23-07_qm.pdf

This box is for the examiner only.

Question:	1	2	Total
Points:	30	30	60
Score:			

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.

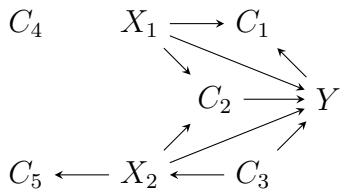
   (b) Discuss the "Difference in Difference (DiD)" method as one potential strategy to address the fundamental problem of causal inference.

   (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.



   (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

This box is for the examiner only.

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.

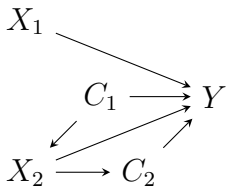
3. (20 points) **[Observational data]**

   (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...



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

24-07_qm.pdf

This box is for the examiner only.

Question:	1	2	3	4	Total
Points:	24	15	12	9	60
Score:					

1. **[Various]**

   (a) (9 points) The terms *internal validity*, *external validity*, and *generalizability* describe important features of research. Define briefly what each of these terms describes.

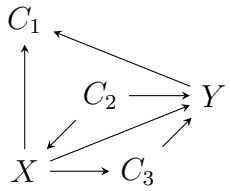
   (b) (5 points) Briefly describe the *fundamental problem of causal inference* as discussed in the *Rubin causal model*.

   (c) (10 points) The identification of a causal effect is one of the core challenges of empirical research. Explain in detail how a *Randomized Controlled Trial (RCT)* can help address this challenge.

2. (15 points) **[Publish]** Ensuring the validity and scientific integrity of research is a significant challenge. In recent years, there has been a widespread adoption of an open research culture to minimize the likelihood of disseminating and endorsing incorrect empirical findings. The paper by Nosek et al. (2015), which we reviewed in the lecture notes, outlines eight standards of science, each aimed at enhancing the transparency of scientific communication. Identify and explain five of these standards in detail.

3. (12 points) **[Observational data]**

   In the lecture, we discussed the paper by Wysocki et al. (2022), which taught us that statistical control requires causal justification. Using Directed Acyclic Graphs (DAGs), the paper discusses three types of variables associated with a variable of interest: a *Confounder*, a *Mediator*, and a *Collider*. In the DAG shown below, you will find these three variables. Assign the variables to these three categories and discuss which variable we should include if we are interested in the impact of X on Y.



4. (9 points) **[Types of research methods]**

   In the lecture, we discussed the paper by Paldam (2021), which classifies different types of research methods:

   1. Economic theory
   2. Statistical technique, incl. forecasting
   3. Surveys, incl. meta-studies
      a) Assessed surveys
      b) Meta-studies
   4. Experiments in laboratories
   5. Event studies (field experiments and natural experiments)
      a) Field experiments
      b) Natural experiments
   6. Descriptive, deductions from data
   7. Classical empirical studies
   8. Newer techniques

   Below, you will find abstracts from three different papers. Assign each paper to one of the categories discussed in Paldam (2021) and briefly explain your decision.

   (a) Abadie (2005): *"Semiparametric Difference-In-Differences Estimators"*

   > **Abstract:** The difference-in-differences (DID) estimator is one of the most popular tools for applied research in economics to evaluate the effects of public interventions and other treatments of interest on some relevant outcome variables. However, it is well-known that the DID estimator is based on strong identifying assumptions. In particular, the conventional DID estimator requires that, in absence of the treatment, the average outcomes for the treated and control groups would have followed parallel paths over time. This assumption may be implausible if pretreatment characteristics that are thought to be associated with the dynamics of the outcome variable are unbalanced between the treated and the untreated. That would be the case, for example, if selection for treatment is influenced by individual-transitory shocks on past outcomes (Ashenfelter's Dip). This paper considers the case in which differences in observed characteristics create non-parallel outcome dynamics between treated and controls. It is shown that, in such case, a simple two-step strategy can be used to estimate the average effect of the treatment for the treated. In addition, the estimation framework proposed in this paper allows the use of covariates to describe how the average effect of the treatment varies with changes in observed characteristics.

   (b) Szymanski et al. (1995): *"Order of Entry and Business Performance: An Empirical Synthesis and Reexamination"*

   > **Abstract:** One stream of research for order of entry effects focuses on the possibility that the order of entry exerts a direct impact on business performance. A second stream of research, the contingency perspective, debates the merits of whether the order of entry, in combination with other market strategy and marketplace variables, is what actually drives business performance. The findings from studies focusing on possible direct effects offer only mixed evidence in favor of a pioneering advantage. The contingency perspective, however, has not been subjected to systematic, empirical scrutiny. Against this backdrop, the authors conduct a meta-analysis of the

Please go on to the next page…

pioneering-market share findings and an examination of the contingency perspective of order of entry effects. The findings from the meta-analysis reveal that, on average, earlier entry is associated with greater market share. The findings from the contingency analysis, however, offer evidence suggesting that the contingency perspective is the more valid perspective for capturing the association between order of entry and market share. The authors discuss the theoretical and managerial implications of their findings and several directions for further research.

