Worksheet: Measuring Your Variables

Prepared by: Matthew E. Vanaman

09-19-2021

Table of Contents

Introduction

When many people hear the phrase "taking data", they imagine someone in lab coat extracting small amounts of liquid from a beaker. Few would think of social scientists as dealing with "data". But believe it or not, any data-driven work, including the social sciences, use data. Anyone studying *a thing* will not be able to understand that thing without measuring it in some way.

You might use this worksheet as a reference as you plan your project out. Alternatively, you might copy-and-paste it, and fill it out once for each variable in your study. You might even find yourself making changes to your variables upon attempting to fill this worksheet out - that is fine too. After all, planning a research project contains some trial and error. However you use them, I hope the steps in this worksheet will make the process of planning your project a bit more concrete and organized.

This checklist assumes that you know what variables you want to measure. Its goal is to help you determine how you will go about getting data of your variables, and does so by helping clarify the nature of the thing you are trying measure by breaker down the process into a series of smaller steps. Some people refer to this as the process of *operationalizing* your variables: you are taking an abstract idea of a variable and figuring out how you can represent or express that variable as a number.

I recommend this order for the steps:

1. Determine the scope (person- or aggregate-level) and observability (direct or indirect) of your variable.
2. Identify sources of data for your variables - or strategies for obtaining data yourself - based on their scope and observability.
3. IF you have more than one source of data for your variables, decide which source to choose based on how valid and reliable you judge the data to be.

The Glossary of Terms provides detailed definitions and examples of all the key terms used.

1. Determining Scope and Observability

What is the variable you want data for?

Which scope best matches the scope of the data you want?

☐ Person-level: I want data of individual people. Each data point (or row in my dataset) should be traceble to an individual person.

☐ Aggregate-level*: I want data of some larger entity, like an institution. Each data point (or row in my dataset) should be traceable to e.g., an institution, zip code, or something larger than an individual person.

*Beware the ecological fallacy.

Which level of observability best matches the observability of your variables?

☐ Direct: I want, and can get, data that represent my variable exactly. The number in the dataset maps directly onto the variable I am measuring.

☐ Indirect: I want data that serve as indirect data of something that cannot be directly measured. The number in the dataset does not map directly onto the variable I am measuring, but can indicate something important about my variable (for example, its presence or absence, or how high or low it likely is, etc.)

Which combination is your variable? Figure this out for each variable.

☐ Person-Level and Direct
☐ Person-Level and Indirect
☐ Aggregate-level and Direct
☐ Aggregate-level and Indirect

2. Finding Measures and data

I wish I could point you to a handbook for finding measures and data that meets the needs of every situation, but there is no such thing as far as I know. In general, you will have to do some sleuthing to find what you need.

The reason Step 1 is helpful (scope and observability) is because it helps you narrow down your potential sources of data. The figure below can help you get started. For example, if I determine that my variable is person-level and indirect (such as an attitude), I know that I will need an attitude measure distributed to individual people using a survey.

