

Stat 401 A XM Homework 12

This is an optional homework designed to help you prepare for the final. Don't feel obligated to write out answers to every one of these questions. Use them to help you as much or as little as you like. Solutions are available on the course web site.

1. A scientist studied the effect of amount of fertilizer and amount of moisture on the yield of a new tomato hybrid. The investigation was carried out in a greenhouse where the two factors of interest could be precisely controlled. Three fertilizer levels (10, 15, and 20 units) and three moisture levels (1, 2, and 3 inches) were considered. Eight pots of tomato plants were randomly assigned to each fertilizer/moisture combination. The mean and variance (rounded to the nearest tenth) of the 8 yields associated with each fertilizer/moisture combination are provided in the tables below.

Means	Amount of Moisture			Variances	Amount of Moisture		
Fertilizer Amount	1	2	3	Fertilizer Amount	1	2	3
5	48.7	54.4	53.7	5	41.0	67.6	59.7
10	54.5	56.1	57.2	10	40.3	78.2	63.7
15	53.6	65.3	59.9	15	86.8	65.2	52.2

You may assume that all of the required assumptions are satisfied for this data set. In fact, the data is not from an experiment as described above. Instead, it has been randomly generated from a model in such a way that all the assumptions are satisfied. The homework solutions show the true model that generated the data.

- (a) Conduct a one-way analysis of variance for this data. Use each combination of fertilizer amount and moisture amount as a treatment. Test

$$H_0 : \text{The treatment means are equal.} \quad \text{vs.} \quad H_A : \text{Some treatment means differ.}$$

Give a test statistic, its degrees of freedom, a p -value, and a conclusion in nontechnical language.

- (b) Compute a 95% confidence interval for the mean yield of tomatoes exposed to 10 units of fertilizer and 2 inches of moisture.
 - (c) Compute a 95% confidence interval for the mean yield of tomatoes exposed to 5 units of fertilizer and 1 inch of moisture minus the mean yield of tomatoes exposed to 5 units of fertilizer and 2 inches of moisture.
 - (d) The Tukey method shows that two treatment sample means must differ by 12.6 to be significant at an overall significance level of 0.05. Make a line plot that allows you to see significant and nonsignificant differences among the 9 means.
 - (e) Compute a confidence interval for the difference between the mean yield of tomatoes that received 5 units of fertilizer and the mean yield of tomatoes that received 10 units of fertilizer.
2. Conduct a two-factor analysis of variance for the data described in problem 1. Use the SAS code and incomplete output below to help you answer the following questions.

```

proc glm;
  class fert moisture;
  model yield=fert moisture fert*moisture;
  means fert / tukey;
  means moisture / tukey;
run;

```

The GLM Procedure

```

      Class Level Information
Class      Levels      Values
fert              3      5 10 15
moisture         3      1 2 3
Number of observations      72

```

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	??	1373.798	??	??	??
Error	??	3883.332	??	??	??
Corrected Total	??	5257.130	??	??	??

R-Square	Coeff Var	Root MSE	yield Mean
0.261321	14.04015	7.851127	55.91912

Source	DF	Type III SS	Mean Square	F Value	Pr > F
fert	??	645.1400831	??	??	??
moisture	??	516.4980308	??	??	??
fert*moisture	??	212.1598440	??	??	??

Tukey's Studentized Range (HSD) Test for yield

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	63
Error Mean Square	61.64019
Critical Value of Studentized Range	3.39458
Minimum Significant Difference	5.4402

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	fert
A	59.584	24	15

	A			
B	A	55.921	24	10
B				
B		52.252	24	5

Tukey's Studentized Range (HSD) Test for yield

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	63
Error Mean Square	61.64019
Critical Value of Studentized Range	3.39458
Minimum Significant Difference	5.4402

Means with the same letter are not significantly different.

