Measuring

Pablo Balán

GOV 1263, March 29 2018
Housekeeping

- **Assignment #3**: Due 9 hours and 7 minutes ago
- **Assignment #4**: Next Thursday at 9 am
The menu

- How do we measure concepts?
- What makes a good indicator?
- Where do we get our data from?
- Key principles of survey design
- Non-survey measurement instruments
CONCEPTS AND INDICATORS
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>A change or impact caused by the program we are evaluating</td>
<td>Increase or decrease in women’s empowerment, child health, corruption</td>
</tr>
<tr>
<td>Indicator</td>
<td>An observable signal used to measure outcomes</td>
<td>The number of women who spoke at a meeting, child arm circumference</td>
</tr>
<tr>
<td>Instrument</td>
<td>The tool we use to measure indicators</td>
<td>A survey question, achievement test, direct observation record</td>
</tr>
<tr>
<td>Variable</td>
<td>The numeric values of indicators</td>
<td>Self-evident</td>
</tr>
<tr>
<td>Respondent</td>
<td>The person or group of people we interview, test, or observe to measure</td>
<td>Individuals; their teachers, colleagues, or family</td>
</tr>
<tr>
<td></td>
<td>the indicators</td>
<td></td>
</tr>
</tbody>
</table>
Social science concepts are messy

- In **natural science** (real science) concepts are the indicators (E.g.: temperature is whatever the thermometer says)
- **Social science** concepts are complex and multidimensional (a mess)
- Examples:
  - Democracy
  - State capacity
  - Social capital
  - Civil war
- Some concepts are contested on normative grounds, some are just hard to measure
Indicators

- Indicators: Observable signals of these conceptual changes
- Usually no consensus on how to measure concepts
- In some discipline, there is more consensus (Examples: psychological tests, malnutrition)
- Look at the literature for most current measurement standards
Example from lecture: Women as Policymakers

Concept: “Women empowerment”
What are the observable consequences of political quotas for women?

1. Even after approved by the Supreme Court, quotas had to be passed in each state (look at records)
2. More women actually hold chairs in village councils
3. More women participate by expressing their views and submitting complaints in assembly meetings
4. Investments in public goods now reflects the preferences of women
5. The quality of these public goods actually increases
Good indicators are:

1. **Observable**: A behavior or state that can be observed in the real world. Happiness or learning are not indicators.

2. **Feasible**: Cost can be prohibitive.

3. **Detectable**: Our measurement instruments and power should be able to detect them.

4. **Reliable**: People forget or deliberately misreport. SDB
Create an index

- When you have many outcomes measuring many dimensions of the same concept
- Avoids fishing for significant results by “picking” outcomes that work
- Increases power

**Standardized index:** Weighted average of the z-score of each item

**Cronbach’s alpha:** \[ \alpha = \frac{K \bar{c}}{\bar{\nu} + (K - 1)\bar{c}} \]

- \( K \) is the number of items
- \( \bar{c} \) is the average of all covariances between items
- \( \bar{\nu} \) is the average variance of each item
Levels of measurement

We can measure stuff at different levels:

1. **Categorical:** typologies
2. **Ordinal:** can rank alternatives but don’t know the distance between them (dichotomous variables or 5 point scales)
3. **Interval:** We know the degree of difference (temperature in Celsius, dates, coordinates)
4. **Ratio:** Scale has meaningful, non-arbitrary zero (mass, length, duration, temperature in Kelvin scale)
Units of measurement

“As Fuck” is my favorite unit of measurement
TYPES OF DATA
Types of Data

- **Administrative data** (if you work in countries like Sweden or Norway, or Brazil)
- **Produce your own data** (if you work in countries like Congo)
Administrative Data: Examples

- Increase tax compliance by sending messages to taxpayers (paying taxes is an obligation/if you don’t pay your taxes, you pay a fine). Need information about who’s paying and who’s not paying
- Ferraz and Finan on corruption audits. Electoral data to assess the effect of audits
- Fujiwara on the effect of electronic voting on health outcomes
- Horacio’s Mexico experiment: Official data from audits to get the malfeasance figures included in the information treatment leaflets
- **Problem:** Sometimes not reliable in developing countries. Countries like Argentina manipulate their statistics!
- Take advantage of the fact that administrative data is doctored for your own experiment. Examples: Callen and Long and Ichino and Schündeln
SURVEY INSTRUMENTS
Endline, for sure
Baseline, sometimes

