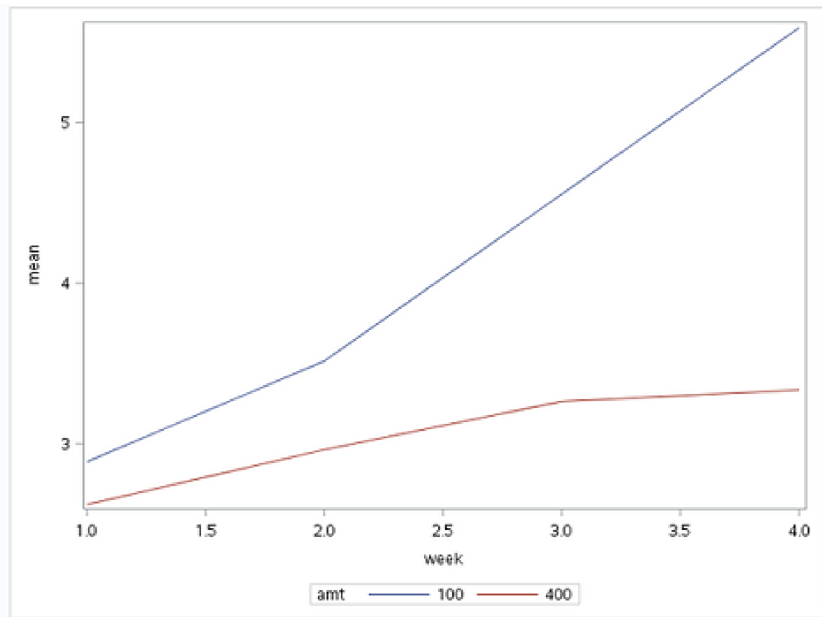
Analysis of Preservative Data

Two different amounts of a preservative were applied to packages of a food product. There were 32 packages in all, 16 that received 100 units of preservative and the other 16 that received 400 units of the preservative. Four of the packages with 100 units of preservative and four with 400 units of preservative were selected for analysis each week, and this was done for a total of four weeks. Bacterial counts were made on each package. The logarithms of bacterial counts were analyzed.



The interaction plot (above) indicates that bacterial counts increase more rapidly under amount 100 than amount 400. Thus one would expect to see interaction in the analysis. The two amounts produce about the same bacterial counts in the first two weeks but appear to separate after that. An important question is this: At what week is there a significant difference between the means?

There is a significant interaction between amount and week ($F = 10.99$, $p<.0001$), so we do not try to interpret the main effects. Instead, we should compare the mean bacterial counts week by week. For week 1 and again at week 2, there is not a significant difference in log bacteria count between amounts 100 and 400 ($p=0.3283$ and $p=0.0503$). Significant differences emerge at week 3 and continue to week 4 (both $p<.0001$).

To put this in practical terms, if the food product is stored for 2 weeks or less, either 100 or 400 units of the preservative will results in (statistically) the same bacterial count. If the food product is stored for longer periods, 400 units of the preservative should be used to inhibit bacterial growth.

Explanation

The preceding page is all you need to turn in for the data analysis problems. I am including the following additional information to help you understand why I included (and excluded) some of the information in the output.

We should look first at the overall F test ($F = 27.62$, $p<.0001$). I did not include this on the previous page because the interaction is significant ($F = 10.99$, $p<.0001$). If the interaction is significant, then the overall F test will be significant, too.

Now that we know the interaction is significant, **IGNORE THE MAIN EFFECTS AND MARGINAL MEANS**. These are meaningless in the presence of interaction.

In the context of the problem, we are using a preservative which presumably inhibits bacterial growth. Common sense tells us that the amount of bacteria will increase with time. From the interaction plot, it is clear that the two levels of preservative have different trajectories (but this difference may not be statistically significant). Since there is interaction, the logical choice is to compare the two levels of preservative at each week. There are 4 weeks, so this will result in 4 tests. If we apply a Bonferroni correction for multiple comparisons, the level of significance for each test should be $0.05/4 = 0.0125$.

The p-values for the weekly comparisons are obtained from the table of pairwise differences (which is the result of the 'pdiff' option on the 'lsmeans' statement). The following pages contain the SAS code and most of the output. Note that there is a great deal of the output that we <u>do not consider</u> because there is a significant interaction.

