# Econometrics 1

## TD4

## Exercise 1

Does sleep have an impact on a tennis player's performance? To answer this question, we use the dataset Tennis-Sleep.dta. Two tennis players, P1 and P2, played 73 matches against each other on Sunday mornings between November 2019 and May 2022. An observation in the database corresponds to a tennis match. For each match, we observe:

- the alcohol consumption of both players the evening and night preceding the game (measured in number of drinks): `AlcoholP1` and `AlcoholP2`,

- the number of hours slept by both players the night preceding the game: `SleepP1` and `SleepP2`,

- the outcome of the match measured by the difference between the number of games (un "jeu") won by P1 and the number of games won by P2: `DifgamesP1P2=GamesP1-GamesP2`.

1. Compute some descriptive statistics and discuss them.

2. Make the linear regression of `DifgamesP1P2` on `SleepP1` (and a constant/intercept as usual). Is the variable `SleepP1` statistically significant at 5%? at 10%?

3. The previous regression omits an *a priori* relevant variable, namely the consumption of alcohol of P1, `AlcoholP1`. It entails an omitted variable bias in the regression of Question 2. What sign do you expect for that bias? You will explain your answer in-depth.

4. Now, make the linear regression of `DifgamesP1P2` on `SleepP1` and `AlcoholP1`. Compare the estimated coefficient of `SleepP1` with the one obtained in the simple linear regression of Question 2. What do you conclude? Was your guess of Question 3 correct?

5. (a) Explain in-depth why there is no reason that the omission of `SleepP2` and `AlcoholP2` entails an omitted variable bias in the previous regression of Question 4.
   (b) Explain why it is nonetheless useful to include them in the regression.

6. Make the new regression that includes the additional control variables `SleepP2` and `AlcoholP2`. Comment on the results carefully (effects of the different variables, $R^2$, statistical significance tests, overall conclusion on the effect of sleep on performance).

7. We would like to know whether the effect of sleep on performance is identical for both players. Write the test of that hypothesis formally. Propose a regression that enables you to test the hypothesis directly (meaning here with a "simple test" in contrast to "multiple/joint tests"). What is the conclusion of the test?

## Exercise 2

The *hprice1.dta* database, that we use in this exercise, contains information on real estate sales in Boston. We know the selling price of each good, as well as some of its characteristics (number of rooms, land area in square feet, living area in square feet, etc.). We also know the good's valuation that was made by some real estate agents.

0. Make some basic descriptive statistics to apprehend the database.

First, we would like to know if the assessment (the variable *assess* in the database) is "rational", i.e. faithful to the selling price.

1. How can you use a simple linear regression to test that?
   *Hint: Think about (i) a relevant regression, (ii) a suited multiple/joint test. Write formally the null and alternative hypotheses of that test using the matrix R and vector b introduced in Chapter 2.*

2. Estimate the regression proposed in Question 1 and perform the test. What do you conclude? Perform the corresponding simple bilateral tests (Student t-tests). What do you remark? Is there any problem here?

Second, we would like to predict how selling prices respond to a change of the house area with an extra bedroom.

3. We consider the multiple linear regression of the logarithm of the house price on the area of the house and the number of bedrooms. Estimate that regression. What is the effect of adding a bedroom of 150 square feet on the price?

4. Propose a new regression that allows estimating the previous effect of adding a bedroom of 150 square feet on the price directly. Construct a 95%-level confidence interval on that effect.