Exercise Sheet 4

This exercise sheet is to be done prior to your actual Support and Feedback class. The Stata do file/R script file, which you write as a group, should then be submitted by one of your group members via e-mail to your class tutor by Monday 10am of Week 6 (Term 1). You MUST take the output generated from your Stata/R programme to your tutorial classes. The data file is **march_cps_1976.dta/RData**, which is based on the replication files of "Hours Worked and the U.S. Distribution of Real Annual Earnings 1976-2019" Fernández-Val et.al., Journal of Applied Econometrics, 2024.

The dataset consists of a sample of U.S. female workers in 1976. For each worker, we observe their age (*age*), whether the worker has a child younger than 5 years old (*chlt5*), the number of weeks worked last year (*wkswork*), total earnings last year in thousands USD (*eanings*), and the region in which the worker lives (*region*; one of Northeast, Midwest, South, and West).

1. Consider the regression equation

$$wkswork = \alpha + \beta_1 age + \beta_2 chlt5 + \varepsilon.$$

   (a) Estimate this regression. In Stata, use the factor option for the *chlt5* variable.

   For your understanding: what is the interpretation of the coefficient on the *chlt5* dummy?

   (b) To the above regression, add the categorical variable for the region and report the estimation result. Use South as the baseline.

   (c) Test the hypothesis that the coefficients on "Northeast" and "West" variables are jointly zero at 5% significance level.

   For your understanding: what do you conclude from this test?

   (d) Re-estimate the equation in (b) using "West" as the default region.

   For your understanding: compare the coefficient estimate on the "Northeast" in this equation to that reported in the equation estimated in (b).

   (e) For the regression in (b), test the hypothesis that the coefficient on "Midwest" is zero. For the regression in (d), test the hypothesis that the coefficient on "Midwest" is zero.

   (f) Estimate the model

$$wkswork = \alpha + \beta_1 age + \beta_2 chlt5 + \beta_3 age \times chlt5 + \varepsilon$$

   without creating the new variable $age \times chlt5$. In Stata, use factor and c. option. In R, use "$*$" or ":".

   (g) Using the regression in (f), test the null hypothesis $H_0 : \beta_3 = 0$.

   (h) Estimate the regression model

$$\begin{aligned} wkswork = \alpha &+ \beta_1 chlt5 + \beta_2 Northeast + \beta_3 Midwest + \beta_4 South \\ &+ \gamma_1 chlt5 \times Northeast + \gamma_2 chlt5 \times Midwest + \gamma_3 chlt5 \times South + \varepsilon. \end{aligned}$$

NOTE:
Before you e-mail your Stata/R file to your class tutor:

(i) Check the file runs all the way through from start to end without intervention from the marker.

(ii) Annotate your Stata/R file with the appropriate question number to which your answer relates to help the marker understand what part of the file is answering which question.

(iii) Ensure that your Stata/R file only contains the commands necessary to answer the questions from the exercise sheet (in the Stata/R file, do not include commands related to earlier exercise sheets, which are not needed for this exercise sheet).

All exercise sheets will be marked out of 3 points. Failure to follow the instructions noted above will involve you losing marks.