Name Key
Please ALSO write your name on the back of the last page.

Ag 400 - Quiz 5
Makeup - Fall 2006

A. Let us suppose that you are interested in assessing regional differences in income levels in the U.S. You randomly select 10 counties in each of four regions of the country and collect information from each county on the income level of the residents. You compile the following data:

	Region			
	Northeast	South	Midwest	West
Number of counties sampled (N)	10	10	10	10
Mean income score ($\bar{X}$)	20	14	16	20
Sum of the scores (ΣX)	200	140	160	200

(20) 1. Test the statistical significance of the income differences among these four regions. Use the .05 level to determine statistical significance. Specify the following:

a. State the Null Hypotheses and the Alternative Hypothesis in words, not statistical symbols.

H_0: there is no relationship between region and income

H_A: there is a relationship between region + income

b. Set up an analysis of variance table (including sums of squares, degrees of freedom, mean squares and F) to carry out your test. (The "Total Sum of Squares" in the analysis of variance table equals 594.)

	SS	df	MS	F
Total	594	39		
Among	270	3	90	10.00
Within	324	36	9	

c. Indicate the degrees of freedom you would use for testing the significance of the calculated F.

3, 36

d. What is the critical value of F at the .05 level.

Between 2.84 and 2.92

e. Reject or do not reject the Null Hypothesis.

Reject H_0

f. What is the probability that you have made a Type I error in (e) above.

$\alpha < .001$

g. What is the probability that you have made a Type II error in (e) above?

0

h. Interpret the meaning of your findings precisely in terms of the problem.

Regions differ in regard to income.
South has lowest income, followed by Midwest.
Northeast and West have highest income

(20) 2. Indicate whether each of the following statement is true (T) or false (F) in terms of the above data. If any part of a statement is untrue, it should be marked false (F). Add comments, if you wish, to clarify your answers.

T a. The statistical unit in the above problem is a county.

F b. The interactive effect of region and income is statistically significant at the .05 level of significance.

F c. The Null Hypothesis for the significance test in #1 above can be stated as follows:

Ho: There is no difference between region and income in the U.S.

F d. The above F-test tests the statistical significance of the differences among the variances in incomes for the four regions.

F e. If the number of counties sampled in each of the four regions were doubled (i.e. 20 counties in each region), but the mean income scores remained the same for each county, Eta would increase.

T f. Eta squared, calculated for the above analysis measures the proportion of the total variation among the 40 counties that is associated with differences among the four defined regions.

F g. The above analysis could have been carried out using a quadratic correlation/regression analysis where X=region and Y=income.

F h. The "sum of squares within the region categories" in the Analysis of Variance Table for the above problem measures the variation in income scores that is accounted for by differences between the regions.

T i. The statistical significance of the difference between the mean income score for the Northeast and the mean income score for South could be tested using a Scheffe's Test.

T j. The mean income score for the Northeast Region is significantly greater than the mean income score for the South.

(8) B. The following table shows the results of additional analysis of the data presented in (A) above. In addition to "Region," the sampled counties have been classified as "within a metropolitan area" and "not in a metropolitan area."

Relationship of Region and Metro/Nonmetro Status to Income Scores of 40 Counties.

Metropolitan Status	Region			
	Northeast	South	Midwest	West
	----------Mean Income----------			
Metropolitan	24 (5)a	18 (5)	20 (5)	24 (5)
Nonmetropolitan	16 (5)	10 (5)	12 (5)	16 (5)

[a]Numbers in parentheses are the number of counties in the cell.

Indicate whether each of the following statements is true (T) or false (F) in terms of the above data.

F 1. This is an example of a randomized block design in which the counties are paired in regard to metropolitan status.

T 2. In the above table, there is no relationship between region and the metropolitan status of the county.

T 3. There is no interactive effect of region and metropolitan status of the county on Income in this sample.

T 4. For counties in metropolitan areas and those not in metropolitan areas, the relationship of region to Income Score can be described as follows:

The Northeast and West have the highest Income Scores and the South has the lowest Income Scores.

C. To ascertain the nature of consumer preferences in regard to three brands of cola soft drinks, an experiment was conducted in which 5 subjects were each asked to sample the three cola drinks and to rate the taste of each on a scale of 1 to 10, where 1 = terrible and 10 = excellent. You compile the following data. Assume that the reported figures have been correctly calculated.

Subjects	Cola Brand X	Cola Brand Y	Cola Brand Z
	----------Taste Ratings----------		
1	6	5	4
2	7	10	7
3	9	9	6
4	6	7	5
5	7	9	8
Total	35	40	30

Analysis of Variance Table

Sources of Variation	SS	df	MS	F
Total	42	14		
Among cola brands	10	2	5.0	5.00 sig .05
Among Raters	24	4	6.0	
Error	8	8	1.0	

(6) 1. Complete the above analysis to test the statistical significance of the differences among the three cola brands in regard to taste ratings. Use the .05 level to determine statistical significance.

There are sig diffs among these 3 cola brands in taste ratings. Brand Y had the highest rating; Brand Z had the lowest rating

(10) 2. Answer the following by indicating whether each statement is true (T) or false (F) in terms of the above data. If any part of a statement is untrue, it should be marked false (F).

