Carleton University

Department of Economics
ECON 3880, Summer 2024

Assignment 5

Due date: 23 June 2024, 11:59 pm

Professor: Firouz Fallahi

Q1 Using the **statcanR** package download the English version of the Table: 14-10-0106-01 (formerly CANSIM 282-0138) and call it **LF**.

- Clean the names of the columns
- keep only ref_date, geo, labour_force_characteristics, population_centre_and_rural_areas, sex, age_group, and value
- filter for 2019 in the ref_date; "Total, all population centres and rural areas" in the population_centre_and_rural_areas; "unemployment rate" in labour_force_characteristics; and "15 to 24 years" and "25 to 44 years" in the age_group.
- What are the unemployment rates of males and females aged 25-44 years old in Canada and provinces? To answer, create a table that shows the geo in rows, sex in the columns, and the unemployment rates in different cells
- create a point plot to visualize the above table. Where do we observe the biggest gap between the females and males' unemployment rate?

Q2 Write code to import the excel data file named **assignment5_data** from Brightspace and call it df. Estimate the following linear model and show the estimated coefficients. Are these independent variables statistically insignificant?

$$y = \beta_0 + \beta_1 * x1 + \beta_2 * x2 + \beta_3 * x3 + \beta_4 * z + u$$

Q3 Use the df data:

- create a new variable called **y_group,** based on the values for y and define y_group = 1 when y > average(y) and 0 otherwise
- estimate a logistic regression using the y_group as the dependent variable and x1, x2, x3, and z as the predictors
- calculate and show the confusion matrix and the accuracy rate for this model.

Good luck!