Name ______________________________ Spring, 2007

Applied Statistics Comprehensive Examination

Regression Methods & Linear Models

1. (20 pts) Consider the simple linear regression model, $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where β_0 is a known constant and $\epsilon \sim IIDN(0, \sigma^2)$.

   (a) Derive the least squares estimator for β_1.

   (b) Determine if this estimator is unbiased.

2. (30 pts) The following tables contain summary information taken from an intermediate step using one of the "stepwise" variable selection methods. Note that in the previous step, variable X_3 was added to the model.

Variables Currently in the Model

Variable	Estimate	F Statistic	P–value
Intercept	71.65		
X_1	1.45	154.01	0.0001
X_2	-0.24	1.86	0.2054
X_3	0.42	5.03	0.0087

Variables Not Currently in the Model

Variable	Estimate	F Statistic	P–value
X_4	1.25	3.75	0.0325
X_5	9.43	4.89	0.0114

(a) Suppose that a Forward Selection Method is being used to choose variables for this model. Describe what will happen in the next step.

(b) Suppose that a Stepwise Selection Method is being used to choose variables for this model. Describe what will happen in the next step.

(c) Based on the information available to you in the two tables, state whether you think multicollinearity is a problem in this dataset. Justify your answer.

3. (30 pts) Suppose an experiment was run in a completely randomized manner and the following data were obtained:

Treatment

A	B	C	D
7	8	2	5
3	2	9	4
3	3	5	
	1		

(a) Write the normal equations assuming the usual fixed effects model, $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ where $\epsilon \sim IIDN(0, \sigma^2)$.

(b) Briefly describe why restrictions are often used to obtain a solution to the normal equations. Name another method which may be used to solve the normal equations.

(c) Determine if $\tau_A - \tau_D$ is estimable and justify your answer.

4. (20 pts) Suppose a (2×3) factorial design is run in a completely randomized manner and results in the data which appear below. The data will be analyzed assuming a model of the form: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$ where α_i is a fixed effect corresponding to rows, β_j is a fixed effect corresponding to columns, γ_{ij} is a fixed effect corresponding to the interaction between rows and columns and ϵ_{ijk} is random error term which is assumed to be $IIDN(0, \sigma^2)$. Note that $i = 1, 2$, $j = 1, 2, 3$ and $k = 1, 2, \ldots n_{ij}$.

6,7	8,8,9	9,11
3,3	5	8,9,10

(a) Calculate the lsmeans for both rows and all three columns.

(b) Write any two degree of freedom contrast in interaction effects. State whether or not these two contrasts are orthogonal and justify your answer.