ECON 340
Economics Research Methods

Div Bhagia

Research Project Guidelines

Overview

- As a part of this class, you will write an empirical research paper *using R* in groups of 2-3 students
- You will pick a question and a dataset and use the tools from this class to answer your question
- You can pick a dataset from the Econ340 Datasets Dropbox folder compiled for this class or use an external dataset
- There are also some external data sources on Canvas on the *Research Project Overview* page.

Milestones

- Aug 31: Form groups of 2-3 students
- Sep 26: First submission worth 5% (pick dataset and question)
- Oct 17: Feedback on your research question
- Oct 31: Second submission worth 10% (preliminary analysis)
- Dec 05: Final paper due worth 15%
- Dec 14: Last day to fill out the (optional) peer survey

Submission 1

Submission 1: Overview

- **Goal**: pick your dataset and research question
- You need to describe your research question and the data you will use
- Highlight why your question is interesting and important
- Mention the source of your data
- Highlight the variables you will use

Submission 1: Research Question

- Your question should concern the impact of one variable on another
- The variable whose impact you will be studying will be your *primary independent variable*, while your outcome of interest will be your *dependent variable*
- In addition, pick two more variables that you think might correlate with both the dependent and the independent variables
- We will call these additional variables *control variables*

Submission 1: Research Question

Your research question should be *well-defined* and *feasible*.

- Well-defined: Clearly specify your dependent and independent variables
- Feasible: Your dataset should contain all the variables

Submission 1: Dataset

- You can pick a dataset from the Econ340 Datasets Dropbox folder compiled for this class or use an external dataset
- There is a list of all the datasets in the folder; each folder contains the codebook (description of variables)
- If you want to use an external dataset, please run it by me in advance of your submissions
- Your dataset should meet the following criteria:
  - Have at least 100 observations
  - Contain your primary independent variable, dependent variable, and two control variables

Submission 1: What to turn in?

You need to submit the following:

- Filled Template for Submission 1 in PDF format
- Dataset in CSV or Excel format if using an external dataset.

If you plan to construct new variables in the data using some existing variables, specify that in the template.

Submissions need to be made on Canvas. Any member of the group can submit on your group's behalf. There is also an example of the filled template on Canvas.

Submission 2

Submission 2: Overview

- **Goal**: obtain a general understanding of the patterns that are present in the data concerning your research question
- Report the essential numerical characteristics of your variables and do some visual analysis
- You will use R to perform this preliminary data analysis.
- Prepare your submission in Microsoft Word and submit it in PDF format, along with R code and data (if changed significantly).

Submission 2: Tasks

1. Describe the data and your variables of interest. Mention if you constructed new variables or deleted observations.
2. Present summary statistics for all your variables and comment.
3. Present a graph showing the relationship between your dependent and primary independent variable
   - Scatter plot if both variables are continuous
   - Bar plot if one variable is binary

   Summarize what you learn.
4. Confirm what you found in (3) by finding the correlation

Submission 2: Tasks

5. In one or more tables, report the correlation between your dependent variable and control variables. Comment.

6. In one or more tables, report the correlation between your primary independent variable and control variables. Comment.

   You can combine 5 and 6 in a single table.

7. Comment: In what direction will your results be biased if you do not control for your control variables?

Final Submission

Final Paper: Overview

- The final paper should reflect:
  - your ability to perform and interpret statistical analyses
  - your writing ability (use Grammarly)
- Carefully read all the instructions on Canvas
- Incorporate the feedback you have gotten so far
- Rubric provides a breakdown of the score for different components

Final Paper: Submission

- The paper should be in essay format divided into sections (around 6-8 pages)
- Comply with the formatting requirements
- Final submission on Canvas, should include
  - write-up in PDF format
  - data (if required)
  - R script that reproduces the analysis (should run without error)

Research Paper: Sections

- Structure and content:
  - Introduction
  - Literature Review
  - Descriptive Analysis and Motivational Evidence
  - Empirical Strategy
  - Results and Analysis
  - Conclusion
  - References
  - Appendix (optional)

Instructions include the *suggested* length for each section.

I. Introduction

- What is your research question?
- Why is your research question interesting? Why is your topic important?
- Preview your results
- A statement that explains why you think your results are correlational or causal.

