STA502: Math & Stat for MBA I
Final Exam 2021

Question 1. The galaxy of SIAI consists of 34 planets. Each planet has their own university and students attend uni on their home planet. Students in SIAI either have to attend lectures in person or watch them online through hyper-space real-time communication. Researcher A has obtained data for the students from all the universities and would like to study the effect of watching lectures online on the students' exam scores. For each student, she has the exam score (0 - 100, 70+ is an A, 40 and below a fail), which she uses as her left hand side variable, the fraction of lectures watched online, and how many days the student visited the library each week. Researcher A obtains the following regression results:

Regressor	(1) - OLS	(2) - IV	(3) - IV	(4) - IV	(5) - IV
Fraction online	-4.91 (0.11)	-2.08 (0.22)	-0.56 (0.40)	-0.75 (0.90)	6.09 (2.22)
Library visits	–	–	0.91 (0.08)	–	2.09 (0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each universities. Robust standard errors are reported in parentheses.

1. What is the interpretation of the coefficient in column (1). If this were a causal effect, would it be big or small? Explain whether this estimate is likely to have a causal interpretation. (5 marks)

2. Researcher B observes that some halls of residence are close to the lecture rooms while others are further away and suggests to use the distance of halls from lecture halls as an instrumental variable for the fraction of lectures watched online.

   Results for this IV regression are displayed in column (2). Explain why instrumental variables may produce a better estimate of the causal effect of watching lectures online, and which assumptions need to be satisfied for this to be the case. Discuss the validity of the assumptions in this case. Can you ascertain whether any of these assumptions are true from the results in the table above? (5 marks)

3. Researcher C points out to B that students can pick which hall they want to live in and that some halls are located in Study Village, close to lecture halls and the library while others are located further away in Party Town, surrounded by pubs. How does this information affect your assessment of the IV strategy? (5 marks)

4. Researcher D realises that the data also include a variable for the number of times a student has checked into the library per week. He suggests to rerun the instrumental variables regression adding this variable as a control. Results for this regression are displayed in column (3). Assess D's strategy. (5 marks)

5. Researcher E notices that there are five universities which assign students to their halls of residence by a lottery. She suggests to run the IV model from columns (2) and (3) for the subsample of students from these universities only. Results are displayed in columns (4) and (5). Assess E's regressions. (10 marks)

6. Drawing on the results in the table above, what have you learned from this exercise about the causal effect of watching lectures online on students' exam results? (5 marks)

7. Thanks to massive developments in physics, there is a warping device that teleports a person to another place in a nano second, regardless of distance within a planet. What will be the consequence to above estimation model? (5 marks)

---

8. In the presence of new technology, the union of university councils in all planets announce that they will no longer keep dormitories and libraries. Students are encouraged to study at their home but must come to school for all classes. If regression model (5) has high enough R^2, in terms of students performance in exams, what do you expect in the following year? If you were the principal, what would you do? (10 marks)

Question 2. SIAI is interested in the effect of computer & console game playing on future earnings in a country where gaming covers a large portion of IT business and IT is the only growing industry. It has a survey of 1,000 students aged 25 to 35 after SIAI's MBA in AI programs. The key variable "game playing" (GP) is a dummy variable indicating whether the individual has been playing games regularly over the past three years. Running regressions with the log of grades on the left hand side he obtains the following results:

Regressor	(1)	(2)	(3)	(4)	(5)
Game playing (GP)	0.118 (0.021)	0.182 (0.025)	0.138 (0.030)	0.169 (0.021)	0.027 (0.022)
Male * GP	–	-0.120 (0.023)	0.015 (0.044)	–	–
Male	–	–	-0.135 (0.036)	-0.123 (0.019)	-0.246 (0.019)
Age	–	–	–	0.070 (0.053)	0.065 (0.053)
Age squared	–	–	–	-0.0002 (0.0009)	-0.0001 (0.0009)
Response time in millisecond (ms)	–	–	–	–	0.050 (0.009)

All regressions also contain a constant term. Robust standard errors are reported in parentheses.

1. Give a story why playing games may have a positive causal effect on earnings, and a different story why it may have a negative causal effect. (5 marks)

2. Interpret the coefficient in column (1). Explain whether this estimate is likely to have a causal interpretation. (5 marks)

3. SIAI2 reacts to these results: "The results in column (2) say that men who play games have lower earnings than those who don't. That's really strange." Explain why SIAI2 is wrong on two counts. (5 marks)

4. How would you carry out a statistical test that playing games have no effect on the earnings of females using column (3)? If you can, derive the result of this test from the information in the table or explain why you can't. (10 marks)

5. For the regresson (4), what is the reasoning behind the regressor Age? What's the statistical point of adding $Age squared$? What can you conclude for the use of squared term? (5 marks)

6. Explain whether you can interpret the results in column (5) causally. (5 marks)

7. Do men have faster or slower in response time than women? Explain how you arrive at your answer. (10 marks)

8. SIAI3 claims that his undergrad study in genetics and biology has taught him that response time in millisecond (ms) to a quick motion is largely affected by an individual's testosterone level, which is not so much an acquired skill like hard practice in game playing as it is mostly determined at the DNA level. How would you leverage his knowledge in SIAI's research? For that what other data sets do you need? (5 marks)