Reasons to run a baseline survey:
1. Increases power (more obs)
2. If data is at the individual level
3. To collect data on variables on which to block
4. Subgroup analysis
   - Example: Divide people by income and specify income-specific effects
Surveys: When to do the endline

- When to do the endline survey
- Some treatments don’t have an immediate effect. We need to wait for effects to show up
- If we wait too long effects can go away, or attrition might be an issue
Survey logistics

- **Online Surveys**
  - Amazon Mechanical Turk (mTurk)
  - Harvard Digital Labs for the Social Sciences (DLABSS)

- **Field Surveys**
  - Write a form using XLSform
  - Use in tablet with ODK collect
Persuading respondents

Hey, girl,
Take the survey
Persuading respondents

NO, I WOULD NOT LIKE TO TAKE YOUR BRIEF SURVEY
Beware of biases in measurement

- Sometimes people give some answer due to social pressure: Social desirability bias
- Or they guess what the experimenter wants to hear: Experimental demand effects
- Or their behavior changes just because they are observed: Hawthorne effect
NON-SURVEY INSTRUMENTS
Non-survey instruments

1. Direct observation
2. Physical tests
3. Biomarkers
4. Mechanical tracking devices
5. Purchase decisions (revealed preference)
6. Games
7. Implicit Association Tests/Vignettes
8. List experiments
9. Social networks
Direct observation: Examples

- Callen and Long: Take a picture of the vote counts

Panel A. Election returns form at the polling center  
Panel B. Election returns form at the national aggregation center

Figure 2. Election Returns Form for the Same Polling Center before and after Aggregation
Physical tests: Examples

Olken: Measuring missing materials

B. Measuring Quantities of Materials

In the engineering survey, an engineer and an assistant conducted a detailed physical assessment of all physical infrastructure built by the project in order to obtain an estimate of the quantity of materials used. In the standard road design, known as a *telford* road, the road consists of three types of materials: a base of sand, a layer of large (10–15 centimeters) rocks, and a top layer of gravel to provide a smooth running surface. To estimate the quantity of each of these materials, the engineers dug 10 40 centimeter × 40 centimeter core samples at randomly selected locations on the road. By combining the measurements of the volume of each material per square meter of road with measurements...
List experiments as a solution to SDB

- SDB especially prevalent when measuring: Racial prejudice, clientelism, sexual behavior, illegal behavior
- A solution: list experiments
- Divide respondents in T and C
- Control group receives a list of options
- Treatment group receives the same list, plus the sensitive option
- Don’t ask which but how many
- The result of the difference in means is the causal effect
List experiments: Examples

- “I’m going to hand you a card that mentions various activities, and I would like for you to tell me if they were carried out by candidates or activists during the last electoral campaign. Please, do not tell me which ones, only **HOW MANY**”
- Control group gets: they put up campaign posters or signs in your neighborhood/city; they visited your home; they placed campaign advertisements on television or radio; they threatened you to vote for them.
- Treatment group gets: all these plus “they gave you a gift or did you a favor”
Measuring norms is more challenging. Researchers often use surveys in which they effectively ask respondents if they possess a certain norm (e.g., “Do you think people are generally trustworthy?”). Program staff typically stress the importance of prosocial norms in their interactions with treated communities and, of course, they do not operate in the control communities at all. Thus respondents in treated communities may subconsciously feel more compelled to give the “right” answer to such questions than do control-community members, who may not even know what the “right” answer is since they have not received this training. Behavioral measures are designed to elicit responses based on the subjects’ true preferences and beliefs, which is why Casey, Glennerster, and Miguel (2012) and Fearon, Humphreys, and Weinstein (2009) used them.
Games

- **Dictator game**: Player A decides how to split X dollars between himself and Player B. Player B does nothing. Rational answer is to give zero. Player A’s offer is a measure of **altruism**.

- **Ultimatum game**: Player A decides how to split X dollars. Player B can accept or refuse. If Player B refuses, they both get nothing. Player B’s response measures **costly punishment**.