People who do not have time to go through the entire paper should get the general idea and conclusion of the paper through the introduction.

II. Literature Review

- What other work has been done on these issues? What has been found?
- Look for papers on Google Scholar (ChatGPT makes up papers that don't exist!)
- The literature review must preferably include at least two academic papers.
- Cite sources in APA format

How do your findings relate to these other papers?

APA Citation Format

Say I want to cite the paper, “The Rise of American Minimum Wages, 1912–1968” by Price V. Fishback and Andrew J. Seltzer published in the Journal of Economic Perspectives in 2021.

The way to include this citation:

- At the end of your paper on a new page start a References section and include

  > Fishback, P.V., and Seltzer A.J. (2021). The Rise of American Minimum Wages, 1912–1968. *Journal of Economic Perspectives*.

- In text: “Fishback and Seltzer (2021) find...”

III. Descriptive Analysis

Tasks 1-3 from Submission 2:

1. Describe the data and your variables of interest. Mention if you constructed new variables or deleted observations.
2. Present summary statistics for all your variables and comment.
3. Present a graph showing the relationship between your dependent and primary independent variable
   - Scatter plot if both variables are continuous
   - Bar plot if one variable is binary

   Summarize what you learn.

IV. Empirical Strategy

Explain how you plan to answer your question(s).

- Shortly recap your research question.
- Write down the main regression model and explain why it makes sense.
    - Interpret each coefficient. Do you expect the main coefficient to capture a causal impact?
    - Why did you pick the controls that you picked? Talk about how excluding these controls may impact your main coefficient of interest.

Empirical Strategy: Example

The objective of this paper is to investigate whether class size has an impact on test performance. To address this question, we estimate the following regression model:

$$testscr = \beta_0 + \beta_1 str + \beta_2 high\_comp\_stu + \beta_3 meal\_pct + u$$

Here, β_0 represents the intercept term. The main coefficient of interest is β_1, which corresponds to the student-teacher ratio variable. This coefficient captures the change in test scores associated with an additional student per teacher, holding other variables in the model constant. We expect β_1 to be negative if smaller class sizes are associated with higher test scores.

Empirical Strategy: Example (cont.)

However, it is possible that β_1 could be capturing the impact of other factors that are correlated with class size and test scores, such as better infrastructure or funding in the school district. To control for these potential confounders, we include two additional variables in our regression model. These variables are *high_comp_stu* and *meal_pct*.

high_comp_stu is an indicator variable that takes the value of 1 if the school has more computers per student than the median school and 0 otherwise. We expect that schools with more computers per student are associated with higher test scores, as computers can provide students with access to a wider range of educational resources and facilitate learning. Additionally, we expect schools with more computers also to have smaller class sizes. Therefore omitting this control variable could overestimate the effect of class size on test scores.

We also control for *meal_pct*... Interpret β_2 and β_3...

V. Results and Analysis

- In one or more tables, present the results of your regression analysis. Adding controls one by one.
- For this example, the code would be:

```
model1 <- lm(testscr ~ str, data)
model2 <- lm(testscr ~ str + high_comp_stu, data)
model3 <- lm(testscr ~ str + high_comp_stu + meal_pct,data)
stargazer(model1, model2, model3, type="text", keep.stat =
    c('n', 'adj.rsq'))
```

V. Results and Analysis

- Interpret the sign, value, and significance of the main coefficient across all specifications.
- Comment on how and why the coefficient changes or does not change as you add each control.
- Comment on the coefficients on your control variables.
- Interpret the adjusted R^2 of your models
- Discuss what you can conclude from your regression results. Correlation vs causation? Are there other omitted variables you are worried about?

	testscr		
	(1)	(2)	(3)
str	−2.28***	−1.74***	−1.00***
	(0.48)	(0.49)	(0.25)
high_comp_stu		7.99***	1.92**
		(1.84)	(0.94)
meal_pct			−0.59***
			(0.02)
Constant	698.93***	684.32***	699.36***
	(9.47)	(9.86)	(5.00)
Observations	420	420	420
Adjusted R^2	0.05	0.09	0.77

VI. Conclusion

- Summarize your results. Given this summary, answer your research question.
- Emphasize again why you think your results are causal or correlational.
- Compare with the literature
- Are there any policy implications of your findings? Are there suggestions for future research?