PAI 705

McPeak

Lecture 1

Epistemology – the science of knowing.

Methodology – a subfield of epistemology, the science of finding out.

How do we know what we know?

How much is direct experience?

How much is passed on?

Do you know the dark side of the moon is cold?

Do you know they speak Portuguese in Guinea Bissau?

Agreement Reality. The things we 'know' as part and parcel of the culture we share with those around us.

We agree that since I am up here with chalk, I am the instructor.

We agree that you need to be on campus to take this class.

In what ways does this help?

In what ways does it hinder?

Role of Tradition in establishing agreement reality; don't necessarily have to experience to have knowledge.

Role of Authority in establishing agreement reality; experts stating assessments about what is a fact.

Agreement reality is not the same as proven reality, which is part of what we try to test in social science inquiry.

Problems in Social Science inquiry.

Inaccurate observation / lack of attention to specific detail. We need to be deliberately paying attention to particular facts and recording them correctly. Problem of recall of a specific item if we were not paying attention to this item.

- Recall example in the Senegal and Mali data

Overgeneralization. We see one case and we think it applies to all cases.

Selective Observation. We see what we expect to see. We dismiss contrary evidence to confirm the version of reality we choose to see rather than allow the contrary evidence to refute our favored interpretation.

Illogical reasoning. The exception proves the rule. Dismiss evidence that refutes one interpretation as exceptions.

The foundation of social science research.

Theory, the logical foundation that provides systematic explanations.

Data collection is the observational aspect.

Social theory in this context has to do with understanding what is (positivist), not what should be (normative).

That is the theme of the Friedman article; normative compared to positive approaches.

Understanding what “is” as compared to understanding “what should be”.

Social science also had a normative dimension. What is economically just? What is the goal of public education? In terms of your program here, "what should be" is drawing on your Management class: what is the objective of our project? What are the indicators? What is your theory of change?

To contrast things, we have to first agree on measurable criteria with which we will evaluate. That is the monitoring indicators approach we will develop in this course.

Looking for patterns of regularity; average overall behavior that may have exceptions, but describes the overall tendency. That is our research part of the overall approach.

To define what is meant by "overall tendency" is where you draw on the statistics training.

Then you take the findings from your research and go back to the management part. How can we take these findings from the research treated by the statistics to do a better job in realizing our objectives?

In doing analysis and understanding behavior we need to understand the context. What are the formal rules and regulations? Where can you park your car and on which side of the street today? Can you eat food in this classroom?

Informal norms and customs. Which side of the stairs is for going up and which side is for going down? Which side of the sidewalk for east to west crossing of the quad?

A critique of social science is that when we identify a regularity, that regularity is trivial. "Female headed households in rural Senegal are poorer than male headed households". I have statistical support of what seems obvious.

However, what makes things interesting is that which is understood as true might not hold up to investigation. “Actually, they are not statistically poorer due to the role of male labor outmigration and remittances.”

That which seems trivial *ex post* was not understood *ex ante*. “Fulani female headed households are poorer but Wolof are not due to differing land inheritance customs that vary across ethnic groups.”

Figuring out which is the exception and which is the general pattern is not always easy, especially if not treated with some systematic design. We get systematic design by adopting the formal structure of probability, and the law of large numbers. What is the general tendency and what is an outlier (and what does the outlier tell us).

An example from my work:

Table 2: Average herd size and income measures by sub-group

Group	Herd Size TLU	Total Income per capita per day	Cash Income as % of Total Income	Total Income variability (cv)
1) Left out	7.3	$0.20	29%	1.32
2) Moving From	7.2	$0.27	46%	0.90
3) Staying With	23.7	$0.34	21%	0.82
4) Combining	26.0	$0.46	35%	0.63
Significant difference in means by groups, t-statistics	t12 , t13 ***, t14 ***, t23 ***, t24 ***, t34 ,	t12 , t13 , t14 **, t23 , t24 ***, t34 ,	t12 ***, t13 **, t14 *, t23 ***, t24 ***, t34 ***,	t12 **, t13 ***, t14 ***, t23 , t24 ***, t34 ***,

Observation and measurement.

The knowledge by people we are observing that they are being observed may lead to changed behavior.

- The observed behavior may not be the 'natural' behavior that would have happened in the absence of observation.
- It may be a response to being observed.
- They might be performing for us so we are recording their construct of what they think we want to see/ they want us to see, not their natural behavior.
- The "Hawthorne Effect". Observing workers in a plant in Illinois in the 20s and 30s. They changed working hours, break times, lighting in the plant. Worker productivity went up when these changes were made. Eventually they figured out the productivity went up because people were aware they were being observed, not due to the changes in the working environment.
- Hard to observe behavior if behavior changes in response to observation.
- People perform.
- Note challenge to informed consent. There is an inherent conflict. I have to tell you the objectives of my study and that I am doing a study to start collecting observational data.

