1 Research questions

A good research paper begins with a good research question. Good research questions can be drawn from your work, your studies, the sports you play (or watch), your hobbies, your personal life, the news, prior research, and/or careful observation of the world around you. For the purpose of this assignment, it is important that the research question you chose be 1) of interest to you, 2) empirical (meaning that it is a question that can be answered with data), 3) hypothesis-driven (more on this below), and 4) sufficiently limited in scope that you can acquire (through a survey, the internet, web-scraping, a research institution, an app, etc) data to answer the research question.

A hypothesis-driven research question is one that posits a testable relationship between (at least two) observable variables. For example, we may want to assess the statistical correlation between variables x and y, or evaluate the causal impact of x on y. A hypothesis-driven research question is one that states a relationship such that it can be tested with data via statistical or causal inference.

For example, "Why has the fertility rate in the U.S. fallen?" is too exploratory and broad to make a good research question for an empirical paper; it does not posit a testable relationship, a hypothesis. A better question would be, "Is there a correlation between household income and fertility?", or "Does an increase in the availability of public education reduce fertility rates?", or "Are income and education correlates of fertility rates?"

Once you have identified your research question, draft an introduction to your paper that makes clear to the reader that the research question driving your paper is an interesting one. Note that your job here is not to explain why the question is interesting to you, but to explain why the question is interesting to the reader. How do you know what the reader is interested in? You don’t. Therefore, you need to make the question interesting to a broad audience by connecting the (necessarily narrow) question you’re asking to larger themes/contexts.
2 Data

In addition to a good, hypothesis-driven, research question, you will need to identify the data you will use to empirically answer your research question. A number of resources are available to assist you in acquiring data to answer your research question; some good starting points include:

- Google Dataset Search
- Williams College Resource Page
- Yale Law Library List of Economic Data
- American Statistical Association Sports Data Page

If you have are having trouble locating data online, you can reach out to your professor, the course Teaching Assistant, and/or a Librarian for assistance. You may also collect your data via survey, web-scraping, experiments, various apps, etc.

Avoid basing your research on data that will require a lot of effort to get into the proper shape for analysis. In addition, data that are available only in summary or aggregate form and data stored in a way not immediately available for analysis (e.g., data that are reported in a summary statistics table in a pdf of a published paper) are not a good choice.

Once you have found your data, make note of where you found it. At a minimum, you will need to identify the source of your data in your research paper; at a maximum, you may find that you need to download the data again.

2.1 Data storage and analysis

You should set up a folder for this project on your personal computer or on your school storage drive. You might title the folder, EconResearch, Econ_Research or something along these lines. Within the project folder, you should include the following folders: Data, Drafts, and Literature. Download your data to the data folder. When you begin to work with and analyze the data, you must take care to not save over your original dataset. One great strategy to do this is to have both an original_data folder and a working_data folder within your Data folder; see, for example, Figure 1 below. With this folder structure, you can draw the original data into your RStudio workspace via a path to the original_data folder; you can then output any revised data as well as any analyses, figures, tables, or other outputs to the working_data folder. If, however, you are working with a single data set (and two different data subfolders seems like overkill), the key thing to not saving over your original data will be to save any work you do to the data under a different name than that given to the original data set.
3 Literature review

The purpose of a literature review is to identify research produced by other researchers who have posed a research question similar to yours. In reading the research relevant to your hypothesis, you will arrive at a better understanding of the hypothesis, the methods that are generally used to test the hypothesis, and the progress that has been made so far. In addition, you will be able to identify any shortcomings in the current literature and articulate the contributions your own paper will make. Research is never done in isolation; it is always done in communication with a larger body of research produced by others. The literature review allows you to identify, and start communicating with, that larger body of research.

Ordinarily, literature reviews will try to review all prior relevant research; however, in this class, your literature review will be limited in scope. To produce your literature review, you will need to identify, summarize, and analyze, at least three papers related to your topic. These papers should be research papers (not news articles or op-eds or other types of non-research media) that rely on data and statistics and are relevant to the hypothesis you have posited.

To identify the relevant research papers, you can search for key words related to your research topic in Google Scholar (a search engine dedicated to scholarly work) and/or use our library’s search options and/or seek out the help of a Librarian. Save the papers you’ve identified to your Literature folder (see Figure 1), as you will need to refer to them multiple times as you develop and revise your paper.

If your research question is very narrow and you are having trouble finding relevant sources, broaden the scope of your search. For example, if your hypothesis is that the Washington Nationals win more home games than away games but search terms “Washington Nationals home team advantage” yields nothing, then try searching “baseball home team advantage” or even “sports home team advantage”, as research that has been conducted on the phenomenon of home team advantage in, e.g., basketball is sufficiently relevant to your research question for you to include it in your literature review.

To summarize the relevant research papers, you should first read them thoroughly. Once you have read the papers, write (at least) a paragraph about each paper that describes the research question that was posed in the paper, the data and methods used to answer the research question, and the paper’s findings. Do not include extensive quotes from the paper you are summarizing in your literature review; use of long quotes
suggests that you don’t understand the material well enough to put it in your own words.

**To analyze** the relevant research papers, you should discuss the contributions and limitations of each of the papers in terms of how well they have addressed the research question. You should also discuss any similarities or differences among the different papers; e.g., do the papers use similar data and methods? If the data and methods differ, why do they differ? Do the papers have similar findings and conclusions? If the findings and conclusions differ, why do they differ? What are the limitations of each of the papers? What have they left unaddressed, and why?

Once you have discussed the contributions, limitations, similarities, and differences among the papers, discuss the contribution you intend to make with your paper: how does your research question and the data you will use relate to these other papers? What is “new” about your paper in light of the research that has already been done on this topic? What are you adding to the discussion?

Your write up of this portion of your paper should be coherent and well written. It should not take the form of an annotated bibliography and it should not be a discussion of what you have learned from the papers. Your literature review is not a diary of your reading of the literature. Rather, it is a summary and analysis of scholarly research, guided by your research question. Once you have completed your literature review, you should have a better understanding of the data needed to answer your research question as well as the methods commonly used to answer your research question. You may find that you need to refine your data search, or even your research question, at this point. Don’t worry, this is part of the process; engagement with the literature often helps us to revise and refine our research questions, data, and methods.

3.1 In-text citations and references

Each discipline has its own citation format (e.g. English majors tend to use MLA, Economist tend to use APA); proper use of the conventional citation format signals to the reader that you are a member of that discipline. Many online resources are available for assistance with the APA citation format; see, for example, the [Purdue Online Writing Lab](https://wwwWritingLab).

Be sure to use in-text citations for each of the papers you include in your literature review. In-text citations include the author’s name, or authors’ names, as well as the year of publication. For example, I might write, ”Pollard (2002) finds that the home team advantage effect is reduced when a team moves to a new stadium.” In addition to providing in-text citations, you must list all your references at the end of your paper in a References section. A general rule of thumb with in-text citations and references is that you need to provide enough information about the source that an interested reader can locate the source.
Please see the Writing Center for help if you are struggling with in-text citations and references. Note that Google Scholar will produce APA citations for you if you click on the quotes ("”) symbol under a paper. See Figure 2 for an example. Make sure to check the Google Scholar generated citation before you paste it into your paper; sometimes it will have errors or missing information. It is your responsibility to make sure the citation is correct.

(a) Click on the quotes ("”) symbol below paper

(b) Copy and paste the APA reference into your paper’s references section

Figure 2: Google Scholar generated references

4 Summary statistics

Before conducting an analysis of your data, you need to understand your data and describe your data to your reader. Some basic details about your data that you should understand and communicate to your reader include: Are your data cross-sectional, time series, or panel data? Who collected the data and how did they collect it? When were the data collected and what time period do they cover? What (relevant) variables are included in your dataset? Include labels and units when describing your data (e.g. income in 100,000 USD). How many observations do you have? Are any data missing? If they are missing, why are they missing?
Provide some basic **summary statistics** of the variables that are most important in your analysis; present these summary statistics in a table or figure. Generally a reader of your empirical economic research will want to be able to get an idea of how the data are distributed and what the sources of the variation in your data are; therefore, meaningful summary statistics might include a table of the means and the standard deviations, or the medians and the interquartile ranges, of the key variables. In addition, you might include density plots or histograms. If your analysis involves the relationship between two variables, then the correlation coefficients and/or scatter plots may also be meaningful. Note any outliers in these figures and tables and the impact you expect them to have on the analysis.

Provide a discussion of your summary statistics that explains what you have found in the tables or figures you produced. In other words, describe the relevant key features of the data, using the tables and figures to assist you.

### 4.1 Tables and figures

Tables and figures should never speak for themselves. In addition, a reader should be able to understand what your table or figure conveys from the text alone, without ever looking at the figure. There are two reasons for this: 1) you want to retain rhetorical control over your tables and figures; 2) some readers will not look at your tables and figures at all. Why are the tables and figures included then? As with references, they are provided to support the claims you make in the text. Similarly, you need to provide enough information in the label and/or notes to your tables and figures that a reader can understand it without looking at the text.

In fact, just as with providing references in your paper, you should provide sufficient information about your table/figure that the interested reader could reproduce it. For example, you should include the time period and location to which the data pertain, any units of measurement, and the source of the data or figure. If you did not create the figure yourself, this should be clearly noted in both the figure and in the text, and a proper reference for the original figure should be provided in the references section.

Label and number all tables and figures. For example, a labeled figure might look like that shown in Figure 3. Note that the measure of life expectancy (years) and the measure of income (USD) are indicated in the figure’s label. In addition, the source of the figure is clearly indicated in the label.

