Experimental economics

Lecture 4: Research question and hypotheses

Matej Lorko
matej.lorko@euba.sk
Materials: www.lorko.sk/lectures

References:

- Weimann, J., & Brosig-Koch, J. (2019). Methods in experimental economics. Springer International Publishing. Chicago
- Jacquemet, N., & l'Haridon, O. (2018). Experimental economics. Cambridge University Press.
- Johnson, J. B., Reynolds, H. T., & Mycoff, J. D. (2015). Political science research methods. Cq Press.

Research paper in economics - basic structure

- Abstract: A brief summary of the whole paper
- Introduction: Pose an interesting question or problem and explain your motivation
- Literature Review: Survey the literature on your topic
- Methods/Data/Design: Describe your data, or **experimental design**
- Hypotheses / Theoretical predictions
- Results: Present and interpret your results with the help of graphs and charts
- Discussion: Critique your method and/or discuss any policy implications
- Conclusions: Summarize what you have done; pose questions for further research
- List of references

Finding a research question

- Economists view the world through the lens of efficiency, starting from the assumption that individuals behave rationally and focusing on the problem of allocating scarce resources. From this common analytical perspective, economists study a wide range of topics, involving the behavior of individuals, organizations and nations. The economic approach can be applied so broadly that choosing a topic to write on can be difficult. Indeed, once you start looking at the world through the eyes of an economist, almost anything can be analyzed in terms of choice under constraint.

- A good term paper will ask an interesting question and offer a plausible answer. It should be plausible in that it is (probably) true, but also not obviously or patently true; and it should be supportable in that it is subject to factual observation or logical demonstration.

- Though there is no one way to find a topic, thinking of the issues that interest you is a great place to begin. While the range of possible topics is large, there are some well-defined fields in economics, and your own interests are likely to fit into one of these.

- In addition to finding something that interests you, you will also need a project that can be done within the parameters of the assignment (for example, length, due date, access to research materials). If the topic doesn't interest you, you probably won't put in the effort needed to do a good job or ask the right questions along the way. On the other hand, a profoundly interesting topic may not be manageable given the time and other constraints that you face.

Finding a research question

- You should focus on a single, manageable question. In any case, it has to be:
  - Relevant - has to deal with some real (economic) issue
  - Important - the issue (e.g. inefficiency) it is dealing with has to be significant
  - Interesting - has to have an audience - there has to be somebody who should be interested in / benefitting from your findings
  - Testable - there has to be a way to answer your question
  - Novel - there has to be some novelty involved, so that you are contributing to the advances in the field
- As a rule of thumb, I think the question is potentially good if it is a “Yes/No + Why” question. That is, you can answer your research question with either “Yes” or “No”, and you can also show the mechanism of the answer - why is it a Yes OR why is it a No.

Imagine that you want to know more about why students in your class are late for lectures…

1. “The relationship between student life and class attendance.”
    - Not a great RQ… I mean… it is not even a question :)
2. “What factors influence the student class attendance?”
    - Also not great… even if you find a couple of factors, how can you be sure that these are all? Maybe there are thousands of more important ones you have not identified…
3. “Does the way of how students manage their time influence their class attendance?”
    - Better. You can answer with yes/no. But.. what is “time management”? Your variable is too vague, very unspecific.
4. “Do reminders have an effect on class attendance?”
    - Good one - for empirical research. You can say yes/no, and also have a mechanism - reminders as a counterstrategy for forgetting/procrastination etc.
    - However, if you want to also show causality, then your problem will be that the main variable of interest (setting reminders) is out of your control.
5. “Do teacher e-mail reminders have an effect on class attendance?”
    - If you want to do experiments, here is your research question. For example, you can send reminders to half of the class and do not send reminders for the rest and then count what percentage of each group shows up.

Research question sources

- Look for…
  - **Curiosity** - maybe you have read/heard about something and you are wondering -> *How does that work?*
  - **Ineffectiveness** - sometimes you may find yourself pissed about something that does not work very well.. -> *Why are they doing it that way? Is there a better way?*
  - **Searching for truth** - sometimes you may feel that somebody is just wrong… -> *I don't believe what they are saying! Is it really how they say it is?*
  - **Competing conclusions** - sometimes there are two ideas/theories that contradict each other… -> *Who/what is lying and who/what is telling the truth?*
  - **New theories** - sometimes you find an attractive theory making clear predictions, but nobody tested the theory yet… -> *Is that a good theory?*
  - **New applications of theories/insights**… -> D*oes it also work in my environment? Does it also apply here?*
  - **Papers that really need replication study** - maybe because of small sample sizes or obsolete methodology -> *Your hypotheses, my design/data - will your results still hold?*