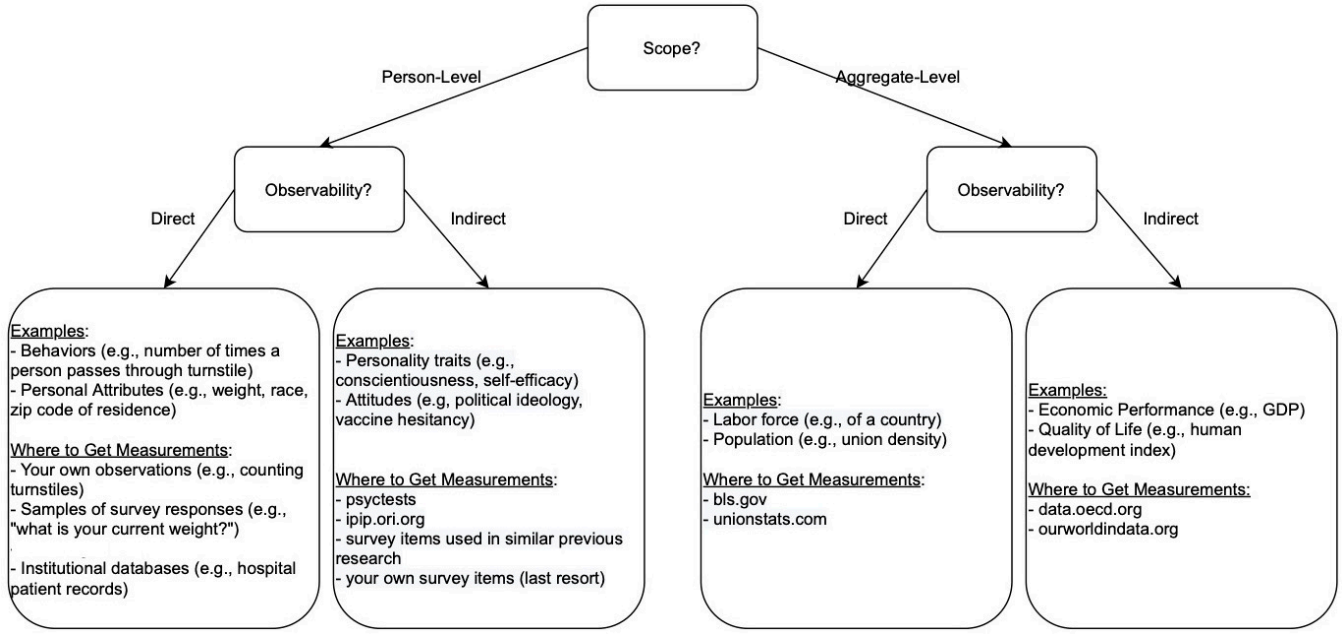

Figure 1: Flowchart for Finding Sources of Measures and data

The table below provides some sources organized by your type of variable. It is by no means an exhaustive list, but could get you started.

Table 1: Clickable Links to Example Sources for Measures and Existing data

		Scope	
		Person	Aggregate
Observability	Direct	PEW Research Polls	US's Public Data
		US Census	NYC Keeping Track
		APA Social Science Datasets	Business and Industry Data
	Indirect	International Personality Item Pool	OECD
		PsychTests (through GC Library)	Our World in Data
		APA's Psychological Datasets	ProPublica

Here are a few more tips for getting data for your variables:

1. **If you are measuring an indirect, person-level variable, do these steps in order:**
    1. Identify a previously-validated measure. "Validated" means that some science has gone into vetting the measure to prove that it measures a specific variable and ruled out that it measures something else. Usually, these are their own articles, completely dedicated to validating the measure itself (as opposed to answering another research question).
    2. If no previously-validated measure is available, see if someone else has made their own measure or a particular data source. In these cases, the variable being measured can be uncertain, but at least you can

match your own data to an existing precedent. The benefit of this is that different studies are at least comparable, because their data both measure the same thing (even if this thing is not exactly what the researchers wanted).

3. Make up completely brand-new measures or identify new data as a measure of you variables. For example, for an attitude measure, you might make up some questions to ask in a survey about how people feel about X, to which they can respond on a scale of 1 to 5 (strongly disagree to strongly agree), for example. Or you might choose to let median income of a neighborhood be a proxy for neighborhood wealth, even though wealth technically has more to do with non-income based characteristics (like whether they own land, what their network is, etc). Technically income and wealth are different, but they are highly correlated enough that income can be used as an indirect indicator of wealth in cases where wealth data are not available.

3. Evaluating Measures and data

Once you are looking for data, you might find that there is more than one way to get the data you need. In such cases, you can use reliability and validity to guide your choice.

See the Glossary of Terms for detailed definitions and explanations of reliability and validity.

