625.661 Statistical Models and Regression

Test 1 for Modules 1, 2, 3, 4

H.M. James Hung

1. Assume that a linear regression model has three regressors.
   a) Derive a test statistic to test the null hypothesis that their respective regression coefficients are all zero. [5 points]
   b) If the null hypothesis in a) is rejected, would this mean that the regression model has value for prediction? [5 points]
   c) Can the importance of each regressor be assessed by simply looking at the respective magnitudes of the t-statistics? [5 points]

State the assumptions for each step of your discussion or derivation in a), b), c).

2. In a simple linear regression analysis where the regressor x is non-random, we have random errors and residuals.

   a) What are the differences between residuals and random errors? [10 points]

   b) Derive the variance of residual and the variance of random error. [15 points]

   c) Now assume x is random. Derive the variance of residual. [15 points]

State the assumptions for each step of your discussion or derivation in a), b), c).

3. In a multiple linear regression analysis, the response variable y is studied with two non-random variables x_1 and x_2. This regression model that is fitted to the data on (y, x_1, x_2) of n subjects is given by

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \varepsilon \quad ,$$

   where $z_1 = w_1 x_1 + (1 - w_1)x_2$, $z_2 = w_2 x_1 + (1 - w_2)x_2$, ε is a random error with mean zero and variance σ^2 (its value is unknown), and the weights w_1 and w_2 have known values.

a) Derive the ordinary least-squares estimators for $(\beta_0, \beta_1, \beta_2)'$ as functions of (y, x_1, x_2), not (y, z_1, z_2) . [15 points]
b) Derive a statistical test for testing statistical significance of the regression model for (y, x_1, x_2), not for (y, z_1, z_2) . [15 points]

State the assumptions for each step of your discussion or derivation in a), b), c).

4. Suppose that n subjects give data following the **true** model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

where x_1 and x_2 are non-random regressors, and ε is the random error.

Now suppose that the reduced model

$$y = \beta_0 + \beta_1 x_1 + \varepsilon,$$

is also fitted to the same data to obtain the least-squares estimator $\tilde{\beta}_1$.

Discuss by mathematical arguments whether this least-squares estimator $\tilde{\beta}_1$ is biased for β_1 . If yes, discuss by mathematical arguments the conditions under which $\tilde{\beta}_1$ is unbiased for β_1 . [15 points]

State the assumptions for each step of your discussion or derivation.