Presentation of your research question

- When presenting your research question, you should give the answers to at least 3 questions to your audience:
    - What? - what is your topic/question?
    - So what? - why is it important? Why should we care?
    - Now what? - how do you plan to do your research and find the answer to your question?
- Don’t forget about main assumptions, concepts and theories in economics - the assumption that people respond to incentives, the assumption of rationality, the law of supply and demand, the problem of scarcity and opportunity costs, the marginal principle, the notion of spillovers or externalities…
- Also, when making economic arguments, don’t forget that economists prefer:
    - evidence on what people actually do, rather than on what they say
    - large numbers of observations
    - random samples

Literature review

- Once you have well-defined research question, it is time for a literature review - a short “essay” about previous research that is relevant to your research question - what has already been done? What has already been found? How does my research question connect to & expand the literature?

- Literature review will also help you by stopping you from doing things that are not worth your time (if it has already been done a million times and everybody knows what the results will be) and on the other hand, will help you find gaps that are worth digging into. Really good papers actually include ideas for future research/expansion in the discussion section.

- Finally, literature review will provide you good guidance on the methodology/design and analysis of data . If everybody in your field uses same methodology and statistical analysis techniques, there is a good chance that you should use what they use - because there is probably a reason why they use what they use.

- Please use as good sources as possible. That means, you should strive for citing quality peer-reviewed journals only. No wikipedia or “internet” sources.

- Things to try: EconLit/Google Scholar for paper search, Mendeley for citations and bibliography

So you have your research question…

- What do you do after figuring out your research topic (motivation), doing your literature review and narrowing down your research questions?

  - Propose a suitable explanation for the phenomena you are interested in (the “why” of your research question)

  - Formulate testable hypotheses

  - Define the concepts identified in the hypotheses (what exactly it is that you will measure and analyse)

Proposing explanations

- Once your research question is developed, you need to propose an explanation (e.g. identify a phenomena/ concept) that may provide you with the answer. Often you will need to identify more phenomena and specify how they are related.

- To help clarify relationships between phenomena, scientists refer to phenomena as variables and identify several types of variables.

- A phenomenon that we think will help us explain our observations or behavior is called an **independent variable**. Independent variables are thought to influence, affect, or cause some other phenomenon.

- A **dependent variable** is thought to be caused, to depend upon, or to be a function of an independent variable.

- Thus, if a researcher has hypothesized that acquiring more formal education will lead to increased income later on (in other words, that income may be explained by education), then years of formal education would be the independent variable, and income would be the dependent variable.

- As the word variable connotes, we expect the value of the concepts we identify as variables to vary or change. A concept that does not change in value is called a constant and cannot be used to investigate a relationship.

Formulating Hypotheses

- A hypothesis is an explicit statement that indicates how a researcher thinks phenomena of interest (variables) are related. It proposes a relationship that subsequently will be tested with empirical observations of the variables.

- A hypothesis is a guess (but of an educated nature) that indicates how an independent variable is thought to affect, influence, or alter a dependent variable.

- Since hypotheses are proposed relationships, they may turn out to be incorrect and not supported by the empirical evidence.

- For a hypothesis to be tested adequately and persuasively, it must be stated properly. It is important to start a research project with a clearly stated hypothesis because it provides the foundation for subsequent decisions and steps in the research process. A poorly formulated hypothesis often indicates confusion about the relationship to be tested or can lead to mistakes that will limit the value or the meaning of any findings.

- A good hypothesis has six characteristics: (1) it is an empirical statement, (2) it is stated as a generality, (3) it is plausible, (4) it is specific, (5) it is stated in a manner that corresponds to the way in which the researcher intends to test it, and (6) it is testable.

Good hypothesis is…

1. Empirical statement - open question, proposing a relationship that can be empirically observed. Not a normative statement (opinions, preferences).

2. General - it should propose a relationship pertaining to many occurrences of a phenomenon rather than just to one occurrence.

3. Plausible - there should be some logical reason for thinking that it might be confirmed.

4. Specific - it should not simply state that variables are associated; rather, it should indicate the direction of the expected relationship (positive - if X inceases then Y increases, or negative - if X increases then Y decreases) between two or more variables. Also, it must be clear what variables X and Y mean (it must be easy to say how to measure variables).

5. Testable - it must be possible and feasible to obtain data that will allow one to test the hypothesis.

Conclusion

- A research project must provide - to both the producer and the consumer of social scientific knowledge - the answers to these important questions:
  - What phenomenon is the researcher trying to understand and explain?
  - What explanation has the researcher proposed for the behavior or attributes in question?
  - What are the meanings of the concepts used in this explanation?
  - What specific hypothesis relating two or more variables will be tested?
  - What is the unit of analysis for the observations?