Reliability

How to Assess Reliability

When choosing between measures (e.g., of attitudes or traits):

1. **With more than 3 questions**: in the validation article or previous article where the measure was created, look for Cronbach's alpha (represented with the alpha symbol α). Alpha ranges from 0 to 1, with 0 = "the responses are entirely due to measurement error" and 1 = "the responses are pure measures of the variable with no measurement error". Generally, a measure is said to have acceptable reliability when above 0.7, though this cutoff is arbitrary. When choosing measures, look to see whether previous researchers reported alpha for that measure. You can calculate it yourself on your own responses with any statistical software, including Microsoft Excel, if you choose to do that.
2. **With 3 or fewer items**: Cronbach's alpha does not work well with small numbers of items. Technically you can calculate alpha for only two items but that number will, perhaps ironically, not be a very reliable number. Instead, look at the correlations among the items. Generally, they should each be at 0.5 or higher, though again this is an arbitrary cutoff.
3. **For measures with 1 item**: you won't be able to quantify reliability from only one item. But, you can get a subjective impression. I once did a study using homeless participants, who are less likely to give reliable responses due to their circumstances (being stressed, distracted, fatigued, etc.). If such cases, you might note that your data are a bit less reliable.

When choosing between existing sources of data (e.g., online databases): you won't be able to quantify reliability for these either, but you will be able to get a subjective impression based on the source of the data. In the description of the data and how they were acquired and/or calculated, look for the following:

- Is there any reason to believe that data collection might be vulnerable to errors in recording or processing? If so, could be worth noting as potential source of unreliability.
- Whether or to what extent the data contains imputed values (filling in missing values with educated guesses).
- Whether or to what extent there is a lot of missing data, as this could indicate something about the way data were entered, whether participants were paying attention as they filled out the survey, etc.

Reliability Assessment

To what extent are the data free of error? More specifically (answer the question that is relevant to you). Note that this may be difficult to answer:

You are using a measure to collect your own data: how well do responses to the individual questions on the measure with each other (not applicable if the measure is only one question)?

☐ Not well at all ☐ Somewhat well ☐ Well ☐ Extremely well

You are using previously-gathered data: look to see how the data were collected. Based on that, how precise (i.e., accurate, not too many mistakes or guesses) was the procedure for collecting the data?

☐ Not well at precise ☐ Somewhat precise ☐ Precise ☐ Extremely precise

Validity

How to Assess Validity

Face validity:

- If using a measure to obtain your own data, such as a survey, look to see whether the wording in the relevant questions on the survey contain superficial descriptions of the definition of your variable. For example, if I define depression as "a mood or emotional state marked by feelings of low self-worth or guilt and a reduced ability to enjoy life", a face-valid survey question could be "do you feel like you have low self-worth?", on a scale of 1 to 5 with the responses 1 = never, 5 = all the time.
- For existing data coming from e.g., an online database, look to see whether the *description* of the data superficially seem to describe your variable, based on your quick impression of the description and how well it seems to match the definition of you variable. For example, if I want data that measures quality of life, I might choose Human Development Index over Gross-Domestic Product, since human development index is defined by things like access to education, healthcare, and average length of life (among others), whereas gross domestic product is defined by the total value of goods produced and services provided in a country during one year. Human development index seems superficially more relevant to quality of life, so has better face-validity.

Content validity:

- If using a measure to obtain your own data, such as a survey, look to see whether the wording in the relevant questions on the survey contain theoretical important aspects of your variable, even if they aren't mentioned in the definition. For example, if I define depression as "a mood or emotional state marked by feelings of low self-worth or guilt and a reduced ability to enjoy life", a content-valid survey question could be "do you have difficulty getting out of bed?". Although depression is not defined by difficulty getting out of bed, it follows from the definition that people who are depressed will often have difficulty getting out of bed.
- For existing data coming from e.g., an online database, look to see whether the *description* of the data contains the theoretically important key concepts of your variable, even if they are not inherent to the definition of your variable. In either case, rely not so much on your quick, subjective impression, but rather on your cerebral, knowledge-based expertise on what the definition of the variable is and what its various key concepts are. Then, look to see whether these key concepts are represented in the data. Note that a measurement can lack face validity, yet still have good content validity; usually they go hand in hand, but not always.