Discuss your summary statistics and explain what is depicted in the table or figure by referencing the table or figure’s number. For example, you might write, "Figure 3, produced by GapMinder (2019), displays a positive linear relationship between life expectancy and income..."
Figure 3: Life expectancy (years) by income (USD), 2018 (source: GapMinder (2019))

5 Methods

After describing your data and discussing your summary statistics, provide a detailed explanation of the statistical and/or econometric methods you’ll apply to your data to test your hypotheses and answer your research question. This discussion should include, the hypothesis(es) you wish to test, the method(s) you’ll use to test the hypothesis(es), and any formulas or equations that are relevant to your selected analytical method including, if relevant, the regression model you will estimate. You should also identify any assumptions you’re making about the data that might support or undermine the use of your selected methods. In particular, if you are writing an econometrics paper, you should explain how you will address failures of the homoscedasticity and exogeneity assumptions. Note that the identification of appropriate methods for your research question and data may require reading ahead in the textbook and/or reaching out to me or the course TA for guidance.

You must **formally state your research question as a testable hypothesis** (or a set of testable hypotheses), clearly indicating your null and research hypotheses. For example, suppose my specific, hypothesis-driven, research question is, ”Is there a negative correlation between income and fertility rates across countries between 1950 and 2010?” I would perform my hypothesis test as a test of the null hypothesis of positive or no correlation between income and fertility as follows, where \( \rho \) indicates correlation,
\[ H_0 : \rho_{\text{income}, \text{fertility}} \geq 0 \]
\[ H_A : \rho_{\text{income}, \text{fertility}} < 0 \]

Or suppose my specific, hypothesis-driven, research question is, "Are income and education correlates of fertility rates across households within the U.S. between 1980 and 2010?" I would first specify my regression function as follows, where fertility is the dependent variable, education and income are the independent variables, observed at the household level, and the \( \beta \)'s are the coefficients,

\[ \text{fertility} = \beta_0 + \beta_1 \text{education} + \beta_2 \text{income} + u \]

I would then indicate my hypotheses in terms of the regression coefficients,

\[ H_0 : \beta_1, \beta_2 = 0 \]
\[ H_A : \beta_1, \beta_2 \neq 0 \]

Or suppose my hypothesis-driven research question is, "Did the Washington Nationals win more home games than away games during the 2018 season?" I would formally state my equivalence of means test as follows, where \( \mu \) indicates the average number of games won during the season,

\[ H_0 : \mu_{\text{home}} \leq \mu_{\text{away}} \]
\[ H_A : \mu_{\text{home}} > \mu_{\text{away}} \]

Please type up any hypotheses, equations and/or formulas using LaTeX or the equation editor in Word or Google Docs.

6 Analysis and discussion of findings

In the analysis and discussion of findings portion of your research paper, you will implement the methods you specified in the methods section to test the hypotheses driving your research question. Report coefficient estimates, test-statistics, p-values and other relevant results of your analysis that will allow the reader to assess the statistical validity of your results. When reporting your results, be sure to quantify any uncertainty about your findings using standard errors and/or confidence intervals, where appropriate.

If you have multiple coefficient estimates (or other findings), you might summarize these in a single table. Remember to label any figure or table you have produced as a part of your analysis, and to include and discuss these in your write up. As with summary statis-

\(^1\)Note that I could also present the prior example as a regression, using the same approach as I use here, just dropping the education variable.
tics, a figure/table displaying the results of your analysis should never speak for itself. As the researcher, you should describe the important information that the figure/table contains to help guide your reader’s understanding of your findings.

In discussing your findings, you should not only discuss whether you reject or fail to reject the null, you should also discuss the implications of the rejection or failure to reject on your research question, linking your findings back to your introduction, literature review, and hypotheses. Do your findings differ from the findings of those whose work you reviewed in your literature review? If so, why? What new insights do your findings offer for those who are interested in your research question?

After discussing your findings and their implications, you should note any limitations of your work and/or methods. For example, did you make any assumptions in the course of your analysis? How tenable are those assumptions? Are you concerned about omitted variables bias? Selection bias? Were the data insufficient or biased in some way? Are there other shortcomings to your research methods? Finally, what further research should or could be done in this field to address your research question?

7 Author’s voice

As you write your paper, imagine a reader to whom you are communicating your research question, methods, and findings. You want to make this reader’s life easy by communicating your work as clearly and coherently as possible.

- Write your paper as a scholar communicating with other scholars. The tone should be formal.
- You do not need to use the royal ”we” if you are the sole author.
- It’s okay to use ”I”, though avoid ”I believe” as this just undermines whatever claim you’re trying to make.
- Use headings throughout your paper to help the reader navigate (e.g. Introduction, Literature Review, etc)
- Break long paragraphs apart and signal transitions between paragraphs with transitional words.
- Proofread.

8 Outline

Your research paper should follow a general outline that includes:

1. Title
2. Introduction

3. Literature review

4. Data and methods

5. Findings and discussion

6. Conclusion

7. References

8. Appendix (if necessary)

If you are considering submitting your paper to a journal or conference, you should try drafting and including an abstract as this is a great exercise in identifying and communicating the important aspects of your paper. Abstracts are generally 100 word (or less) summaries of the key features of your paper: the research question, data, methods, and findings.

As you pull your full research paper together, keep in mind that your research question may have shifted over the course of the semester. Therefore you may have to spend more time revising the work you produced earlier in the semester; save time for this. Before you turn your paper in, make sure that you have clearly labeled all your tables and figures, cited all sources (both in-text and in your references section), and written transparently about your methods as well as their limitations.