Measurement

- Before testing hypotheses, we must understand some issues involving the measurement of the concepts we have decided to investigate and how we record systematic observations using numerals or scores to create variables that represent the concepts for analysis.

- How researchers measure their concepts can have a significant impact on their findings; differences in measurement can lead to totally different conclusions.

- It is useful to think of arriving at the definition of the variables as being the last stage in the process of defining a concept precisely. We often begin with an abstract concept (such as democracy), then attempt to define it in a meaningful way, and finally decide in specific terms how we are going to measure it.

- At the end of this process, we hope to attain a definition that is sensible, close to our meaning of the concept, and exact in what it tells us about how to go about measuring the concept.

- To be useful in providing scientific explanations for behavior, measurements of phenomena must correspond closely to the original meaning of a researcher's concepts.

- They must also provide the researcher with enough information to make valuable comparisons and contrasts. Hence, the quality of measurements is judged in regard to both their accuracy and their precision.

Accuracy of measurement

- There are two major threats to the accuracy of measurements. Measures may be inaccurate because they are unreliable and/or because they are invalid.

- Reliability describes the consistency of results from a procedure or measure in repeated tests or trials. In the context of measurement, a reliable measure is one that produces the same result each time the measure is used. An unreliable measure is one that produces inconsistent results-sometimes higher, sometimes lower.

- The reliability of social science measures can be calculated in many different ways.

    - The test-retest method involves applying the same "test" to the same observations after a period of time and then comparing the results of the different measurements.

    - The alternative-form method of measuring reliability also involves measuring the same attribute more than once, but it uses two different measures of the same concept rather than the same measure.

    - The split-halves method of measuring reliability involves applying two measures of the same concept at the same time. The results of the two measures are then compared. This method avoids the problem that the concept being measured may change between measures.

- A valid measure is one that measures what it is supposed to measure. Unlike reliability, which depends on whether repeated applications of the same or equivalent measures yield the same result, validity refers to the degree of correspondence between the measure and the concept it is thought to measure.

Sampling

- Suppose we want to assess national level of support for some proposed government policy. Since it is impossible to interview everyone, a more practical approach is to select just a "few" members of the population for further investigation. This is where sampling comes in.

- A sample is any subset of units collected in some manner from a population. The sample size and how its members are chosen determine the quality (that is, the accuracy and reliability) of inferences about the whole population.

- A researcher's decision whether to collect data for a population or for a sample is usually made on practical grounds. The advantages of taking a sample are often savings in time and money. The disadvantage is that information based on a sample is usually less accurate or more subject to error than is information collected from a population.

- Once a sample has been gathered, features or characteristics of interest can be examined and measured. The attributes of most interest in empirical research are numerical or quantitative indicators such as percentages or averages. These measures, or sample statistics, as they are known - are used to approximate the corresponding population values, or parameters.

- In order to mitigate the sample bias, ideally each element in the total population should have a known probability of being included in the sample. This knowledge allows a researcher to calculate how accurately the sample reflects the population from which it is drawn.

What can be learned from samples

- Samples provide only estimates or approximations of population attributes. Occasionally these estimates may be exactly right, but most of the time, however, they will differ from the true value of the population parameter.

- When we report a sample statistic, we always assume there will be a margin of error, or a difference between the reported and actual values.

- Where does the loss of precision or accuracy come from? The answer is chance, or luck of the draw. If you flip a coin ten times, you probably won't get exactly five heads, even if the coin is fair or the probability of heads is one-half. Randomness seems to be an innate feature of nature, at least on the scale at which we observe it.

- Just as with our coin toss, a random sample of ten (or even much larger) is not likely to produce precisely the value of a corresponding population parameter. But if we follow proper procedures and certain assumptions have been met (for example, the sample is a simple random sample from an infinite population), a sample statistic approximates the numerical value of a population parameter.

Types of Data and Collection Techniques

- Researcher collects data on behavior by observing either the behavior itself (direct observation) or some physical trace of the behavior (indirect observation).

- Data collected through firsthand observation is an example of primary data, that is, data recorded and used by the researcher making the observations.

- Data from interviews or the written record can be primary data or secondary data - data used by a researcher who did not personally collect the data.

- Students will often find suitable data generated through interviews or the written record for free in publicly available data archives, but students wishing to use data generated through direct or indirect observation must usually rely on their own ability to make the observations.

- Structured and Unstructured Observation - In structured observation, the investigator looks for and systematically records the incidence of specific behaviors. The researcher will have decided, based on theory, the relevant behaviors before starting data collection.

