625.661 Statistical Models and Regression

Test 3 for Modules 7 - 10

H.M. James Hung

1. Consider the multiple regression model for n y-data $y_1, \ldots, y_n$,

$$y = X_1 B_1 + X_2 B_2 + \varepsilon$$

where $y = (y_1, \ldots, y_n)'$, $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2 V$, X_1 and X_2 are fixed, and each element of the vector y is normally distributed. B_1 and B_2 are vectors of regression coefficients (i.e., B_1 has at least two regression coefficients, so does B_2) under estimation. Assume that σ^2 and V are known. Construct a test statistic for the hypotheses

H_0: $B_1 = 0$ versus H_1: $B_1 \neq 0$.

1) Derive the best linear unbiased estimator for the regression coefficients vector $B = [B_1', \ B_2']'$ and prove that your estimator is indeed the best linear unbiased estimator. [15 points]
2) Derive the test statistic under the null hypothesis, statistical distribution of the test and its rejection region for H_0. [20 points]

2. Can **any** linear regression model be checked for model adequacy by statistical testing for lack of fit or goodness of fit? Why or why not? Please provide your answer with detailed justification (i.e., by mathematical proof or by a numerical example). [10 points]

3. Prove that in selection of p regressors, minimizing $MS_{res}(p)$ will lead to maximizing the adjusted $R^2(p)$. [10 points]

4. Critique the following statement: "For selection of the regressors to include in a linear regression model, the best strategy is always finding the model with the largest possible value of R^2 ". [10 points]

5. There are multiple methods or criteria for selection of regressors (e.g., forward selection, all possible regressions, R^2, C_p).

1) Describe all the methods we have learned in this course. [10 points]

2) Is there the best method or criterion? [5 points]

3) Create a dataset (for a model with response variable y , two regressors x_1 and x_2, and its interaction x_1x_2) to illustrate how to apply each method or criterion to selection of regressors. Provide details about how you get the dataset. Do not use any data in the Textbook. [20 points]