Tukey				
Grouping				
		Mean	N	moisture
	A	58.583	24	2
	A			
B	A	56.919	24	3
B				
B		52.255	24	1

- (a) Determine the missing degrees of freedom in the output.
 - (b) Conduct tests for significant interactions and, if appropriate, significant differences among the fertilizer means and moisture means. Give test statistics, p -values, and conclusions.
 - (c) What does the Tukey output suggest?
3. Both of the factors in the study discussed in the previous two problems are quantitative. It may be appropriate to analyze the data using multiple regression. Partial output from the fit of one model is provided below. The variable *msquared* is moisture squared.

Source	DF	Sum of Squares	Mean Squares	F	P
Model					
Error					
Corrected Total		5257.13017			
R-Square	0.2210				
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t

Intercept	1	30.60263	7.25943	4.22	<.0001
fert	1	0.73322	0.22403	3.27	0.0017
moisture	1	18.31663	7.84108	2.34	0.0224
msquared	1	-3.99622	1.94016	-2.06	0.0433

- Give the degrees of freedom for “Model”, “Error”, and “Corrected Total”.
 - Find the model and error sum of squares.
 - Compute the mean squares for model and error.
 - Compute the overall F -statistic, p -value, and a conclusion.
 - Is there a significant linear relationship between moisture squared and yield when accounting for the relationship between yield and the explanatory variables fertilizer and moisture? Give a test statistic, p -value, and conclusion.
 - Is there a significant linear relationship between moisture and yield when accounting for the relationship between yield and the explanatory variables fertilizer and moisture squared? Give a test statistic, p -value, and conclusion.
 - Is there a significant linear relationship between fertilizer and yield when accounting for the relationship between yield and the explanatory variables moisture and moisture squared? Give a test statistic, p -value, and conclusion.
 - Provide an estimate of the regression function.
 - Estimate the mean yield for tomatoes that receive 7 units of fertilizer and 2.5 inches of moisture.
 - Explain how the negative coefficient on moisture squared is consistent with the comparison of the moisture level means made with the Tukey procedure in the previous problem.
 - The simple linear regression of yield on fertilizer has a mean squared error of 65.88557. Use this information to test $H_0 : \beta_2 = \beta_3 = 0$, where β_2 and β_3 are the coefficients of *moisture* and *msquared* for the multiple regression of yield on fertilizer, moisture, and moisture squared.
 - Does the two-factor analysis of variance model fit the yield data significantly better than the multiple regression of yield on fertilizer, moisture, and moisture squared? Conduct an appropriate test.
4. Hogs were randomly divided into two groups of 20 hogs each. Group A was fed diet A while hogs in group B were fed diet B. Data on the weight gains for hogs in each group are provided below.

Group	Average Weight Gain	Standard Deviation of Weight Gains
A	135	29
B	146	38

- Is there a significant difference between the average weight gains associated with the two diets? Conduct an appropriate test assuming the data are “well behaved.”
- You could have used either a two-sample t -test or one-way ANOVA to answer part (a). Explain the relationship between the test statistics and p -values for the two approaches.
- Conduct a confidence interval for the difference in means and explain how it can be used to answer part (a). Also explain why it is nice to have both a p -value and a confidence interval.
- Suppose that the same table of summary statistics was obtained in a similar experiment that used 200 hogs in each group. Would there be a significant difference between the average weight gains associated with the two diets in this case?

- (e) Suppose the standard deviation for group B was 68 rather than 38. How might your data analysis strategy change?
- (f) How might you analyze the data if the standard deviations were as reported in the table but both groups had some extreme high and low outlying observations?
- (g) What sort of graphical displays might be useful for comparing the distributions of weight gains between the two groups?
5. Is it possible to change a slope coefficient from significantly positive to significantly negative by adding a single explanatory variable to a regression model? If possible, sketch a plot to indicated how this could happen.
6. We once saw an example indicating that males tend to claim to have higher maximum driving speeds than women. Would the correlation between height and claimed maximum driving speed be negative, zero, or positive?
7. What is the difference between an observational study and an experiment? Explain how the type of conclusion that can be drawn from data depends on whether the data came from an observational study or an experiment.
8. Suppose in a two-factor problem the population treatment means are as described in the table below.

