Ph.D. Qualifying Exam of Statistical Methods, September 2008

1. The following is the result of ANOVA table of a 2^3 factorial design:

Source of variation	Sum of square	df	Mean square	F-value	P-value
A	36	1	36	57.60	<0.001
B	20.25	1	20.25	32.40	<0.001
C	12.25	1	12.25	19.60	0.002
AB	2.25	1	2.25	x	0.094
AC	0.25	1	0.25	0.40	0.545
BC	1	1	1	1.60	0.242
ABC	1	1	1	1.60	y
Error	5	z			
Total	78	15			

(i) How many replications does it has? (2%)

(ii) What are the values of x, y, z, respectively? (3%)

(iii) Fit an appropriate regression model using the factors which showed significant under 0.10 nominal. (Note: You should give the corresponding values of the parameters estimate. It can be done through the results in the above table.)(10%)

2. Assume that X and ε are two independent random variables which have finite second moment. However, we can only observe (X, Y) in which $Y = \beta_0 + \beta_1 X + \varepsilon$ and both β_0 and β_1 are unknown. Consider the least square estimation to estimate β_1 based on the random sample (X_i, Y_i) $i = 1, \ldots, n$. Show that the estimator is an unbiased and consistent estimator of β_1. Can we estimate β_0 without any external information? (Note: You should not assume $E(\varepsilon) = 0$.) (15%)

3. Consider a 16-run 2^{9-5} fractional factorial experiment in nine factors with generators $E = BD, F = BCD, G = AC$, $H = ACD$, and $J = AB$.

(i) Find the defining relation and the alias relationships in this design. (5%)

(ii) What is the resolution of this design? (5%)

(iii) What is the resulting design if it projects into four factors which is not a word in the defining relation? (5%)

4. Assume that the correct linear regression model is as following

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

where X_1 is a full rank $n \times p$ matrix, X_2 is an $n \times r$ matrix, and β_j $(j = 1, 2)$ are the corresponding vector parameters. Let $\hat{\beta}_1$ be the least square estimator of β_1 by fitting the regression model

$$Y = X_1\beta_1 + \varepsilon.$$

(i) What is the expectation of $\hat{\beta}_1$? (10%)

(ii) Give the condition in which $\hat{\beta}_1$ is an unbiased estimated of β_1. (5%)

5. Assume that $Y^*|X\tilde{\ }Poisson(\ln(X^\beta+1))$ and $X\tilde{\ }f(.)$ which takes value greater than 1. Let $Y = I(Y^* > 0)$.

(i) Describe the corresponding logistic regression model in terms of (X, Y). (5%)

(ii) Write down its corresponding likelihood function based on (X_i, Y_i) $(i = 1, \ldots, n)$. (5%)

(iii) Suppose that Y^* is observable, give the advantages and disadvantages of using (X, Y) to estimate β instead of (X, Y^*). (5%)

6. Explain thoroughly the following items:

(i) Forward selection, backward elimination and stepwise selection. (6%)

(ii) Principal components analysis. (4%)

(iii) Nonlinear regression. (4%)

(iv) Response surface methodology and center composite design. (6%)

(v) ANOVA and ANCOVA. (5%)