See the Glossary for detailed definitions and explanations of validity.

Validity Assessment

Face validity: How well do the measures/data *superficially appear* to capture your variable (as opposed to some other variable)?

☐ Not well at all ☐ Somewhat well ☐ Well ☐ Extremely well

Content validity: Based on your knowledge of your variable, well do the measures/data capture the full range of key concepts relating to your variable (as opposed to some other variable)?

☐ Not well at all ☐ Somewhat well ☐ Well ☐ Extremely well

Some Tips for Reporting

When you write up your project, you will want to give an argument for why you chose the measures and/or data that you did.

- You might report reliability, including Cronbach's alpha (α) as it has been reported in previous research as well as your own calculation from your own data.
- Add a few sentences about face and content validity, why you think they are good, along with an example question (if using a measure) and/or description of the how the data were obtained and/or calculated (if using existing data).
  - For content validity, if there is previous research describing the definition or conceptualization of the variable you are measuring, you can talk briefly about how your measure(ments) logically follow from the definition.
- In general, you'd want discussion of the strengths of reliability and validity in the methods section, and discussions of lack of reliability and validity in the limitations section. If you think reliability and/or validity could be weakness, add "(but see limitations section for discussion of the weaknesses of reliability/validity)" or something along those lines to the methods section, then talk about it in more detail in the limitations section.

The figure below does a very nice job of describing what to cover when writing up your methods section.

Question	**Information to transparently report**
What is your construct?	Define construct Describe theories & research supporting the construct
How do you operationalize your construct?	Describe measure and administration procedure Match measure(s) to construct(s)
Why do you select your measure?	Justify measure selection Report existing validity evidence
How do you quantify your measure?	Describe item response coding & transformation Describe items/stimuli per score Describe calculation of score Describe all conducted psychometric analyses
Do you modify the scale? If so, how and why?	Describe any modifications Describe if modifications occurred before or after data collection Provide justification for modifications
Do you create the scale on the fly?	Justify why you do not use an existing scale Report all measurement details listed above Describe all available validity evidence and if there is no evidence, report that

Figure 2: Tips for transparent reporting of measurement.

Glossary of Terms

- **Variable:** A variable is the thing you want to study. A variable "varies", meaning it changes over time, differs between people/locations/countries/etc. Gentrification is a variable, as is height, eye color, a personality trait, or an economy.
- **Measurement:** A measurement is a number assigned to an instance of the variable. Otherwise known as a "data point". Every data point is just a number that has been assigned to a variable. To use the height example, the number on a tape measure, in inches, would be a measurement of height.
- **Measure:** An instrument used to obtain a measurement. In social science research, this usually is synonymous with "questionnaire" or "survey".
- **Scope:** The "level of analysis" of your data. Although there are many potential scopes, the most important distinction is between *person* and *aggregate* levels.
  - **Person-level scope:** data are of individual people, so claims are about individual people. In these kinds of data, each individual measurement can be traced to an individual person. Used when you are interested in describing people, such as a sample of participants from a neighborhood.
    - Example: You want to see how a people's wealth relates to their subjective well-being. In this case, you'd measure *each person's* wealth and subjective well-being.
  - **Aggregate-level scope:** you want data of some large entity that is beyond the scope of a person. In aggregate-level data, each measurement is traced to an institution, country, economy, etc., but are themselves averages of groups of people in some way. You either don't have access to, or do not want, data of individual people.
    - Example: You want to see whether there is a relationship between a country's wealth and well-being. In this case, you'd get a measurement of wealth (such as GDP) and well-being (such as HMI, or Human Development Index). Beware the ecological fallacy.
- **Observability:** whether a variable is directly or indirectly observable.
  - **Direct:** Your data represent your variable exactly. Each measurement "is" the variable. The variable can be measured without relying on "proxies" or go-between variables.
    - Example: Subway ridership, as measured by the number of people who used the subway within the period of interest.
  - **Indirect:** Your data do not represent your variable exactly, either because of practical or conceptual constraint.
    - A practical constraint could be that the means to measure a variable are not available. For example, you want to compare well-being in wealthy and less wealthy neighborhoods, but do not have the ability to directly measure and quantify the wealth of people who live in each neighborhood. You decide to use zip code of known wealthy and less wealthy neighborhoods as an indirect measure of wealth.
    - A conceptual constraint, on the other hand, is when the variable itself is not a tangible "thing", but can be represented by smaller things. Personality traits are a good example; self-efficacy is the general tendency for a person to have a "can-do" attitude toward challenges in their life. Self-efficacy, though, is not something you can pick up and look at; it's probably a psychological or biological mechanism, which we have to measure based on its indicators (such as to what extent a person endorsements statements like "I feel like I can solve most problems in my life").
- **Ecological fallacy:** The ecological fallacy is the mistake of using aggregate-level data to make the same kind of research claims that come from person-level data.
  - Example: When I have person-level data for subjective well-being and individual wealth, and find that they are positively correlated, it would be OK to conclude that well-being and wealth are correlated. On the other hand, if I have person-level data for subjective well-being and measurement's of their home countries' wealth, and find that they are positively correlated, it would NOT be OK to conclude that well-being and wealth are correlated. This is because, unlike when both data are person-level, I would have no way to tell whether the people who are high in well-being are the same people who are high in