- **Public goods game**: **obligation**
Games: Example from lecture

- Avdeenko and Gilligan: Willingness to share with the needy

**Willingness to Share with the Needy.** We measured subjects’ willingness to share with the needy with a simple alteration of the standard dictator game. Subjects were given 3 Sudanese pounds in six half-pound coins. They were asked to decide how much, if anything, of that amount to donate to a local needy family. The subjects were not told the name of the needy family to protect the family’s privacy and avoid any differences between subjects in their affinity with the needy family. Each subject was called individually to the games area. The six half-pound coins were set side by side on sheet of paper with a line drawn across the middle. The subjects were instructed to push the amount they wished to donate to the needy family across the line on the paper and they were told that any remaining amount would be added to the lump sum that they received at the end of the session. The average amount sent was roughly 1.47 pounds, about half the pot.
Games: Example from lecture

- Avdeenko and Gilligan: Trust

- Player A’s behavior: trust
- Player B’s behavior: trustworthiness
Implicit Association Tests (IAT)

- Widely used in social psychology to measure implicit bias
- The association between two concepts is stronger, the faster you categorize together
- Example: black/white and pleasant/unpleasant
- A word appears in the middle and the respondent has to categorize it
- A participant will be able to categorize more quickly when White and Pleasant are paired together than when Black and Pleasant are paired if he or she has more positive associations with White people than with Black people
IAT Example: Powerful Women

Figure A1: Female/Male Politician Good/Bad Stereotypic Block Example
Social networks: Example from lecture

- Avdeenko and Gilligan on social capital in Sudan
- Ask each respondent about their relationship with all the other respondents (requires full list)

| TABLE 2. Summary Statistics of Key Network Variables |
|---------------------------------|---------|---------|
|                                  | (1)     | (2)     |
|                                  | Relative to Group Size | Absolute |
|                                  | mean    | sd      | mean    | sd    | max |
| Basic Social Relationships       |         |         |
| Are you family members with ...? | 0.22    | 0.18    | 4.46    | 3.53  | 16  |
| Are you neighbors with ...?      | 0.16    | 0.10    | 3.24    | 2.08  | 10  |
| Do you get together socially with ...? | 0.13    | 0.11    | 2.53    | 2.10  | 11  |
| Do you attend the same mosque with ...? | 0.39    | 0.38    | 7.49    | 7.21  | 19  |
| Economic Relationships            |         |         |
| Do you buy or sell products or services with ...? | 0.09    | 0.19    | 1.90    | 4.00  | 23  |
| Are you employed at the same farm or shop with ...? | 0.01    | 0.03    | 0.18    | 0.66  | 4   |
| Do you work for ...?              | 0.00    | 0.01    | 0.02    | 0.16  | 2   |
| Voluntary Groups                  |         |         |
| Are you members of the same producers group with ...? | 0.04    | 0.15    | 0.70    | 2.78  | 15  |
| Do you attend PTA meetings with ...? | 0.03    | 0.08    | 0.52    | 1.42  | 7   |
| Are you members of the same women's group with ...? | 0.03    | 0.11    | 0.56    | 2.32  | 13  |
| Favor Exchange Relationships      |         |         |
| In the last year have you sought advice about an important personal matter from ...? | 0.06    | 0.09    | 1.10    | 1.95  | 22  |
| In the last year has ... watched your children for a short period of time? | 0.01    | 0.04    | 0.28    | 0.74  | 4   |
| Trust-based Groups                |         |         |
| Are you members of the same revolving credit group with ...? | 0.07    | 0.18    | 1.36    | 3.63  | 14  |
| Do you exchange labor with ...?   | 0.09    | 0.25    | 1.73    | 4.54  | 17  |

Note: Each subject was asked whether he or she is in the specified relationship with each other subject in his/her village subject pool. We crosschecked reported relationships with the other subject in the reported relationship. In the two columns under (1) Relative to Group Size we report the percentage of the subject pool that shared the specified link, i.e., we divide each subjects' total number of relations in the specified category by the number of subjects in the their village and report the average of the values across the entire subject pool. The number of subjects varied across villages due to a small amount of subject attrition. The three columns under (2) Absolute report the mean, standard deviation, and maximum number of relationships in each category across the entire subject pool.