F a. The above analysis is a factorial analysis of variance.

F b. The interactive effect of cola brand on taste ratings is statistically significant at the .05 level.

F c. The relationship between cola brand and taste rating is negative in the sample.

T d. If there were 10 raters instead of 5, the degrees of freedom for "Among Cola Brands" would be 2.

T e. The above analysis could have been done using a paired t-test if there were only <u>two</u> brands of cola being evaluated.

(36) D. The attached output reports the results of the following analysis for students enrolled in Stat 200 last semester.

Q1 = Performance on Quiz 1 measured in terms of the following categories.

Q1 Score	Q1 Code
<60	1
60-74	2
75 and over	3

FINAL = Actual Score on the final exam.

Answer the following questions in terms of this output by indicating whether each statement is true (T) or false (F). If any part of a statement is untrue, it should be marked false (F).

T 1. In the sample, students who received grades of less than 60 on Quiz 1 overall had lower final exam scores than did those who scored 60 points or more on Quiz 1.

F 2. Degrees of freedom for testing the statistical significance of the F-value of 3.397 are 2 and 131.

T 3. The F-value of 3.397 on the output is used to test the following Null Hypothesis:

H_o: In the population students with scores of less than 60 points on Quiz 1 have the same Final Exam score as those who scored between 60-74 points on QUIZ 1 and those who scored 75 points or more on Quiz 1.

T 4. If you reject the following Null Hypothesis based on this analysis, the probability that you have made a Type I error is less than .05.

H_o: Final exam scores are not related to Quiz 1 performance as measured by these three categories (<60; 60-74; 75 and over)

T 5. In the sample, the standard deviation in final exam scores is larger for students who scored <60 points on Quiz 1 than are the standard deviations in final scores for those who scored 75 points or more on Quiz 1. However, this difference is not statistically significant at the .05 level.

F 6. If you had access to the numerical scores for Quiz 1 as well as the numerical scores for the Final Exam for these students, you could have tested the statistical significance of the relationship between Quiz 1 Score and Final Exam Score using a paired t-test.

F 7. This analysis is a randomized block design since the same students' Quiz 1 scores and Final Exam Scores are used in the analysis.

F 8. Even if you had information on the numerical scores for Quiz 1 as well as the numerical scores for the Final Exams of every student in the sample, you should not use a linear correlation-regression to test the significance of the relationship between these variables because the relationship in the sample is curvilinear.

F 9. Eta for these data equals about .05. This means that approximately 5% of the total variation in FINAL is accounted for by the variation in Quiz 1 scores as measured by the three categories defined above.

T 10. Approximately 95% of the variation in Final Exam Scores is <u>not</u> associated with differences among the Quiz 1 categories of (<60; 60-74; and 75 or more).

T 11. The sum of the deviations of the 132 individual final exam scores about the grand mean of 75.48 equals zero.

F 12. On the basis of the analysis presented here, you know that the homogeneity of variance assumption of the analysis of variance is met by these data.

F 13. Levene's Test Shows that the variances in final exam scores for the three categories measuring Quiz 1 are not significantly homogeneous and hence we should question the validity of the homogeneity of variance assumption.

F 14. The "Post Hoc Test" reported on this output is the Scheffe's Test. Here Scheffe's Test is used to test the significance of the difference between Quiz 1 and Final Exam Scores for the various categories.

F 15. There is an interactive effect of Quiz 1 and Final in this data set, but this output does not test the statistical significance of this interaction.

T 16. In this analysis, final exam score, the dependent variable, is measured by an interval score. Quiz 1 score is treated as a nominal scale in this analysis.

T 17. The statistical unit here is a student enrolled in Stat 200.

F 18. Using Scheffe's Test, you can conclude that Quiz 1 scores in the first category differ significantly (.05 level) from Quiz 1 scores in the third category.

Oneway

Descriptives

FINAL

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	48	73.19	9.65	1.39	70.39	75.99	44	92
2.00	51	75.67	8.74	1.22	73.21	78.12	50	91
3.00	33	78.55	8.87	1.54	75.40	81.69	54	93
Total	132	75.48	9.28	.81	73.89	77.08	44	93

Test of Homogeneity of Variances

FINAL

Levene Statistic	df1	df2	Sig.
.261	2	129	.770

ANOVA

FINAL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	564.142	2	282.071	3.397	.037
Within Groups	10712.828	129	83.045		
Total	11276.970	131			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: FINAL
Scheffe

(I) Q1	(J) Q1	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	-2.48	1.833	.403	-7.02	2.06
	3.00	-5.36*	2.061	.037	-10.46	-.25
2.00	1.00	2.48	1.833	.403	-2.06	7.02
	3.00	-2.88	2.036	.371	-7.92	2.16
3.00	1.00	5.36*	2.061	.037	.25	10.46
	2.00	2.88	2.036	.371	-2.16	7.92

*. The mean difference is significant at the .05 level.

Homogeneous Subsets

FINAL

Scheffe[a,b]

Q1	N	Subset for alpha = .05	
		1	2
1.00	48	73.19	
2.00	51	75.67	75.67
3.00	33		78.55
Sig.		.458	.350

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 42.406.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.