wealth. There is no reason why people from a wealthier country could not have *lower* subjective well-being, even if the correlation between well-being and wealth is positive at the person level (such as if the top 5% generated most of country's wealth and were also miserable).

- **Validity:** Validity is the idea that a set of data should represent, or "capture" the correct variable - that is, higher values reflect higher levels of the variable and lower scores reflect lower levels of the variable. Valid data do not conflate data of the intended variable with data of a different variable.
  - **Content Validity:** A measurement is said to have content validity if it thoroughly captures the full range of conceptual content relevant to the variable.
    - Example: Sometimes researchers use GDP (Gross Domestic Product) as a measure of well-being of a country's citizens, but some argue, as in this article, that the process used to calculate GDP does not include many of the factors that lead to genuine well-being, such as "the health of our children, the quality of their education, or the joy of their play". GDP could be said to lack content validity as a measure of well-being. However, it has high content validity as a measure of economic output, as it captures a lot of the relevant factors that indicate how well an economy is doing (though some economists disagree about that too).
  - **Face Validity:** A measurement is said to have face validity if it "looks" like your variable. For measures (e.g., questionnaires), the survey questions should contain wording that reflects the key words and phrases describing the variable.
    - Example: Self-efficacy is a person's general tendency to feel like they can deal and cope with challenges in life. If I want data of my participant's self-efficacy within a health context, I should choose a measure of self-efficacy for which the wording looks and sounds like "health-based self-efficacy", as much as possible. A survey question that asks "In general, I feel like I can overcome the challenges I face" would have good face validity. "In general, I feel like I can overcome the challenges *to my health* that I face" is better.
- **Reliability:** Reliability refers to the overall error in the data. Error should be minimal, so that individual data can be considered maximally trustworthy. For a measure that you are using to obtain data of something (e.g., a personality questionnaire), a person's response to one question should correlate with their response to another. In cases where data were already gathered or calculated, reliability refers to how precise the measurement procedure was (e.g., how many mistakes were make in recording the data), or how much the calculation relies on "rough guesses" or "imputations" (substituting missing data with educated guesses).
  - Example: census data are considered highly reliable because the government uses a lot of quality control to minimize mistakes. The actually send people out to make sure citizens are filling out the census accurately. On the other hand, as this article points out, GDP (Gross Domestic Product) measure of economic output contains a lot of guesses and approximations, and so is considered an unreliable measurement by some (this article argues that GNP, or Gross National Product, is more reliable and should be used instead).