23-07_qm.pdf

1. (30 points) **[Fundamental problem of causal inference]**

   (a) Explain what is meant with the fundamental problem of causal inference. In that context, define the meaning of average treatment effect (ATE) and the individual treatment effects (ITE). Discuss why the problem can be seen as a challenge related to missing data.

   (b) Discuss the "Difference in Difference (DiD)" method as one potential strategy to address the fundamental problem of causal inference.

   (c) Apart from the DiD method, explain another empirical strategy of your choice that can be used to tackle the fundamental problem of causal inference.

2. (30 points) **[Causal justification]**

   The directed acyclic graph (DAG) presented below originates from a research group that conducted qualitative research. They firmly believe that two variables, X_1 and X_2, have a causal impact on the outcome variable Y. Furthermore, they have evidence suggesting that these two variables have causal relationships with five other variables, as depicted in the DAG. The researchers have collected a large dataset consisting of seven metric variables: $X_1, X_2, C_1, C_2, C_3, C_4$, and C_5. Now, they aim to analyze the magnitude of the causal impact by performing a multiple regression analysis to quantify the effects of the two causes, X_1 and X_2.

   The group knows that you have some idea on quantitative research and they ask you to provide advice on which variables they should consider in their regression model and which variables they should handle with caution or disregard entirely.

   $$C_4 \qquad X_1 \longrightarrow C_1$$
   $$C_2 \longrightarrow Y$$
   $$C_5 \longleftarrow X_2 \longleftarrow C_3$$
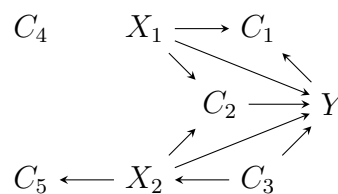
   (a) Give advice and explain your decision in detail. In that context, explain a *confounder*, a *mediator*, a *collider*, a *proxy*, and an *independent* variable.

   (b) Based on your recommendations, the researchers proceeded with the regression analysis and discovered that, ceteris paribus, X_2 has a statistically significant negative impact on Y. However, they have doubts about these findings. They firmly believe that X_2 and Y are positively associated due to a high positive correlation between these two variables. In other words, they observed that nearly all instances with relatively high values of X_2 also had high values of Y. This leads them to question whether the regression result can be accurate. In essence, they are asking whether X_2 can have a negative impact on Y despite their positive correlation. Please discuss this matter in detail.

End of exam

23-11_qm.pdf

Question:	1	2	3	Total
Points:	15	25	20	60
Score:				

1. (15 points) **[Features of good research]**

   (a) Briefly explain validity, reliability, and generalizability as important features of research.

   (b) Explain briefly the difference between the replicability and the reproducibility of studies.

2. (25 points) **[Experiments]**

   (a) Please complete the sentences:

   While randomized controlled trials (RCTs) assume the concept of ignoreability, most observational data present challenges in drawing causal conclusions due to the presence of _____

   that affect both

   (1) the _____ and

   (2) the _____ .

   (b) Explain what is a *Randomized Controlled Trial (RCT)* and why is it often referred to as the gold standard of research.

   (c) Discuss briefly the differences between natural experiments, laboratory experiments, and field experiments.

   (d) Briefly explain an example for a scientific natural experiment and a laboratory experiment.

3. (20 points) **[Observational data]**

   (a) Explain what is meant with the fundamental problem of causal inference. In that context, define the meaning of average treatment effect (ATE) and the individual treatment effects (ITE). Discuss why the problem can be seen as a challenge related to missing data.

   (b) The directed acyclic graph (DAG) presented below originates from a research group that conducted qualitative research. They firmly believe that two variables, X_1 and X_2, have a causal impact on the outcome variable Y. Furthermore, they have evidence suggesting that these two variables have causal relationships with two other variables, as depicted in the DAG. The researchers have collected a large dataset consisting of several metric variables: $X_1, X_2, C_1,$ and C_2. Now, they aim to analyze the magnitude of the causal impact by performing a multiple regression analysis to quantify the effects of the two causes, X_1 and X_2.

   The group knows that you have some idea on quantitative research and they ask you to provide advice on which variables they should consider in their regression model and which variables they should handle with caution or disregard entirely.

Please go on to the next page...

$$X_1$$
$$C_1 \longrightarrow Y$$
$$X_2 \longrightarrow C_2$$
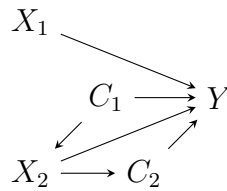
i) Give advice and explain your decision in detail.

ii) Based on your recommendations, the researchers proceeded with the regression analysis and discovered that, ceteris paribus, X_2 has a statistically significant positive impact on Y. However, they have doubts about these findings. They firmly believe that X_2 and Y are negatively associated due to a high negative correlation between these two variables. In other words, they observed that nearly all instances with relatively high values of X_2 also had low values of Y. This leads them to question whether the regression result can be accurate. In essence, they are asking whether X_2 can have a positive impact on Y despite their negative correlation. Please discuss this matter in detail.

References

End of exam