Graduate Statistics **Fall, 1998**
Second Mid-term Exam

QUESTION A

As a result of having a large number of salary disputes, a large academic department decides to do an analysis of their salary data. The following variables are available:

SALARY:	Salary in $1000s
YEARS:	Years since receiving Ph.D.
PUBS:	Average number of publications per year
GRANT:	Average amount of grant dollars per year
GENDER:	Male or female
RANK:	Assistant, Associate, or Full Professor (Note: Associate and Full Professors have tenure, but Assistant Professors do not)

Use the variables to specify the model comparisons that would allow the department to ask the following questions. Be sure to specify any transformations of variables and to indicate how you are coding any categorical predictor variables. In each case, use the simplest model comparison that will answer the question. Also, indicate PA – PC (and note that some of the questions may require PA – PC > 1).

1. (3 pts.) Do men have higher salaries on average than women?

2. (3 pts.) Do tenured professors have higher salaries than non-tenured professors do?

3. (3 pts.) Some professors complain that a promotion in rank doesn't result in a permanent step increase in salary. That is, they contend that salary is essentially based on years since Ph.D. Over and above years since Ph.D., does higher rank predict higher salaries?

4. (5 pts.) In the PROC REG output for MODEL A from the previous question, the department chair notices that neither the regression coefficient for YEAR nor either of the RANK codes is statistically significant. The department chair therefore wonders if both YEAR and the RANK codes could be left out of the model of salary. Which model comparison would answer the chair's question? The department chair is confused after doing this comparison to find that both YEAR and the RANK codes cannot be left out. Explain the situation to the department chair in a couple of sentences.

5. (5 pts.) If we were able to compare professors who had the same number of years since graduation, the same average number of publications, the same average amount of grant dollars, and the same rank, would the male professors have a higher salary on average than the female professors? In addition to specifying the model comparisons, state in a sentence or two why this model comparison might have a different answer about a gender difference than the one in #1 above.

6. (3 pts.) The Dean wants to be competitive in hiring assistant professors. She knows that the average assistant professor at comparable institutions has a salary of about $43,000. Assuming that years since Ph.D. and rank are the only useful predictors of salary, specify the model comparison that would tell the Dean whether the average assistant professor's salary in this department differs from $43,000. (Note: the easiest answer may not involve using contrast codes.)

QUESTION B

The analysis on the following pages uses data taken from the 1995 U.S. News & World Report's Guide to America's Best Colleges. Most of the data are for the 1993-94 school year.

School officials at places like the University of Colorado are concerned about relatively low graduation rates and the disparity in such rates between private and public universities. To try and understand these graduation rates (and the public versus private university difference in them), they have done a series of regressions. In these GRADRATE (graduation rate, varying between 0 and 100) has been regressed on sets of the following predictor variables:

PUBPRI	Public university (+1) versus Private (-1)
NFULL	Number of Fulltime Students Enrolled
SFRATIO	Student/Faculty Ratio
SPENDPER	Total Expenditures/Student
COMBSAT	Average Combined SAT score of Entering Class
TOP10HS	Percent of Entering Class in Top 10% of HS Class

Based on these analyses, answer the following questions.

1. (3 pts.) What is the mean graduation rate at public universities? What is the mean graduation rate at private universities? What are the values of Pre and F* for testing whether these mean graduation rates differ?

2. (3 pts.) Does there remain a difference in graduation rates between public and private universities even when we control for NFULL, SFRATIO, SPENDPER, COMBSAT, and TOP10HS? What are the values of Pre, F*, PA-PC, and N-PA for answering this question?

3. (3 pts.) The variables PUBPRI, NFULL, SFRATIO, SPENDPER, COMBSAT, and TOP10HS can be thought of as measuring two different kinds of things. The first four measure characteristics of the school (whether it's public or private, its size, student faculty ratio, and expenditures per student), the last two measure the quality of the entering student body (SAT and high school standing).

A reasonable question is whether the student characteristic variables (COMBSTAT and TOP10HS) are useful predictors of GRADRATE even when we control for the characteristics of the school. Provide Pre, F*, PA-PC, and N-PA for this comparison in addition to your conclusion.

4. (3 pts.) Suppose we wanted to know whether the public and private universities attracted different sorts of students. One way to answer this question might be to predict PUBPRI using the student characteristic variables (COMBSTAT and TOPHS). What would be the value of Pre if we compared such a model to a model that simply estimated the mean of PUBPRI?

5. (3 pts.) If we did the model comparison suggested in question D, what assumption might inferential tests based on such a model comparison violate? Why?

Model: MODEL1
Dependent Variable: GRADRATE   Graduation Rate

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	28817.46945	28817.46945	100.894	0.0001
Error	626	178799.16112	285.62166		
C Total	627	207616.63057			

Root MSE	16.90035	R-square	0.1388
Dep Mean	63.92675	Adj R-sq	0.1374
C.V.	26.43705		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	61.070384	0.73189965	83.441	0.0001
PUBPRI	1	-7.351634	0.73189965	-10.045	0.0001

Model: MODEL2
Dependent Variable: GRADRATE   Graduation Rate

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	56344.43420	14086.10855	58.012	0.0001
Error	623	151272.19637	242.81251		
C Total	627	207616.63057			

Root MSE	15.58244	R-square	0.2714
Dep Mean	63.92675	Adj R-sq	0.2667
C.V.	24.37546		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	54.673729	2.93106526	18.653	0.0001
PUBPRI	1	-8.020792	0.90318888	-8.881	0.0001
NFULL	1	0.000671	0.00016484	4.074	0.0001
SFRATIO	1	-0.348312	0.14323376	-2.432	0.0153
SPENDPER	1	0.000875	0.00012297	7.112	0.0001

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	84485	.	.
PUBPRI	1	19149	0.11236330	0.55824628
NFULL	1	4029.266918	0.02594481	0.58769393
SFRATIO	1	1435.875821	0.00940275	0.74220517
SPENDPER	1	12280	0.07508473	0.78293788

Model: MODEL3
Dependent Variable: GRADRATE   Graduation Rate

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	88936.08772	29645.36257	155.870	0.0001
Error	624	118680.54285	190.19318		
C Total	627	207616.63057			

Root MSE	13.79105	R-square	0.4284
Dep Mean	63.92675	Adj R-sq	0.4256
C.V.	21.57321		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-10.573369	6.45051451	-1.639	0.1017
PUBPRI	1	-6.111834	0.60141405	-10.162	0.0001
COMBSAT	1	0.071282	0.00756746	9.420	0.0001
TOP10HS	1	0.067418	0.04842380	1.392	0.1643

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	511.014264	.	.
PUBPRI	1	19642	0.14200288	0.98618805
COMBSAT	1	16875	0.12449033	0.34894447
TOP10HS	1	368.665386	0.00309675	0.34938007

Model: MODEL4
Dependent Variable: GRADRATE   Graduation Rate

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	89640.41913	14940.06985	78.641	0.0001
Error	621	117976.21145	189.97780		
C Total	627	207616.63057			

Root MSE	13.78324	R-square	0.4318
Dep Mean	63.92675	Adj R-sq	0.4263
C.V.	21.56099		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-7.129076	7.00657554	-1.017	0.3093
PUBPRI	1	-6.347653	0.81133242	-7.824	0.0001
NFULL	1	0.000136	0.00015286	0.889	0.3744
SFRATIO	1	-0.211501	0.12715707	-1.663	0.0968
SPENDPER	1	-0.000141	0.00013629	-1.037	0.3003
COMBSAT	1	0.071644	0.00783890	9.140	0.0001
TOP10HS	1	0.067899	0.05117505	1.327	0.1851

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	196.678849	.	.
PUBPRI	1	11629	0.08972430	0.54127393
NFULL	1	150.127005	0.00127090	0.53467497
SFRATIO	1	525.590874	0.00443530	0.73682691
SPENDPER	1	204.141221	0.00172737	0.49874517
COMBSAT	1	15869	0.11856317	0.32482861
TOP10HS	1	334.436708	0.00282677	0.31246917

QUESTION C

A researcher is interested in whether or not memory can be passed biochemically from one organism to another. In order to study this phenomenon, she takes several planaria (a type of worm) and trains them to constrict in the presence of light. After the worms have been sufficiently trained, they are chopped up and fed to another group of test subject planaria. These test subject planaria are then known as **trained cannibals**. She compares the constriction rate of the trained cannibals group to two other groups who differed from the first in the diet they were fed. One test subject group was fed planaria that had not been trained to react to the light (**untrained cannibals** condition). The other group was fed a normal worm diet (**control** condition). After being fed the appropriate diet for a week, each test subject planaria was exposed to light and the number of constrictions it made during a 10 second period was counted. Following you will find the means, within-cell variances and sample sizes for each group.

	Trained Cannibals	Untrained Cannibals	Control Condition
cell means:	13.31	8.13	6.69
cell variances:	19.54	19.98	10.24
n per cell	16	16	16

The researcher does a one way analysis of variance to ask whether there are any differences among the cell means. She correctly computes an F* statistic, with 2 and 45 degrees of freedom, of 11.69. The Mean Square Error from her model A (also the denominator of her F* statistic) equals 9.38.

She now wants to test two single degree of freedom contrasts among her cell means. First, she wants to know whether the mean in the trained cannibals condition is different from the means in the other two conditions. Second, she wants to know if the untrained cannibals mean differs from the control condition mean.

1. (3 pts.) What contrast coded predictors would you use to conduct these two further tests?

2. (3 pts.) What are the parameter estimates for your model A?

3. (4 pts.) Conduct the two tests. For each one, report the Pre, F*, N-PA, PA-PC, and a substantive conclusion.