625.661 Statistical Models and Regression

Test 2 for Modules 5 & 6

H.M. James Hung

1. Prove that in using a regression model analysis to compare the differences of the expected values of the response variable y between the K levels of a categorical regressor x , the sum of squares, SS_T, SS_R, SS_{Res}, will not change regardless of how the K-1 indicators of x are coded (Recall any dummy variable D can be coded in many ways, e.g., D = 0, 1, or, D = -1, 1, or others). [10 points]

State assumptions in each step of your proof.

2. In a study, there are four treatments (labeled as 1, 2, 3, 4) to compare. Assume that there are m subjects per treatment.

(a) Construct an analysis of variance model to compare the four treatments; that is, test whether there is at least one pair of treatments that differ and construct an estimator of every pair of expected treatment difference. [20 points]

(b) Construct a linear regression model such that the regression analysis is equivalent to the analysis of variance in (a). [20 points]

State assumptions in each step of your proof or derivation in a) and b).

Do not use any math/stat software for calculation for Problem 3, except for obtaining the percentile of standard normal, t, chi-square, or F distribution, matrix operations, or basic mathematical calculations.

3. Ten observations on the response variable y associated with two regressor variables x_1 and x_2 are given in the following table. The model fitted to these observations is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma x_{1i} x_{2i} + \varepsilon_i \ , \quad i = 1, \ldots, n \ , \qquad (1)$$

where ε's are identically and independently distributed as a normal random variable with mean zero and a known variance $\sigma^2 = 4$.

Observation #	y	x_1	x_2
1	7	9	1
2	8	6	1
3	5	10	1
4	4	8	1
5	2	5	1
6	10	7	-1
7	9	6	-1
8	10	5	-1
9	8	5	-1
10	8	4	-1

a) Test the null hypothesis "there is no difference between the y-intercept for $x_2 = 1$ and the y-intercept for $x_2 = -1$ and there is no difference between the slope for $x_2 = 1$ and the slope for $x_2 = -1$" at a statistical significance level of 0.05. [20 pts]

b) Estimate the difference, $E(y \mid x_1 = 5, x_2 = 1) - E(y \mid x_1 = 5, x_2 = -1)$, and calculate its 95% confidence interval. [10 pts]

c) Predict the difference in y value at $x_1 = 5$ between $x_2 = 1$ and $x_2 = -1$. [10 pts]

d) Now fit Model (2): $y_i = \beta_0 + \beta_2 x_{2i} + \varepsilon_i$ to the 10 observations.
Calculate the residual for the observation #8 and its variance. [10 pts].

State assumptions in your derivations and calculations in a), b), c), d).