- In unstructured observation, all behavior is considered relevant, at least at first, and recorded. Only later, upon reflection, will the investigator distinguish between important and trivial behavior.

Survey Research and Interviewing

- How to ensure validity and reliability of survey and interview data?
- Let R stand for the respondent and I for the interviewer:
  - The requested information must be available to R (that is, not forgotten or misunderstood).
  - R must know what is to I a relevant and appropriate response.
  - R must be motivated to provide I with the information.
  - R must know how to provide the information.
  - I must accurately record R's responses.
  - The responses must reflect R’s meanings and intentions, not I's.
  - Other users of the data must understand the questions and answers the same way R and I do.

Survey research

- A group of individuals respond to or fill out more or less standardized questionnaires. The questionnaires may take different forms to investigate different hypotheses, but they do not involve freewheeling or spontaneous conversations.

- Although surveys can be relatively quick and cheap mean to obtain data, the researcher needs to think carefully about:

- Completion rates - If the response rate is low, either because individuals cannot be reached or because they refuse to participate, the researchers' ability to make statistical inferences for the population being studied may be limited. Also, those who do participate may differ systematically from those who do not, creating other biases. Increasing the size of the survey sample to compensate for low response rates may only increase costs without alleviating the problem.

- Sample-population congruence - how well the sample subjects represent the population, is always a major concern. Here we are speaking of how well the individuals in a sample represent the population from which they are presumably drawn. Bias can enter either through the initial selection of respondents or through incomplete responses of those who agree to take part in the study.

- Questionnaire length - if a survey poses an inordinate number of questions or takes up too much of the respondents' time, the respondents may lose interest or start answering without much thought or care.

Response quality

- Response quality = the extent to which responses provide accurate and complete information. It is the key to making valid inferences.

- Response quality depends on several factors, including the respondents' motivations, their ability to understand and follow directions, their relationship with the interviewer and sponsoring organization, and, most important, the quality of the questions being asked.

- Engaging respondents - it is important to get off on a good footing by introducing yourself, your organization, your purpose, your appreciation of their time and .trouble, your nonpartisanship, your awareness of the importance of anonymity, and your willingness to share your findings.

- Since the whole point of survey research is to accurately measure people's attitudes, beliefs, and behavior by asking them questions, we need to spend time discussing good and bad questions.

- Good questions prompt accurate answers; bad questions provide inappropriate stimuli and result in unreliable or inaccurate responses. When writing questions, researchers should use objective and clear wording. Failure to do so may result in incomplete questionnaires and meaningless data for the researcher. The basic rule is this: the target subjects must be able to understand and in principle have access to the requested information.

- Certain types of questions make it difficult for respondents to provide reliable, accurate responses. These include double-barreled, ambiguous, and leading questions.

Closed-ended questions

- The main advantage of a closed-ended question is that it is easy to answer and takes little time. Another advantage is that answers are easy to compare, since all responses fall into a fixed number of predetermined categories. These advantages aid in the quick statistical analysis of data.

- With open-ended questions, by contrast, the researcher must read each answer, decide which answers are equivalent, decide how many categories or different types of answers to code, and assign codes before the data can be analyzed.

- Another advantage of closed-ended questions over open-ended ones is that respondents are usually willing to respond on personal or sensitive topics (for example, income, age, frequency of sexual activity, or political views) by choosing a category rather than stating the actual answer.

- Critics of closed-ended questions charge that they force a respondent to choose an answer category that may not accurately represent his or her position. Therefore, the response has less meaning and is less useful to the researcher.

- Also, closed-ended questions often are phrased so that a respondent must choose between two alternatives or state which one is preferred. This may result in an oversimplified and distorted picture of public opinion. A closed-ended question allowing respondents to pick more than one response (for example, with instructions to choose all responses that apply) may be more appropriate in some situations.

Open-ended questions

- Unstructured, free-response questions allow respondents to state what they know and think. They are not forced to choose between fixed responses that do not apply. Open-ended questions allow respondents to tell the researcher how they define a complex issue or concept.

- Disadvantage of the open-ended question is that respondents may respond too much or too little. Some may reply at great length about an issue - a time-consuming and costly problem for the researcher. On the other hand, if open-ended questions are included on mail surveys, some respondents with poor writing skills may not answer, which may bias responses.

- Furthermore, unstructured answers may be difficult to code, interpretations of answers may vary (affecting the reliability of data), and processing answers may become time-consuming and costly.

Question order

- The first several questions in a survey are usually designed to break the ice. They are general questions that are easy to answer.

- Complex, specific questions may cause respondents to terminate an interview or not complete a questionnaire because they think it will be too hard. Questions on personal or sensitive topics usually are left to the end.