Geneva Graduate Institute (IHEID)

Econometrics I (EI035), Fall 2024

Marko Mlikota

Problem Set 4

Due: Sunday, 10 November, 23:59

- Prepare concise answers.
- State clearly any additional assumptions, if needed.
- Submit your solutions, along with any code (if applicable), in a **single pdf file** through **Moodle**. If you choose to write your solutions by hand, please make sure your scanned answers are legible.
- Grading scale:

5.5	default grade
6	absolutely no mistakes and particularly appealing write-up (clear and concise answers, decent formatting, etc.)
5	more than a few mistakes, or single mistake and particularly long, wordy answers
4	numerous mistakes, or clear lack of effort (e.g. parts not solved or not really attempted)
1	no submission by due date

Problem 1

You can find the data set for this question and a description of the variables on Moodle. The data spreadsheet contains four variables: average hourly earnings *ahe*, age *age*, gender $female$, and education *bachelor*. To answer the questions, use the asymptotic distribution of the OLS/ML estimator to conduct hypothesis tests or generate 95% confidence intervals.

(a) Run a regression of the logarithm of earnings on age, age^2, $female$, and *bachelor*. Based on your results, what are the predicted log-earnings of a 30 year old female with a bachelor degree? Note that you can write the quantity of interest as

$$\mathbb{E}[y_i|age = 30, female = 1, bachelor = 1] = \tilde{x}_i'\beta \,, \quad \text{where} \quad \tilde{x}_i = [1, 30, 30^2, 1, 1]' \,.$$

(b) Using a t-test and a significance level of $\alpha = 0.05$, can you reject the null hypoothesis that the expected hourly earnings of a 30 year old female with a bachelor degree are equal to 20 dollars per hour (i.e. that the expected log-earnings are equal to $\ln 20 \approx 2.99$)?
Hint: Note that we can write $\mathcal{H}_0 : \tilde{x}_i'\beta = 2.99$, with $\tilde{x}_i$ as defined above. Based on the (asymptotic) distribution of β, you can find that of $\tilde{x}_i'\beta$, which allows you to construct a t-test for that quantity.

(c) Using your t-test, construct a 95%-confidence interval for the expected log-earnings of a 30 year old female with a bachelor degree.

(d) Redo exercises (a) and (c) as a function of age. Concretely, plot the regression relation (the so-called age-earnings profile) between expected *age* (on the x-axis) and log *ahe* (on the y-axis) for the age range 20-65 for females with a bachelor degree, i.e. plot

$$\mathbb{E}[\log \, ahe \mid age, \, male, \, bachelor]$$

as a function of *age*. Also, overlay confidence bands around the age-earnings profile by plotting the 95% confidence interval for the above quantity as a function of *age*.

(e) Can you interpret the coefficient in front of *bachelor* as the causal effect of obtaining a bachelor degree on earnings? Discuss.

(f) By virtue of including both *age* and age^2, the regression you interpreted so far assumes a non-linear relationship between age and log-earnings, and this relationship is assumed to be the same for males and females. Keeping the assumption of such a non-linear relationship between age and log-earnings, test whether this relationship is different for males and females.
Hint: construct two covariates as the interactions $female * age$ and $female * age^2$, and test whether they are jointly (!) significantly different from zero.