	Factor B			
Factor A	level 1	level 2	level 3	level 4
level 1	17	12	11	29
level 2	8	3	2	20
level 3	19	14	13	31

- (a) Is there interaction?
- (b) How many numbers would need to be changed in order to change the correct answer to part (a)?
9. The following data is based on information taken from *Winter Wind Studies in Rocky Mountain National Park*, by D. E. Glidden (Rocky Mountain Nature Association). At five weather stations on Trail Ridge Road in Rocky Mountain National Park, the peak wind gusts (miles per hour) in January and April are recorded below.

	Weather Station				
Month	1	2	3	4	5
January	139	122	126	64	78
April	104	113	100	88	61

Does this information indicate that peak wind gusts are higher in January than in April?

10. Suppose an completely randomized experiment is conducted with two factors A and B . Suppose factor A has two levels and factor B has three levels. Due to experimental problems, the data (provided in the table below) are unbalanced.

	B1	B2	B3
A1	3,6,8,11	25,29	36, 42
A2	2	20,22,24	27, 28, 30, 35

- (a) Compute the lsmean for each level of factor A .
- (b) Compute the mean for each level of factor A .

- (c) Which level of factor *A* leads to higher values of the response variable according to the means?
- (d) Which level of factor *A* leads to higher values of the response variable according to the lsmeans?
- (e) Which gives the most appropriate information about the effect of factor *A* on the response, the means or the lsmeans? Explain.
11. An experiment was conducted to compare the effectiveness of three feed additives (A, B, and C) on weight gain in hogs. Three pens of hogs on each of 12 farms were used for the experiment. On each farm, the three additives were assigned in a completely randomized fashion to the three pens available for the experiment. Use the partially completed SAS output below to summarize the important results for the researchers. Note that two analyses of the data are contained in the code and output. One is correct and one is incorrect. Identify the correct analysis, and explain why the other analysis is incorrect.

```
proc glm;
  class farm trt;
  model wtgain=farm trt;
  lsmeans trt / pdiff adjust=tukey;
run;
```

The GLM Procedure

Class Level Information

Class	Levels	Values
farm	12	1 2 3 4 5 6 7 8 9 10 11 12
trt	3	1 2 3

Number of observations 36

Dependent Variable: wtgain

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	??	3453.706944	???????????	255.42	<.0001
Error	??	22.882778	???????????		
Corrected Total	??	3476.589722			

R-Square	Coeff Var	Root MSE	wtgain Mean
0.993418	1.846004	1.019866	55.24722

Source	DF	Type I SS	Mean Square	F Value	Pr > F
farm	??	3213.443056	292.131187	???????	<.0001
trt	??	240.263889	120.131944	???????	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
farm	??	3213.443056	292.131187	???????	<.0001
trt	??	240.263889	120.131944	???????	<.0001

Least Squares Means
 Adjustment for Multiple Comparisons: Tukey

trt	wtgain LSMEAN	LSMEAN Number
1	58.9000000	1
2	53.4833333	2
3	53.3583333	3

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

Dependent Variable: wtgain

i/j	1	2	3
1		<.0001	<.0001
2	<.0001		0.9516
3	<.0001	0.9516	

```
proc glm;
  class trt;
  model wtgain=trt;
  lsmeans trt / pdiff adjust=tukey;
run;
```

The GLM Procedure

Class Level Information

Class	Levels	Values
trt	3	1 2 3

Number of observations 36

Dependent Variable: wtgain

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	??	240.263889	???????????	?????????	0.3068
Error	??	3236.325833	???????????		
Corrected Total	??	3476.589722			

R-Square	Coeff Var	Root MSE	wtgain Mean
0.069109	17.92498	9.903054	55.24722

Source	DF	Type I SS	Mean Square	F Value	Pr > F
trt	??	240.2638889	120.1319444	????	0.3068

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	??	240.2638889	120.1319444	????	0.3068

Least Squares Means

Adjustment for Multiple Comparisons: Tukey

trt	wtgain LSMEAN	LSMEAN Number
1	58.9000000	1
2	53.4833333	2
3	53.3583333	3

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

Dependent Variable: wtgain

i/j	1	2	3
1		0.3838	0.3676
2	0.3838		0.9995
3	0.3676	0.9995	