MSDS UNIT 3 HW 3

PART I

1. Back in Unit 1 we considered a study in which 4 different fertilizers were tested by researchers for their yield (in mm of growth) on a local grass: Red Fescue (Just for fun .. this is a real mountain grass! http://fescue.com/info/creepingred.html#.WIu6rbYrKu4). To conduct their study researchers had enough money to run three replicates of each fertilizer. They knew the red fescue was a mountain grass so they went out to the mountains and carefully fertilized plots of land as you see in the diagram below. The data is contained in the ***growth3*** data set.

Given the information you have available, run a simple ANOVA to test for the effects (if any) of the fertilizers. **From a previous study, we know that the yields from each fertilizer are normally distributed with equal variances. For now assume that independence is not a concern here. You may assume the assumptions are met for all questions in this homework.**

Deliverables: 1. ANOVA table. 2. Means Plot (Interaction Plot from SAS) 2. Conclusion for the ANOVA. 3. Confidence intervals with multiple comparison corrections for **SIGNIFANT** differences (between Fertilizers). 4. SAS Code: proc glm or Proc mixed code. Your answer should fit in the given box.

The GLM Procedure

Dependent Variable: Growth

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	234.2500000	78.0833333	2.83	0.1064
Error	8	220.6666667	27.5833333		
Corrected Total	11	454.9166667			

R-Square	Coeff Var	Root MSE	Growth Mean
0.514929	30.74332	5.251984	17.08333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Fertilizer	3	234.2500000	78.0833333	2.83	0.1064

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Fertilizer	3	234.2500000	78.0833333	2.83	0.1064

Fertilizer	Growth LSMEAN	95% Confidence Limits	
1	14.000000	7.007655	20.992345
2	19.000000	12.007655	25.992345
3	12.000000	5.007655	18.992345
4	23.333333	16.340989	30.325678

With only one factor, no interaction plot is needed. From the ANOVA, there is not sufficient evidence to suggest that the mean growth of plants under different fertilizers is different (p-value .1064 from an ANOVA).

```
proc glm data=growth3 plots=(diagnostics residuals);
```

```
class fertilizer;
model growth = fertilizer;
lsmeans fertilizer / pdiff tdiff  CL adjust=bon;
run;
```

2. Now assume that you get to thinking about it and realize that we may really be looking at three different environments here: Sunny, Wetlands and Mostly Shady. This data has been recorded in the ***growth4*** data set. Conduct a similar analysis as in 1 but now account for the environment variable (ENV).

Deliverables: 1. ANOVA table. 2. Means Plot (Interaction Plot from SAS) 2. Conclusion for the ANOVA (only step 6 is required.) 3. The researchers were specifically interested in (a) Inference on the main effect of the fertilizer. (b) If Fertilizer 4 performed significantly better in one environment than another. (c) Which fertilizer will perform best in the wetlands. Answer each QOI (question of interest) with a confidence interval and a 1 sentence conclusion. Your answer should fit in the space below (on this page.) 4. SAS Code: Proc glm or proc mixed statement. Answer should fit in the given box.

The GLM Procedure

Dependent Variable: Growth

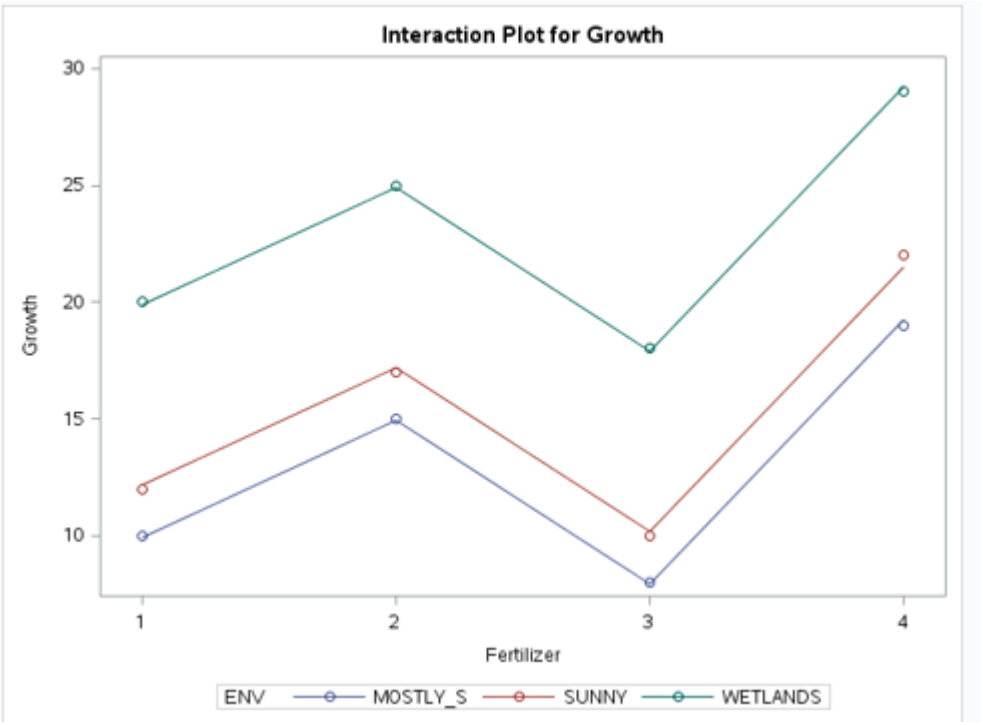
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	454.4166667	90.8833333	1090.60	<.0001
Error	6	0.5000000	0.0833333		
Corrected Total	11	454.9166667			

