

Take Home Exercise 3: Final Research Paper

Course: Statistics and Programming I

Instructor: Dr. Fabian Habersack

Due Date: 17 December 2024

1 Objective

Building on the research concept developed in THE2, this exercise transitions into the empirical phase of your research. For your final research paper, you will collect your own data, which may be unclean or require pre-processing. You will then perform descriptive and inferential statistical analyses, and interpret your findings to test the theoretical claims about your chosen variables, X and Y . This exercise is designed to deepen your knowledge of statistical analyses and data visualization skills through using real-world data.

2 Task Details

2.1 Data Collection

Identify and collect data relevant to your research question, in order to operationalize X_1 , Y , and any control variables identified in the theoretical discussion. A suggested list of data sources is available on Moodle (THE1). Ensure that the data accurately represents the variables defined in your research design (THE2).

2.2 Empirical Analysis

Conduct an empirical analysis in three steps, providing relevant statistics and results in the form of tables of figures in your research paper. In general, make sure to document your code and analytical processes thoroughly to ensure replicability of your results. Do not, however, directly include your code in the main text of your research paper!

2.2.1 Univariate Analysis

Begin your empirical analysis by providing descriptive statistics for each of the primary variables in your study, namely your independent variable (X_1), dependent variable (Y), and any control variables (X_k). Ensure that the statistics and accompanying visualizations are appropriate for the measurement levels of each variable. This section should focus on exploring the extent of variation within your variables — such as the range, median, mean, and standard deviation — and identifying any potential distributional biases or deviations from the normal distribution related to

skewness and kurtosis. Optionally, you may include a (univariate) graph such as a histogram or boxplot to visualize the distribution of selected variables.

2.2.2 Bivariate Analysis

Explore the relationship between your independent variable and dependent variable using bivariate statistics and bivariate graphs. This could include calculating correlation coefficients or creating scatter plots or line graphs to visually represent the relationship. Discuss the degree to which these variables vary together and any visible trends or patterns. Again, ensure that the statistics and graphs match the measurement level of your data.

2.2.3 Multivariate Analysis

Perform a multivariate Ordinary Least Squares (OLS) regression analysis to control for potential confounding variables (X_k) and more rigorously test your hypothesis. Include at least three control variables in your models. Interpret the coefficients (significance, direction, and effect size) to understand the impact of your independent variable on the dependent variable, accounting for the scale of these variables. Then, assess the overall model fit, interpreting summary statistics like R^2 and the adjusted R^2 . Also briefly discuss any limitations of your study, e.g., related to the scope of your data and sample size. This section forms the crux of your analysis and should aim to shed light on the causal relationship between X_1 and Y .

2.3 Conclusion

Briefly conclude your analysis by providing an overview of your chosen research design and empirical results and reflecting on the wider implications of your findings. Discuss how your findings align with or deviate from the existing literature on the topic. Evaluate the implications of your findings, considering how they contribute to the existing body of knowledge and what they suggest for future research. Additionally, acknowledge any limitations of your study, such as potential biases, data collection constraints, or analytical challenges.

3 Evaluation Criteria

The following criteria will be used to evaluate your submission, with a total of 180 points available.

- **Theoretical Discussion (30 points):** This component assesses the clarity and depth of your theoretical argument and hypothesis. Your discussion should define the relationship between variables X_1 and Y , supported by a review of existing research. This review should also address competing explanations and discuss potential confounding variables (X_k) that could impact the relationship between X_1 and Y .
- **Descriptive Statistics (25 points):** Evaluation will focus on the selection and accuracy of descriptive statistics, including measures of central tendency, variability, and distribution shape. Precision in interpreting these statistics is essential. While graphical representations, such as histograms or boxplots, are optional, including and discussing them will be considered a bonus.

- **Bivariate Analysis (25 points):** Points are awarded for correctly selecting and applying statistical methods that explore associations between two variables, such as correlation coefficients or cross-tabulations. Clarity of visual representations (e.g., scatter plots) and the cogency of accompanying explanations regarding the nature of these relationships are critical.
- **Multivariate Analysis (60 points):** Assessment will focus on the correct specification and implementation of regression models to examine the impact of multiple independent variables on the dependent variable. Points will be awarded for the accurate interpretation of regression outputs, including the significance, direction, and effect size of coefficients, as well as the overall model fit.
- **Conclusion (25 points):** The conclusion should synthesize your empirical findings with the theoretical framework presented earlier, clearly linking your results to the existing literature. Evaluations will consider how effectively the conclusion addresses the research question, creating a “dialogue between theory and data.” Discuss the broader implications of your findings and acknowledge any limitations of your analysis, such as methodological constraints or data-related challenges.
- **Form and Presentation (15 points):** This criterion assesses the overall structure and organization of your paper, ensuring a logical and coherent flow of arguments. Proper citation practices, consistent formatting, and the professional presentation of data and figures are crucial.

4 Office Hours and Support

4.1 Scheduled Office Hours

To provide tailored support for THE3, I will offer individual online office hours in the remainder of November. Please register for a slot using the provided Google spreadsheet.

4.2 Walk-In Sessions

In addition to the scheduled office hours, I will host online walk-in sessions where you can join at any time during the session to discuss any aspect of your project or receive help with R. No need to sign up. Online walk-in sessions will be held at: <https://webconference.uibk.ac.at/rooms/fab-fc1-qrr-wzd>

4–6 December, 9:00–11:30 and 13:00–17:00

5 Submission Details

- **Due Date:** 17 December 2024.
- **Format:** R Script, replication data file, and PDF document of approx. 4,000 words (this is a reference value; the word count does not directly influence the overall evaluation). File name: `lastname_firstname_THE3.pdf`
- **Submission Platform:** Moodle.