Name _____ Spring, 2008

Applied Statistics Comprehensive Examination
Regression Methods & Linear Models

Calculators are permitted on this part of the examination.

1. (10 points) Consider the typical linear regression situation where $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ and $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Derive the least squares estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$. Show your work.

2. (20 points – 5 each part) The following are assumptions of a regression model:

   a. The predictor variable and response variable are linearly related (in simple linear regression)
   b. The errors have constant variance
   c. The errors follow a normal distribution
   d. There is no multicollinearity among the predictor variables (in multiple regression)

   For each of the four assumptions above, draw a plot that shows a violation of the assumption. Make sure to label both the axes.

3. (20 points) In the article *Is There a "Gender Affinity Effect" in American Politics? Information, Affect, and Candidate Sex in U.S. House Elections* (by Dolan, Political Research Quarterly, 2008), the author examined elections in the U.S. House of Representatives from 1990-2000 to determine how the gender of the candidate affects views and voting.

   Information is a score from 0-10 that measures how much people know about the candidate and is used as the dependent variable in one of their models. The following table shows the parameter estimates (and standard errors in parentheses) for candidate gender, voter gender, and the interaction of the two for Democratic candidates only (controlling for other covariates as well, not listed).

	Parameter Estimate (Standard Error)
Female Candidate	0.204 (.081)
Female Voter	-0.183 (.042)
Voter & Candidate both Female	0.279 (.103)
n	6,056

   a. Conduct the formal hypothesis test to determine whether there is an interaction effect between *female candidate* and *female voter*.
   b. How much more (or less) *information* does a female voter have for a Democratic candidate relative to a male voter?

4. (25 points) Consider a linear mixed model that involves a three-way treatment structure. Factors *A* and *B* are fixed, while factor *C* is random. Suppose that there are two levels of each factor and three observations on each combination of factor levels. Assume that the only terms in the model are those listed in the ANOVA table given below.

Source	SS	df	MS	Expected MS
A	215.0			$\sigma^2+6\sigma^2_{AC}+Q(A)$
B	85.6			$\sigma^2+6\sigma^2_{BC}+Q(B)$
C	51.4			$\sigma^2+6\sigma^2_{AC}+6\sigma^2_{BC}+12\sigma^2_{C}$
AC	16.5			$\sigma^2+6\sigma^2_{AC}$
BC	9.4			$\sigma^2+6\sigma^2_{BC}$
Error	123.5			σ^2

   a. (15 points) Using the method of moments, find point estimates for all four variance components in the model.

   b. (10 points) Find a 95% confidence interval for the error variance σ^2.

5. (25 points) A company was interested in comparing two different panels for use by air traffic controllers. Each panel was examined under four different emergency conditions. The experiment was run as a completely randomized design, with two professional air traffic controllers assigned to each combination of factor levels. The cell means are given in the table below.

		Emergency Condition			
		1	2	3	4
Display Panel	1	17	25	24	26
	2	14	31	22	29

   a. (15 points) Create an interaction plot, and discuss the presence/absence of interaction.

   b. (10 points) Assume an effects model with interaction. Let α_i be the effect of the i^{th} display panel, β_j be the effect of the j^{th} emergency condition, and γ_{ij} be the interaction effect associated with the i^{th} display panel and the j^{th} emergency condition. Consider the following two linear combinations of parameters. Identify each linear combination as estimable or as not estimable, and justify your answer. (i) $\alpha_1- \alpha_2+ \beta_1- \beta_2$ (ii) $\alpha_1+ \gamma_{13}- \alpha_2- \gamma_{23}$