BST169: Course Work Project

Students should work individually to study the following questions. Every student should submit an electronic copy of the report and the R script file(s) via Learning Central.

The project report should contain no more than 2,000 words (not including titles, appendices, formulas, tables and/or graphs). All pages, equations, tables and figures should be numbered. Tables and figures should be given proper captions and comments of explanation when necessary.

1. Consider the model,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i. \tag{1}$$

   What is the requirement for e_i such that the following test statstics will be valid to test $H_0 : \beta_1 + beta_2 = 1$?

   - $W = N\frac{SSR_R - SSR_U}{SSR_U}$ (Wald),
   - $LM = N\frac{SSR_R - SSR_U}{SSR_R}$ (Lagrange Multiplier),
   - $LR = N \ln \frac{SSR_R}{SSR_U}$ (Likelihood Ratio)

   where SSR_R is the sum of squared residuals obtained from the restricted model, while SSR_U is from the unrestricted model.

2. For the data set `pbp.csv`, can we use the three test statistics mentioned in the previous question to test $H_0 : \beta_1 + \beta_2 = 1$? Why? If W and LM are not valid, how can one modify them for the test? What is your conclusion from the valid test?

3. Generate y_i from the following model,

$$y_i = \beta_0 + \beta_1 x_{1i} + (1 - \beta_1) x_{2i} + \sqrt{x_{1i}}\epsilon_i. \tag{2}$$

   where x_{1i} follows chi-squared distribution with 2 degrees of freedom. Generate ϵ_i from student t distribution with 6 degrees of freedom and $x_{2i} \sim U(0, 10)$. Check whether W, LM and LR in Question 1 follow chi-squared distribution by Monte Carlo.(The R command `ks.test( ,'pchisq',2)` can be used.) If W and LM are not valid, calculate the correct test statistics and also verify them by Monte Carlo. Please consider different sample sizes.

4. Compare the size of different test statistics (frequencies of making Type 1 error) from Monte Carlo using 5% level of significance for different sample sizes. Explain the results.

5. For the data set `pbp.csv`, suppose Equation (2) is the true model. Use proper bootstrapped errors from the true model to study whether different test statistics for $H_0 : \beta_1 + \beta_2 = 1$ in the previous questions follow chi-squared distribution. Explain your results.