R-Square	Coeff Var	Root MSE	Growth Mean
0.998901	1.689806	0.288675	17.08333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Fertilizer	3	234.2500000	78.0833333	937.00	<.0001
ENV	2	220.1666667	110.0833333	1321.00	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Fertilizer	3	234.2500000	78.0833333	937.00	<.0001
ENV	2	220.1666667	110.0833333	1321.00	<.0001

There is strong evidence from the ANOVA that at least one fertilizer and/or at at least one environment group has a deferent mean growth score than the others (pvalue < .0001).



The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Fertilizer	Growth LSMEAN	LSMEAN Number
1	14.0000000	1
2	19.0000000	2
3	12.0000000	3
4	23.3333333	4

Least Squares Means for effect Fertilizer
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Growth

i/j	1	2	3	4
1		<.0001	0.0009	<.0001
2	<.0001		<.0001	<.0001
3	0.0009	<.0001		<.0001
4	<.0001	<.0001	<.0001	

Fertilizer	Growth LSMEAN	95% Confidence Limits	
1	14.000000	13.592181	14.407819
2	19.000000	18.592181	19.407819
3	12.000000	11.592181	12.407819
4	23.333333	22.925515	23.741152

Least Squares Means for Effect Fertilizer

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-5.000000	-5.910516	-4.089484
1	3	2.000000	1.089484	2.910516
1	4	-9.333333	-10.243849	-8.422818
2	3	7.000000	6.089484	7.910516
2	4	-4.333333	-5.243849	-3.422818
3	4	-11.333333	-12.243849	-10.422818

```
proc glm data= growth4 plots = (ResidualPanel);
class fertilizer env;
model growth = fertilizer env ;
lsmeans fertilizer env / CL adjust = bon;
run;
```

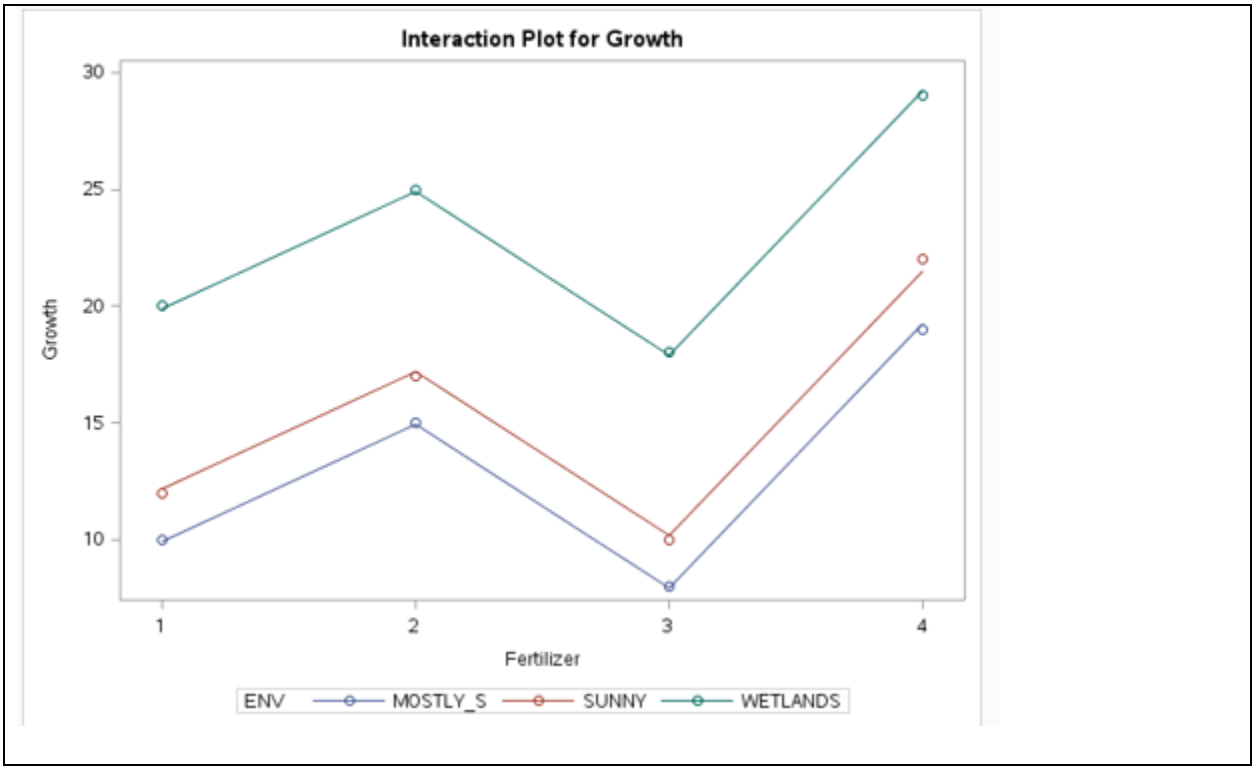
3a. There is strong evidence that, after accounting for the environments, there is a significant difference in mean growth between at least 1 pair of fertilizers from the Type 3 tests. The table above indicates that each fertilizer has a mean growth that is significantly different that the others. For instance, the mean growth with Fertilizer 1 is estimated to be between 5.9mm and 4.1mm less than the mean growth with fertilizer 2, also Fertilizer 3 is estimated to be between 12.2 and 10.4mm less than the mean growth with fertilizer 4(after a Bonferroni adjustment).

```
proc glm data= growth4 plots = (ResidualPanel);
class fertilizer env;
model growth = fertilizer | env ;
lsmeans fertilizer env / CL adjust = bon;
run;
```