(c) Hamilton (2000): *"Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment"*

Abstract: Possible explanations for earnings differentials in self-employment and paid employment are investigated. The empirical results suggest that the nonpecuniary benefits of self-employment are substantial: Most entrepreneurs enter and persist in business despite the fact that they have both lower initial earnings and lower earnings growth than in paid employment, implying a median earnings differential of 35 percent for individuals in business for 10 years. The differential cannot be explained by the selection of low-ability employees into self-employment and is similar for three alternative measures of self-employment earnings and across industries. Furthermore, the estimated earnings differentials may understate the differences in compensation across sectors since fringe benefits are not included in the measure of employee compensation.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. The Review of Economic Studies, 72(1):1–19.

Hamilton, B. H. (2000). Does entrepreneurship pay? an empirical analysis of the returns to self-employment. Journal of Political Economy, 108(3):604–631.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., et al. (2015). Promoting an open research culture. Science, 348(6242):1422–1425.

Paldam, M. (2021). Methods used in economic research: An empirical study of trends and levels. Economics, 15(1):28–42.

Szymanski, D. M., Troy, L. C., and Bharadwaj, S. G. (1995). Order of entry and business performance: An empirical synthesis and reexamination. Journal of Marketing, 59(4):17–33.

Wysocki, A. C., Lawson, K. M., and Rhemtulla, M. (2022). Statistical control requires causal justification. Advances in Methods and Practices in Psychological Science, 5(2).

End of exam

24-11_qm.pdf

This box is for the examiner only:

Question:	1	2	Total
Points:	36	24	60
Score:			

Quantitative Methods (11/2024)

1. **Various**

   (a) (6 points) The terms *validity*, *credibility*, and *reliability* describe important features of research. Define briefly what each of these terms describes.

   (b) (5 points) Briefly describe the *fundamental problem of causal inference* as discussed in the *Rubin causal model*.

   (c) (10 points) Explain the basic setup of a *Randomized Controlled Trial (RCT)*. Why is it often called the gold standard of empirical research?

   (d) (9 points) Explain the difference of the following types of experiment in empirical research: (1) Laboratory Experiment, (2) Field Experiment, and (3) Natural Experiment.

   (e) (6 points) In the lecture we have discussed the following statement of Peng (2011, p. 1226):

   > "*Replication is the ultimate standard by which scientific claims are judged.*"

   Discuss the importance of replication in ensuring scientific standards and quality.

2. (24 points) **Observational data**

   In the lecture, we discussed the paper by Wysocki et al. (2022). Using Directed Acyclic Graphs (DAGs), the paper discusses various types of variables associated with a variable of interest. Assume you are interested in the impact of a variable X on Y.

   For a regression analysis, how should you treat the following types of variables: (1) Confounder, (2) Mediator, (3) Collider, (4) Proxy, and (5) Independent.

   Please visualize these relationships using DAGs and discuss how each of these variable types should be incorporated into a regression analysis.

References

Peng, R. D. (2011). Reproducible research in computational science. *Science*, *334*(6060), 1226–1227.

Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science*, *5*(2). https://doi.org/10.1177/25152459221095823

25-01_qm.pdf

This box is for the examiner only:

Question:	1	2	3	4	Total
Points:	31	12	12	5	60
Score:					

Quantitative Methods (01/2025)

1. **Various**

   (a) (3 points) The terms *serendipity*, *structured data* and *unstructured data* are used in the 'Doing Research' section of the lecture notes and are included in the glossary. Please provide a brief definition of each of these terms.

   (b) (6 points) Briefly describe the *fundamental problem of causal inference* as discussed in the *Rubin causal model*.

   (c) (10 points) Explain the basic setup of a *Randomized Controlled Trial (RCT)*. Why is it often called the gold standard of empirical research?

   (d) (4 points) Briefly explain the difference betwen a *field experiment*, and a so-called *natural experiment*.

   (e) (8 points) In the lecture, we discussed the challenges of drawing insights from anecdotes. Please briefly recall the four key points that were emphasised in the lecture notes.

2. **Good research**

   (a) (8 points) In the lecture, we discussed various tactics that authors commonly employ to manipulate readers in an unscientific manner. Please briefly explain at least four methods that these 'manipulators' use to present their arguments in a polemical fashion while disregarding counter-arguments that could potentially weaken their position. These methods were covered in the lecture notes under '7.1 Information and Insights'

   (b) (4 points) In the lecture we have discussed the following statement of Peng (2011, p. 1226):

   "*Replication is the ultimate standard by which scientific claims are judged.*"

   Discuss the importance of replication in ensuring scientific standards and quality.

3. (12 points) **Observational data**

   Using Directed Acyclic Graphs (DAGs), Wysocki et al. (2022) discuss various types of variables associated with a variable of interest within a regression analysis framework. Assume you are interested in the overall impact of variable (X) on variable (Y).

   Please visualize the relationships between (Y) and (X) with (1) a confounder and (2) a mediator using DAGs, and discuss how each of these variable types should be incorporated into a regression analysis.

4. **Statistical Significance**

   (a) (1 point) After conducting a hypothesis test on the estimated coefficient in a regression analysis, the p-value is found to be 0.04. Is this result statistically significant at the 5% significance level?

   (b) (4 points) A researcher runs a regression and obtains the following results for a predictor variable (X): The estimated coefficient is $\hat{\beta} = 2.5$ and the standard error of the coefficient is 0.5. Calculate the t-value for the predictor (X). Is this estimated coefficient statistically significantly different to zero at the 5% significance level? Explain your decision.

References

Peng, R. D. (2011). Reproducible research in computational science. *Science*, *334*(6060), 1226–1227.

Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science*, *5*(2).