

Programming Assignment # 3.

Please complete the following problems and submit a file named `PA3.R`.

Remember:

- Do not rename external data files or edit them in any way. In other words, don't modify `wage.csv`. Your code won't work properly on my version of that data set, if you do.
- Do not use global paths in your script. Instead, use `setwd()` interactively in the console, but do not forget to remove or comment out this part of the code before you submit. The directory structure of your machine is not the same as the one on Gradescope's virtual machines.
- Do not destroy or overwrite any variables in your program. I check them only after I have run your entire program from start to finish.
- Check to make sure you do not have any syntax errors. Code that doesn't run will get a very bad grade.
- Do not install or require any libraries unless specified in the task.
- Make sure to name your submission `PA3.R`

Tip: before submitting, it might help to clear all the objects from your workspace, and then `source` your file before you submit it. This will often uncover bugs. Before you start, download the file `wage.csv` from the Blackboard and save it in the same folder as the R script `PA3.R`. Make sure you set a correct working directory.

Consider the dataset `wage.csv` from the Blackboard. Download it and read the data first.

1. Import the data stored in 'wage.csv' as a data frame with name `Data`. The data frame contains the cross-sectional data on hourly wages in USD and years of education of 526 individuals.
2. Create new variables: the dependent variable of interest `y = Data$wage` - hourly wage of US workers, `x2 = Data$educ` - years of schooling, `x3 = Data$exper` - years of labour market experience, `x4 = Data$female` - a dummy variable which equals 1 if the individual is a female and equals 0 otherwise.

Verify that the relationship $1 - R^2 = (1 - r_2^2)(1 - r_{y,3|2}^2)$ holds. For that proceed as follows:

3. Estimate the following model specification:

$$y_i = \delta_1 + \delta_2 x_{2,i} + \delta_3 x_{3,i} + u_i, \quad i = 1, 2, \dots, 526.$$

Report the coefficient of determination R^2 saved in a variable called `R2`.

4. Compute the sample correlation coefficient between the dependent variable y and years of education x_2 . Save it as `rho_y2`.
5. The sample partial correlation coefficient $r_{y,3|2}$ can be computed as a sample correlation coefficient between e_2 and $e_{3|2}$, where e_2 denotes the vector of residuals from the regression $y = \alpha_1 + \alpha_2 x_2 + u$ and $e_{3|2}$ denotes the vector of residuals from the regression $x_3 = \gamma_1 + \gamma_2 x_2 + \xi$. Compute $r_{y,3|2}$ and save it as `rho_y32`.

6. Report the values of $1 - R^2$ and $(1 - r_2^2)(1 - r_{y,3|2}^2)$ as `ls` and `rs` respectively.

Verify that the relationship $1 - R^2 = (1 - r_4^2)(1 - r_{y,3|4}^2)(1 - r_{y,2|3,4}^2)$ holds for the following model:

$$y_i = \delta_1 + \delta_2 x_{2,i} + \delta_3 x_{3,i} + \delta_4 x_{4,i} + u_i, \quad i = 1, 2, \dots, 526.$$

Proceed as above and report the following quantities:

7. $1 - R^2$ as `ls2`.
8. r_4 as `rho_4`.
9. $r_{y,3|4}$ as `rho_34`.
10. $r_{y,2|3,4}$ as `rho_234`.
11. $(1 - r_4^2)(1 - r_{y,3|4}^2)(1 - r_{y,2|3,4}^2)$ as `rs2`.
12. Based on the estimation results for the model in specified in 6. :

$$y_i = \delta_1 + \delta_2 x_{2,i} + \delta_3 x_{3,i} + \delta_4 x_{4,i} + u_i, \quad i = 1, 2, \dots, 526.$$

Conduct a test of the following hypotheses pair: $H_0 : \delta_4 \geq 0$ vs $H_1 : \delta_4 < 0$ at the 5% significance level. Save the test statistic as `t1`, corresponding critical value as `cv1` and the p-value as `pv1`.

13. Next conduct a test of the following hypotheses pair: $H_0 : \delta_4 \leq -2$ vs $H_1 : \delta_4 > -2$ at the 1% significance level. Save the test statistic as `t2`, corresponding critical value as `cv2` and the p-value as `pv2`.
14. Which hourly wage does the model predict for a female worker with 12 years of education and 3 years of labour market experience? Save the answer as a variable `Ew_female`.
15. Which hourly wage does the model predict for a male worker with 12 years of education and 3 years of labour market experience? Save the answer as a variable `Ew_male`.