

“If at any point in this exam you feel that anything is unclear, please make additional assumptions that you feel are necessary and state them clearly.” Write concise answers. Precise arguments will be rewarded. Including irrelevant information in your answers will not improve your marks and will reduce the time you have to answer the questions directly. Marks will be deducted for incorrect irrelevant information. Make sure to allocate your time accordingly (on average, 5 marks are worth 3 minutes).

Question 1

The Ministry of Truth is interested in a rumour that air pollution could impact mental health. One of the most harmful pollutants is fine particulate matter PM2.5, which comes from operations that involve the burning of fuels such as wood, oil, coal, gas, or grass fires. A research team is sent to investigate the rumour. The team randomly selects and surveys 19,920 people across 71 districts of the country. The key variable, $Exposure_i$, is a dummy variable equal to 1 if the individual i is exposed to a large amount of PM2.5 in the last two years, and 0 otherwise. The team also conducts a standardised questionnaire to record depressive symptoms in the last month, called the Kessler Psychological Distress scale (K6). The questionnaire results in a score, $Depression_i$, that ranges from 0 to 24; and the higher the score, the more severe the depressive symptoms for individual i . The variable has a sample average of 2.96. Running regressions with $Depression_i$ as the dependent variable, the analyst Winston Smith obtains the following results:

Dependent variable: $Depression_i$						
Regressor	(1)	(2)	(3)	(4)	(5)	(6)
$Exposure_i$	0.834 (0.032)	0.635 (0.039)	0.614 (0.045)	0.598 (0.021)	0.422 (0.020)	0.554 (0.042)
$Exposure_i \times Female_i$		-0.834 (0.013)	0.065 (0.024)	0.054 (0.011)		
$Female_i$			-0.739 (0.036)	-0.732 (0.018)	-0.745 (0.019)	-0.825 (0.066)
$Manual Job_i$				0.324 (0.122)	0.050 (0.008)	
Age_i					0.324 (0.111)	0.452 (0.132)
Age_i^2					0.421 (0.122)	0.524 (0.121)

Notes. All estimations contain a constant term. Robust standard errors are in the parentheses. $Manual Job_i = 1$ if i works in a manual job, and 0 otherwise. $Female_i = 1$ if i is a female, 0 otherwise. Age_i is the age (years old) of individual i , and Age_i^2 is the square of Age_i .

(Question continues next page)

- a) Interpreting the coefficient in Column (1), a journalist, Katherine, claims: “Since participants are randomly selected, we can infer that exposure to a large amount of PM2.5 does cause depression.”
- i. Explain carefully why Katherine is wrong, specifying the direction of bias(es) if there is any. Which assumption(s) would she need to impose for the causality claim to hold?
[13 marks]
 - ii. What is the correct interpretation from Column (1) that Katherine should have made?
[2 marks]
- b) Consider the estimated coefficient of $Exposure_i$ in Column (4). How would you conduct a statistical test that this coefficient is different from 0.5 at the 5% significance level? If you can, derive the result of this test from the information in the table. If you can't, explain why. Note that the critical values for t-statistics are $t_{0.05,\infty} = 1.645$, $t_{0.025,\infty} = 1.96$, $t_{0.01,\infty} = 2.326$, $t_{0.005,\infty} = 2.576$.
[15 marks]
- c) Another analyst, Julia, suggests that the team should include into the regressions a variable, $Health\ Expenditure_i$, which captures individual i 's total expenditures on health-related services in the last two years. Her rationale is that financial distress may cause depression and is related to air pollution in the area where people with financial distress work. Should the team follow her suggestion? Explain why or why not.
[10 marks]
- d) O'Brien, the chief investigator of the project, tells the team that the computers used in the survey were affected by a virus. The virus changes the scores from the depression questionnaire (K6) for some random individuals by random amounts. Should the team be worried about the incidence when interpreting the results? Explain why or why not.
[10 marks]

Question 2

In Britannia, a premium supermarket chain, Moonania, is interested in whether its customer loyalty cards affect its revenues. Customers can voluntarily sign up for a *free* loyalty card. It collects points as credits every time a transaction is recorded. The customers can use the credits for future discounts. A team from Moonania has put together a new database on every Moonania store. The database includes data on the revenues and the number of loyalty cards issued in the last financial year.

- a) Give a plausible story why customer loyalty cards may have a causal effect on revenues.

[5 marks]

- b) The team runs the following regression:

$$\ln(\text{Revenue}_i) = 4.652 + 0.0875 \text{ Loyalty Cards}_i + e_{1i} \quad (1)$$

where Revenue_i is the revenue in million pounds (£) for store i , Loyalty Cards_i is the number of loyalty cards (in '000 units) issued by store i .

- i. What is the interpretation of the estimate 0.0875?

[5 marks]

- ii. Does the estimate 0.0875 reflect the causal effect of customer loyalty cards on stores' profits? Explain your answer. Explain a potential source of bias you would expect in Regression (1). Which directions would this bias go and why?

[15 marks]

- c) A discount supermarket chain, Sunania, is also interested in whether its new loyalty card scheme affects revenues. Sunania decides to have a trial for their stores in the North. Amongst 500 stores in the North, Sunania randomly chooses 250 stores to run an aggressive marketing campaign to promote its loyalty cards. A research team from Sunania uses this marketing campaign as an instrument for the number of loyalty cards to estimate the causal effect of loyalty card on revenues. Explain how this instrument would help mitigate the bias(es) that the Moonania team faces. Discuss and evaluate any assumptions that the Sunania team would need to make.

[15 marks]

- d) Suppose we can obtain an estimate for the causal effect in Question 2c). Could we make a claim that loyalty cards do affect revenues in the same way for every supermarket chain in Britannia? Explain why or why not, providing at least two distinct reasons.

[10 marks]