3b. If we model each Fertilizer / Environment combination separately (the non-addititve model), while fertilizer 4 performed best in the Wetlands, we only had one observation at this fertilizer / environment combination and thus we do not have an estimate of the standard deviation in which to run a test. For the same reason, we cannot construct a confidence interval for the difference between the mean growth with fertilizer 4 in the different environments.

3c. Again, while fertilizer 4 performed best in the Wetlands, we only had one observation at each fertilizer / environment combination and thus we do not have an estimate of the standard deviation in which to run a test. For the same reason, we cannot construct a confidence interval for the difference between the mean growth with the different fertilizers in the Wetlands.

3. Consider the model you used in problem 2. Inspect the Means Plot (Interaction Plot). Does it look like there will be a significant interaction? Explain by interpreting what an interaction is and then comparing that to what you see in the plot.



An interaction in this case would be one in which the effect of the fertilizer depended on the environment that the fertilizer was used in. Since all the lines in the interaction plot above are close to parallel, there is not a lot of evidence of an interaction effect.

4. Fit the full model that includes the interaction term? What do you notice? Why? Discuss.

The GLM Procedure

Dependent Variable: Growth

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	454.9166667	41.3560606	.	.
Error	0	0.0000000	.		
Corrected Total	11	454.9166667			

R-Square	Coeff Var	Root MSE	Growth Mean
1.000000	.	.	17.08333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Fertilizer	3	234.2500000	78.0833333	.	.
ENV	2	220.1666667	110.0833333	.	.
Fertilizer*ENV	6	0.5000000	0.0833333	.	.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Fertilizer	3	234.2500000	78.0833333	.	.
ENV	2	220.1666667	110.0833333	.	.
Fertilizer*ENV	6	0.5000000	0.0833333	.	.

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Fertilizer	Growth LSMEAN	LSMEAN Number
1	14.0000000	1
2	19.0000000	2
3	12.0000000	3
4	23.3333333	4

Least Squares Means for effect Fertilizer
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Growth

i/j	1	2	3	4
1		.	.	.
2	.		.	.
3	.	.		.
4	.	.	.	

Fertilizer	Growth LSMEAN	95% Confidence Limits	
1	14.000000	.	.
2	19.000000	.	.
3	12.000000	.	.
4	23.333333	.	.

Least Squares Means for Effect Fertilizer

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-5.000000	.	.
1	3	2.000000	.	.
1	4	-9.333333	.	.
2	3	7.000000	.	.
2	4	-4.333333	.	.
3	4	-11.333333	.	.

With a degrees of freedom of '0' so we do not have an estimate of the MSE and can't perform any tests or construct any confidence intervals.
This underlines the importance of repeat observations at different treatments (fertilizer/environment combinations).

```
proc glm data= growth4 plots = (ResidualPanel);
class fertilizer env;
model growth = fertilizer | env ;
lsmeans fertilizer env / CL adjust = bon;
run;
```

Extra credit: Repeat number 2. With a regression model instead of a 2 way ANOVA. Show that the results are the same.
Deliverables: 1. Parameter Estimate Table, 2. Written conclusion showing that the estimates from the regression analysis is equivalent to the ANOVA analysis. 3. SAS proc glm or proc mixed code.

PART II

Exercise 13.15 :

Blood–Brain Barrier. Analyze the effect of the design variables—sacrifice time and treatment—on the log of the ratio of brain count to liver count in the data set described in Section 11.1.2 (file: case1102). (a) Ignore the covariates and use an analysis of variance procedure to fit the data. Fit a model that includes interaction terms; plot the residuals versus the fitted values. (b) Test whether there is an interactive effect of treatment and sacrifice time. What are the F-statistic, the degrees of freedom, and the p-value? (c) If there are no interactive effects, test whether there are main effects of treatment and sacrifice time. (d) Complete the analysis by describing the effects of treatment and sacrifice time, either by estimating the appropriate contrasts or by using a regression procedure with indicator variables to model treatment (one indicator) and sacrifice time (three indicators).

PatNum	Brain	Liver	Time	Treatment
1	41081	1456164	0.5	BD
2	44286	1602171	0.5	BD
3	102926	1601936	0.5	BD
4	25927	1776411	0.5	BD
1	31342	1790863	0.5	NS
2	22815	1633386	0.5	NS
3	16629	1618757	0.5	NS
4	22315	1567602	0.5	NS
5	77961	1060057	3	BD
6	73178	715581	3	BD
7	76167	620145	3	BD
8	123730	1068423	3	BD
9	25569	721436	3	NS
10	33803	1019352	3	NS
11	24512	667785	3	NS
12	50545	961097	3	NS
1	84616	48815	24	BD
2	55153	16885	24	BD
3	48829	22395	24	BD
4	89454	83504	24	BD
5	37928	20323	24	NS
6	12816	15985	24	NS
7	23734	25895	24	NS
8	31097	33224	24	NS
9	35395	4142	72	BD
10	18270	2364	72	BD

11	5625	1979	72	BD
12	7497	1659	72	BD
13	6250	928	72	NS
14	11519	2423	72	NS
15	3184	1608	72	NS
16	1334	3242	72	NS

PatNum	Brain	Liver	Time	Treatment	depVar
1	41081	1456164	0.5	BD	-3.568
2	44286	1602171	0.5	BD	-3.5884
3	102926	1601936	0.5	BD	-2.745
4	25927	1776411	0.5	BD	-4.2271
5	31342	1790863	0.5	NS	-4.0455
6	22815	1633386	0.5	NS	-4.271
7	16629	1618757	0.5	NS	-4.5783
8	22315	1567602	0.5	NS	-4.252
9	77961	1060057	3	BD	-2.6099
10	73178	715581	3	BD	-2.2802
11	76167	620145	3	BD	-2.097
12	123730	1068423	3	BD	-2.1558
13	25569	721436	3	NS	-3.3399
14	33803	1019352	3	NS	-3.4064
15	24512	667785	3	NS	-3.3048
16	50545	961097	3	NS	-2.9452
17	84616	48815	24	BD	0.5501
18	55153	16885	24	BD	1.1837
19	48829	22395	24	BD	0.7795
20	89454	83504	24	BD	0.0688
21	37928	20323	24	NS	0.6239
22	12816	15985	24	NS	-0.221
23	23734	25895	24	NS	-0.0871
24	31097	33224	24	NS	-0.0662
25	35395	4142	72	BD	2.1454
26	18270	2364	72	BD	2.0449
27	5625	1979	72	BD	1.0446
28	7497	1659	72	BD	1.5083
29	6250	928	72	NS	1.9073
30	11519	2423	72	NS	1.559
31	3184	1608	72	NS	0.6831
32	1334	3242	72	NS	-0.888

N	32
n	4
#levels Time	4
#levels Trt	2
ybar All	-1.2681

Time	ybar
0.5	-3.9094
3	-2.7674
24	0.35396
72	1.25058

trt	ybar
BD	-0.8716
NS	-1.6645

ybar		
Time\trt	BD	NS
0.5	-3.5321	-4.2867
3	-2.2857	-3.2491
24	0.64553	0.0624
72	1.6858	0.81535

sigma2		
Time\trt	BD	NS
0.5		
3		
24		
72		

Source of Variation	ss	df	MS	F	Pr(F)
Between Groups	150.784	7	21.541	64.073	0.0001
Time	145.594	3	48.531	144.36	0.0001
Trt	5.029	1	5.029	14.96	0.0007
Interaction	0.161	3	0.0536	0.16	0.9224
Error	8.068	24	0.0336		
Total	158.852				