Name Spring, 2009

Applied Statistics Comprehensive Examination
Regression & Linear Models

1. (20 points) Recall that in regression through the origin, the linear regression equation is $Y_i = \beta X_i + \epsilon_i$ for $i = 1, \ldots, n$ and we assume that the X's are fixed and constant and that $\epsilon_i \sim \text{iidN}(0, \sigma^2)$. In the middle of an exam, your mind blanks, and you cannot recall the regression estimator for the slope. However, you know that two points define a line, so instead of the typical estimate for the slope you connect the origin, $(0, 0)$, with the first observation, (x_1, y_1). The slope is defined as the change in y over the change in x, so your estimator of β becomes

$$\hat{\beta}^* = \frac{y_1 - 0}{x_1 - 0} = \frac{y}{x}$$

   (a) (5 Points) Show that $\hat{\beta}^*$ is an unbiased estimator of β.

   (b) (5 Points) Derive the variance of $\hat{\beta}^*$.

   (c) (10 Points) Which is a better estimator, $\hat{\beta}^*$ or $\hat{\beta}$? Explain why. (Recall that $\hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$ for the true regression through the origin equation.)

2. (30 points) An analysis on predicting success in an introductory psychology class was done on 471 students at the University of Wisconsin Madison. The following table presents the model they used to predict final grade:

Variable	**Slope**
Performance-based Goals (Higher = Performance is important)	0.16
Work Avoidance Goals (Higher = Tries to avoid doing work)	-0.15
Ability (SAT or ACT Score)	0.33
High School Percentile	0.33
Instructor (coded as +1 or -1 for the two instructors)	0.09
Gender (coded as 1 for women or 0 for men)	0.18

   (a) (10 Points) The F-statistic for the multiple regression was 26.5. Conduct the appropriate hypothesis test for this statistic.

   (b) (10 Points) The standard error for the slope for the ability variable is .04. Test the appropriate hypothesis to determine whether ability is related to final grade, controlling for performance-based goals, work avoidance goals, high school percentile, instructor, and gender.

   (c) (10 Points) Typically, dichotomous variables are coded as 0 and 1. Instructor was coded as +1 and -1.

      i. What impact does this choice have on the slope estimate?

      ii. What impact does this choice have on the p-value for testing whether this slope is 0?

3. (30 Points) The following table shows the results for a 2x3 factorial design. The table displays the mean of each cell and the sample size in parentheses.

	Treatment 2		
Treatment 1	1	2	3
1	9 (3)	14 (2)	18 (2)
2	16 (4)	31 (2)	13 (3)

The mean squared error is 11.

(a) (10 Points) Construct a set of two orthogonal interaction contrasts for this design.

(b) (10 Points) Test the significance of the interaction $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}$

(c) (10 Points) Using an effects model with interaction and using only one observation per cell, construct the X matrix using set-to-zero restrictions.

4. (20 Points) Consider the additive model $Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$. The data are presented in the following table:

	Treatment 2	
Treatment 1	1	2
1	3	11, 17
2	2, 6	14, 15, 16

(a) (10 Points) Calculate the means and least squares means for both levels of Treatment 1. Which set of means is appropriate for this problem. Explain.

(b) (10 Points) Show that all cell means are estimable using this effects model.