SAS Code and Annotated Output

```
data preservatives;
input package amt week logcount @@;
datalines;
1  100  1  2.85    2  100  1  2.21    3  100  1  3.20    4  100  1  3.29
5  100  2  3.81    6  100  2  3.45    7  100  2  3.20    8  100  2  3.60
9  100  3  4.46    10 100  3  4.47    11 100  3  4.15    12 100  3  5.14
13 100  4  5.60    14 100  4  5.33    15 100  4  5.94    16 100  4  5.50
17 400  1  1.93    18 400  1  2.82    19 400  1  2.93    20 400  1  2.80
21 400  2  2.51    22 400  2  3.05    23 400  2  3.66    24 400  2  2.63
25 400  3  2.99    26 400  3  3.01    27 400  3  3.50    28 400  3  3.55
29 400  4  3.44    30 400  4  3.50    31 400  4  3.25    32 400  4  3.15
;
run;

proc sort data=preservatives;
by amt week;
 run;

proc means data=preservatives noprint;
 by amt week;
 var logcount;
 output out=means mean=mean;
 run;

proc sgplot data=means;
 series x=week y=mean / group=amt;
 run;



proc glm data=preservatives;
 class amt week;
 model logcount = amt week amt*week / ss3;
 lsmeans amt week amt*week / stderr pdiff;
 run;

quit;
```

The following information is about the levels of the factors and the number of observations. It is good to check this to see if you get what you expected. For instance, if an amount had been mistakenly entered asb300, it would show up here.

The GLM Procedure

Class Level Information

Class	Levels	Values
amt	2	100 400
week	4	1 2 3 4

Number of Observations Read	32
Number of Observations Used	32

This is the part of the analysis that is equivalent to doing a one-way ANOVA on the 8 amount by week combinations. Clearly there is something going on because p<.0001 but it is not particularly informative. The Root MSE may be of interest in determining sample sizes in future studies or in just giving an indication of the variability in the data.

The GLM Procedure

Dependent Variable: logcount

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	27.78915000	3.96987857	27.62	<.0001
Error	24	3.44940000	0.14372500		
Corrected Total	31	31.23855000			

R-Square	Coeff Var	Root MSE	logcount Mean
0.889579	10.55651	0.379111	3.591250

Here we see the components of the ANOVA. All factors are highly significant. The most important thing, however, is that interaction is significant. It tell us that both factors are important and that we should look at treatment means rather than marginal means to explain how the two factors affect the response.

Source	DF	Type III SS	Mean Square	F Value	Pr < F
amt	1	9.54845000	9.54845000	66.44	<.0001
week	3	13.50180000	4.50060000	31.31	<.0001
amt*week	3	4.73890000	1.57963333	10.99	<.0001

These are the marginal means for amount. Because of the significant interaction between amount and weeks, we should IGNORE THIS PART OF THE OUTPUT and instead compare the mean bacterial counts week by week.

Least Squares Means

amt	logcount LSMEAN	Standard Error	H0:LSMEAN=0 Pr > |t|	H0:LSMean1=LSMean2 Pr > |t|
100	4.13750000	0.09477770	<.0001	<.0001
400	3.04500000	0.09477770	<.0001	

These are the marginal means for week. The bacteria counts increase on average from week to week, which is what we expect for bacterial growth. SINCE THERE IS AN INTERACTION, THE DIFFERENCES AMONG THE MARGINAL MEANS SHOULD BE IGNORED.

week	logcount LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
1	2.75375000	0.13403591	<.0001	1
2	3.23875000	0.13403591	<.0001	2
3	3.90875000	0.13403591	<.0001	3
4	4.46375000	0.13403591	<.0001	4

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

Dependent Variable: logcount

i/j	1	2	3	4
1		0.0172	<.0001	<.0001
2	0.0172		0.0017	<.0001
3	<.0001	0.0017		0.0074
4	<.0001	<.0001	0.0074	

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.

Here are the means for the 8 combinations of amount by week and the p-values for the pairwise comparisons of means. The means of 100 versus 400 for each week are the comparisons of interest. There are four such comparisons, and their p-values are highlighted in red. These p-values are not adjusted for multiple comparisons (because we did not include an 'adjust=' option on the 'lsmeans' statement). With a Bonferroni adjustment, differences are significant only if the p-value is less than 0.05/4 = 0.0125.

amt	week	logcount LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
100	1	2.88750000	0.18955540	<.0001	1
100	2	3.51500000	0.18955540	<.0001	2
100	3	4.55500000	0.18955540	<.0001	3
100	4	5.59250000	0.18955540	<.0001	4
400	1	2.62000000	0.18955540	<.0001	5
400	2	2.96250000	0.18955540	<.0001	6
400	3	3.26250000	0.18955540	<.0001	7
400	4	3.33500000	0.18955540	<.0001	8

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

Dependent Variable: logcount

i/j	1	2	3	4	5	6	7	8
1		0.0279	<.0001	<.0001	0.3283	0.7820	0.1746	0.1080
2	0.0279		0.0007	<.0001	0.0027	0.0503	0.3556	0.5083
3	<.0001	0.0007		0.0007	<.0001	<.0001	<.0001	0.0001
4	<.0001	<.0001	0.0007		<.0001	<.0001	<.0001	<.0001
5	0.3283	0.0027	<.0001	<.0001		0.2136	0.0247	0.0135
6	0.7820	0.0503	<.0001	<.0001	0.2136		0.2742	0.1774
7	0.1746	0.3556	<.0001	<.0001	0.0247	0.2742		0.7891
8	0.1080	0.5083	0.0001	<.0001	0.0135	0.1774	0.7891	

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.

This is a reminder that these p-values have not been adjusted for multiple comparisons.