Problem of inference. Example of insurance and behaving to meet the targeting criterion?

What you see is also filtered against what you are prepared to see:

http://www.jstor.org/stable/665280?seq=1#page_scan_tab_contents

And ‘truth’ is an elusive target.

Unidentified participant: Mr. Faulkner—

William Faulkner: Yes, sir.

Unidentified participant: In *Absalom, Absalom!* is any one of the people who talks about Sutpen have the right view, or is it more or less a case of thirteen ways of looking at a blackbird with none of them [getting it] right?

William Faulkner: That's it exactly. I think that—that no one individual can—can look at truth. It—it—it blinds you. You look at it, and—and you—you see one phase of it. Someone else looks at it and sees a slightly awry phase of it, but taken all together, the truth is—is in what they saw, though nobody saw the truth intact. So—so these are—are true as far as—as Miss Rosa and as Quentin saw it. Quentin's father saw what—what he believed was truth. That was all he saw. But the old man was—was himself a little too big for—for people no greater in stature than Quentin and Miss Rosa and Mr. Compson to see all at once. It would've taken, probably, a wiser or more tolerant or more sensitive or more thoughtful person to see him as he was. It was, as you say, thirteen ways looking at a blackbird. But the truth, I would like to think, comes out, that when the reader has read all these thirteen different ways of looking at the blackbird, the reader has his own fourteenth image of that blackbird, which I would like to think is the true one.

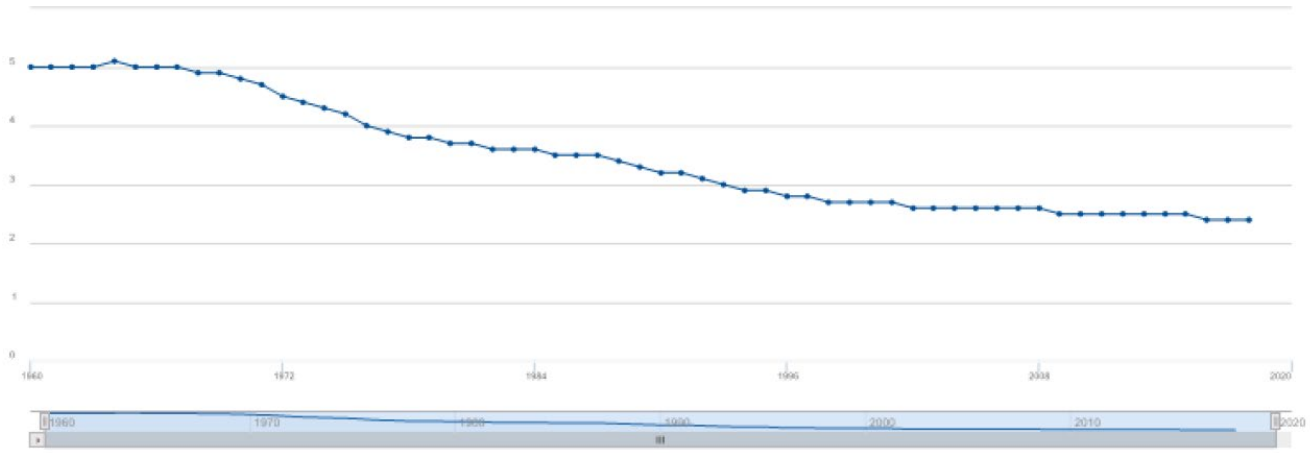
http://faulkner.lib.virginia.edu/display/wfaudio29_1

Social science is study of the aggregate, rather than the individual.

We are looking for collective regularities.

Individual decisions, but in aggregate, they create aggregate facts.

Total Fertility Rate (TFR). TFR is the average number of children a woman will give birth to in her reproductive years conditional upon her birth cohort.



World Development Indicators Online, accessed February 8, 2021

Total Fertility Rate for example, that changes over time but is not a collective decision, but a characteristic of individual decisions aggregated to the collective level. Not an outcome of social planning (though may reflect social programs and policies).

The goal is not to understand an individual person, but 'to understand the systems in which people operate, the systems that explain why people do what they do'.

Variable. Sets of attributes that can have different values.

Attributes. Characteristics of people or things.

Variable	Attributes
Age	Number of years since birth
Gender	Female, Male, Other
Occupation	Professor, Lawyer, Mayor
Social Class	Upper, Middle, Lower
Height	Value in centimeters

Attributes are agreed upon concepts that sort complicated reality into conceptual 'bins'. The bins are constructs, choices. We have to figure out what to do with a tall transgender lawyer who makes candy on the weekends and self-identifies as working class. Redefine bins? Add bins?

Are the attributes mutually exclusive?

Are the attributes exhaustive?

Above, we used constructed groupings to be mutually exclusive and exhaustive:

Group
1) Left out- below median herd, below median cash income
2) