METHODS QUALIFYING EXAM

JANUARY 2005

INSTRUCTIONS:

1. DO NOT put your NAME on the exam. Place the NUMBER assigned to you on the UPPER LEFT HAND CORNER of EACH PAGE of your exam.
2. Please start your answer to EACH QUESTION on a SEPARATE sheet of paper.
3. Answer all the questions.
4. Be sure to attempt all parts of every question. It may be possible to answer a later part of a question without having solved the earlier parts.
5. Be sure to hand in all of your exam. No additional material will be accepted once the exam has ended and you have left the exam room.

PROBLEM #1:

Three different experimental situations are described below. Provide the information requested for each.

A. A study is conducted to examine the effects of three different room temperatures and absence or presence of background noise on the performance of students taking an exam. A total of 48 students are available for the study, with the midterm exam score available for each student. Eight students are to be assigned to each of the six temperature by background noise combinations.

   (i) How would you suggest the students for each temperature by background noise combination be chosen?

   (ii) The score on the exam, taken under the experimental conditions, is recorded for each student. Give the sources and degrees of freedom for the appropriate ANOVA table. Where appropriate, indicate the numerator and denominator mean squares for testing significance of an effect.

   (iii) The exam consists of a multiple choice part, a short answer part, a "work a problem" part, and an essay question. Instead of a single exam score for each student, the scores on the individual parts of the exam are recorded separately. Does this change your ANOVA table? If so, provided the correct ANOVA table and F-ratios.

B. A research assistant to the president of a university collected data for a random sample on n = 100 full time tenure track faculty members. The variables, and their values, are:

SALARY = nine month salary equivalent (in Dollars)

AGE = age, in years

RANK = Academic Rank = 1 for Professor
2 for Associate Professor
3 for Assistant Professor
4 for Instructor, Lecturer, etc.

TENURE = Tenure Status = 0 if not tenured and not tenure track
1 if not tenured, but tenure track
2 if tenured

DEGREE = Final Degree = 1 if bachelors
2 if masters
3 if postmaster, but not doctorate
4 if doctorate

SEX = sex of faculty member = 0 if male
1 if female

TIME = Length of time (in years) since initial appointment to faculty at the university

The president has asked her assistant to determine if there is evidence of discrimination in salaries based on the sex of the faculty member. The research assistant has asked your help.

(i) If you were to fit a model using SALARY as the response variable (the "Y"), which of the other variables would you consider to be class variables (i.e., ANOVA type variables) and which of the variables would you consider to be covariates (i.e., regression type variables)?

(ii) The president asks if there is evidence the change in average salary for each additional year after initial appointment is not the same for males and females. How would you determine if there is evidence of this, i.e., what term or terms would you add to the model?

(iii) Generally, faculty members at the rank of Associate Professor and Professor have tenure, faculty members at the rank of Assistant Professor do not have tenure but are tenure track, and faculty members at the rank of Instructor, Lecturer, etc., do not have tenure and are not tenure track. Could this create any difficulties in fitting a model with all the variables mentioned above included? If so, what is the problem?

C. An analysis of covariance was used to analyze data from an experiment on ten treatments in a Randomized Complete Block Design with five blocks and a covariate, X. Various models were fit to the data. The Error sums of squares (Residual sums of squares) for different models are shown below. The terms in the model are indicated by M = overall mean, BLK = class variable or dummy variables for blocks, TRT = class variable or dummy variables for treatments, and X for the covariate.

Model Includes	SS Error	Error DF
M	393.500	49
M, BLK	332.300	45
M, TRT	321.200	40
M. BLK, TRT	259.000	36
M, BLK, TRT, X	123.120	35
M, BLK, TRT, X(TRT)	82.226	26
M, BLK, TRT, X, X*TRT	82.226	26
M, BLK, X	197.599	44
M, TRT, X	139.928	39
M, X	214.637	48

(i) Give the value of the F-ratio to test if the slopes for the ten lines are equal.

(ii) Assuming a common slope, give the value of the F-ratio to test that the adjusted treatment means are equal.

PROBLEM #2: A programmer claims that $Y_1, \ldots, Y_{50}$, is a random sample of size 50 from a distribution having cdf given by

$$F(y) = 1 - exp(-(y-\theta)/\beta) \quad if \quad y \geq \theta,$$

where θ and β are **known** constants.

(A) Describe a graphical method to evaluate the programmer's claim. Be sure to label your axes.

(B) Describe a test of hypotheses to evaluate the programmer's claim.

(C) Suppose θ and β are **unknown** constants. Describe a graphical method to yield rough estimates of θ and β?

(D) Explain why the Chi-squared Goodness-of-Fit would not be a good choice for your test of hypotheses in part (B)?

(E) Under what circumstances, is the Chi-squared Goodness-of-Fit a good choice for a test statistic for testing the fit of a specified model?

PROBLEM #3:

These are regression problems. Assume all of the usual assumptions are met unless told otherwise. Please give complete answers to all questions.

A. Your client comes to you with the following output and claims: "I am very happy that vehicle weight is more significant than horsepower since the sig value (p value) for weight is less than the sig value for horsepower."

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.822[a]	.675	.674	4.459

a. Predictors: (Constant), Vehicle Weight (lbs.), Horsepower

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16085.855	2	8042.928	404.583	.000[a]
	Residual	7733.138	389	19.880		
	Total	23818.993	391			

a. Predictors: (Constant), Vehicle Weight (lbs.), Horsepower

b. Dependent Variable: Miles per Gallon

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	44.777	.825		54	.000
	Horsepower	-.061	.011	-.299	-5	.00000016227
	Vehicle Weight (lbs.)	-.005	.001	-.551	-10	.00000000000

a. Dependent Variable: Miles per Gallon

In responding to her comment, provide in detail

1) the usual assumption for a regression analysis
2) the model being used.
3) the hypotheses being tested by Horsepower and VehicleWeight.
4) the reasons why, if you agree with her statements; or give the reasons why you disagree with her statements.

B. Another client comes with the following output and asks: "What is my prediction equation?

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	396.985	3	132.328	21.516	.000[a]
	Residual	98.405	16	6.150		
	Total	495.390	19			

a. Predictors: (Constant), x3, x2, x1

b. Dependent Variable: y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	117.085	99.782		1.173	.258
	x1	4.334	3.016	4.264	1.437	.170
	x2	-2.857	2.582	-2.929	-1.106	.285
	x3	-2.186	1.595	-1.561	-1.370	.190

a. Dependent Variable: y

In answering his question, give:

1) the model used
2) the hypotheses being tested by x1, x2, and x3
3) the prediction equation
4) any other comments you might want to make to clarify what you are providing the client