulting samples.

Identify the following components of this experiment:

a. (1.5 points) The whole plots. 16 plots in the grid

b. (1.5 points) The split plots. 48 sub plots (3 in each WP)

c. (1.5 points) The whole plot factor.

Variety

d. (1.5 points) The split plot factor.

Cooking method

e. (1.5 points) The whole plot treatment structure.

One-way

f. (1.5 points) The whole plot design structure.

LS.

g. (1.5 points) The split plot treatment structure.

One-way

h. (1.5 points) The split plot design structure.

RCB w/ reps

i. (1.5 points) The entire design treatment structure.

two-way factorial

j. (1.5 points) The entire design design structure.

Triple block.

5. (14 points) An experiment was conducted to determine the effects of four treatments (A, B, C, and D) on current flow in television tube filaments. The experiment was conducted over several days and it was expected that day-to-day variability might be a significant nuisance factor. Therefore, the experiment was conducted treating days as blocks. Because of the time-consuming nature of the treatments, only three of the four treatments could be observed during any particular day. The resulting incomp lete block design and data (current flow) are given below.

	Treatments					
Day	A	B	C	D	$y_{\cdot j}$	$\bar{y}_{\cdot j}$
1	2	–	20	7	29	9.67
2	–	32	14	3	49	16.33
3	4	13	31	–	48	16.0
4	0	23	–	11	34	11.33
$y_{i\cdot}$	6	68	65	21	$y_{\cdot\cdot} = 160$	
$\bar{y}_{i\cdot}$	2.0	22.67	21.67	7.0		$\bar{y}_{\cdot\cdot} = 13.33$

Refer to the attached SAS program "final5.sas" and output "final5.lst" to complete this problem.

a. (6 points) Test the null hypothesis that the four treatment means are all equal. State your conclusions.

$F = 4.04 \quad p = .08$

do not reject at $\alpha = .05$

no evidence of diffs among mean current flows for the 4 treatments.

b. (8 points) Test a contrast comparing the means for treatments A and D with the means for treatments B and C. State your conclusions. (Hint: LSMEANS gives the $\hat{\mu}_i$s).

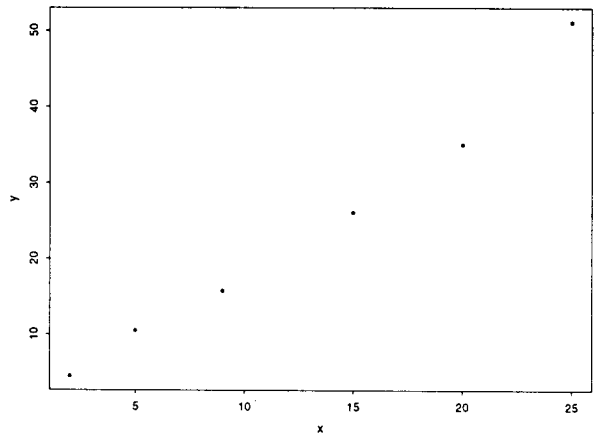
$C = 1.70833 - 22.45833 - 21.95833 + 7.20833 = -35.5$

$$t = \frac{|-35.5|}{\sqrt{\frac{3}{2(4)}(72.633)}\sqrt{4}} = 3.40$$

$t_{.025}(5) = 2.571$

So reject.

6. (9 points) A statistician is analyzing data from a one-way layout for a client. She suspects that the assumption of constant variance may be violated for these data. Therefore, she performed Levene's test of homogeneity of variance and rejected the null hypothesis. A plot of $\log(s_i)$ (the natural logarithm of the sample standard deviation within the i^{th} treatment) versus $\log(\bar{y}_{i\cdot})$ (the natural logarithm of the sample mean within the i^{th} treatment) is as follows:



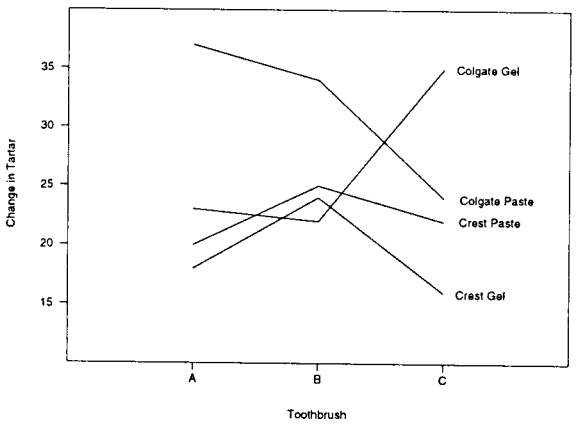
Notice that the scales are different on the x and y axis for this plot. What transformation should she apply to these data to correct the heterogeneity of variances problem? Provide some (brief) justification for your answer (I want to be sure you didn't guess).

$\hat{\alpha} = 2$ $\quad \lambda = 1 - 2 = -1$

$y^{\lambda} = \frac{1}{y}$

inverse transformation

7. (12 points) An experiment was performed to determine the effectiveness of four different toothpastes (Crest Tartar Control Gel, Crest Tartar Control Paste, Colgate Tartar Control Gel, Colgate Tartar Control Paste) and three different toothbrushes (A, B, C) for removing tartar. In the experiment, toothpastes and toothbrushes were crossed to yield 12 treatments. Twenty-four subjects previously determined to have similar baseline levels of tartar were randomized to the 12 treatments such that two sub jects received each treatment. Each subject used the assigned combination of toothpaste and toothbrush for one month. The response measured on each subject was the change in tartar over the one month treatment period. A profile plot of the 12 resulting treatment means is as follows:



A model was fit to these data using the following SAS MODEL statement:

MODEL TARTAR=TOOTHPST TOOTHBSH TOOTHPST*TOOTHBSH;

For each of the following sources of variation state whether or not you expect statistical significance. Briefly justify your answer based on the above profile plot.

a. (2 points) TOOTHPST

Yes obvious diffs among 4 toothpastes (colgate paste higher)

b. (2 points) TOOTHBSH

No mean stays about same

c. (2 points) TOOTHPST*TOOTHBSH

Yes strong interactions w/ lines crossing + going in opposite directions

Using CONTRAST statements, additional effects and interactions were investigated. For each of the following additional sources of variations state whether or not you expect statistical significance. Briefly justify your answer based on the profile plot.

d. (2 points) CREST*TOOTHBSH (interaction between Crest brand toothpastes and toothbrushes)

No, close to //

e. (2 points) PASTE*TOOTHBSH (interaction between paste-type toothpastes and toothbrushes)

Yes,  lines clearly not //

f. (2 points) BRANDS (effects of brands – Crest versus Colgate – ignoring toothpaste-type and toothbrush)

Yes, Colgate clearly higher.

final1.sas:

```
options ls=78;

data diet1;
  infile 'diet1.dat';
  input group $ animalno week bodywgt;
run;

proc glm data=diet1;
  class group week animalno;
  model bodywgt=group animalno(group) week group*week / ss3;
  contrast 'Vit. E vs control' group 2 -1 -1 ;
  contrast 'Low E vs. High E' group 0 1 -1 ;
  lsmeans group;
  lsmeans week;
  lsmeans group*week;
run;
```

final1.lst:

General Linear Models Procedure
Class Level Information

Class	Levels	Values
GROUP	3	Control High Low
WEEK	6	1 3 4 5 6 7
ANIMALNO	5	1 2 3 4 5

Number of observations in data set = 90

General Linear Models Procedure

Dependent Variable: BODYWGT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	276299.5000	9527.5690	17.56	0.0001
Error	60	32552.6000	542.5433		
Corrected Total	89	308852.1000			

R-Square	C.V.	Root MSE	BODYWGT Mean
0.894601	4.166081	23.29256	559.1000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	2	18548.0667	9274.0333	17.09	0.0001
ANIMALNO(GROUP)	12	105434.2000	8786.1833	16.19	0.0001
WEEK	5	142554.5000	28510.9000	52.55	0.0001
GROUP*WEEK	10	9762.7333	976.2733	1.80	0.0801

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Vit. E vs control	1	17940.05000	17940.05000	33.07	0.0001
Low E vs. High E	1	608.01667	608.01667	1.12	0.2940

General Linear Models Procedure
Least Squares Means

GROUP	BODYWGT LSMEAN
Control	539.133333
High	565.900000
Low	572.266667

General Linear Models Procedure
Least Squares Means

WEEK	BODYWGT LSMEAN
1	486.200000
3	535.000000
4	574.266667
5	566.800000
6	579.266667
7	613.066667

General Linear Models Procedure
Least Squares Means

GROUP	WEEK	BODYWGT LSMEAN
Control	1	466.400000
Control	3	519.400000
Control	4	568.800000
Control	5	561.600000
Control	6	546.600000
Control	7	572.000000
High	1	497.800000
High	3	534.600000
High	4	579.800000
High	5	571.800000
High	6	588.200000
High	7	623.200000
Low	1	494.400000
Low	3	551.000000
Low	4	574.200000
Low	5	567.000000
Low	6	603.000000
Low	7	644.000000

final5.sas:

```
options ls=78;

data tv;
  input treat $ day current;
  cards;
A 1 2
A 3 4
A 4 0
B 2 32
B 3 13
B 4 23
C 1 20
C 2 14
C 3 31
D 1 7
D 2 3
D 4 11
;
run;

proc glm data=tv;
  class day treat;
  model current=treat day;
  lsmeans treat;
  means treat;
run;
```

final5.lst:

General Linear Models Procedure
Class Level Information

Class	Levels	Values
DAY	4	1 2 3 4
TREAT	4	A B C D

Number of observations in data set = 12

General Linear Models Procedure

Dependent Variable: CURRENT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	981.5000000	163.5833333	2.25	0.1954
Error	5	363.1666667	72.6333333		
Corrected Total	11	1344.6666667			

R-Square	C.V.	Root MSE	CURRENT Mean
0.729921	63.91889	8.522519	13.33333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TREAT	3	975.3333333	325.1111111	4.48	0.0701
DAY	3	6.1666667	2.0555556	0.03	0.9928

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TREAT	3	880.8333333	293.6111111	4.04	0.0834
DAY	3	6.1666667	2.0555556	0.03	0.9928

General Linear Models Procedure
Least Squares Means

TREAT	CURRENT LSMEAN
A	1.7083333
B	22.4583333
C	21.9583333
D	7.2083333

General Linear Models Procedure

Level of TREAT	N	CURRENT Mean	CURRENT SD
A	3	2.0000000	2.00000000
B	3	22.6666667	9.50438495
C	3	21.6666667	8.62167810
D	3	7.0000000	4.00000000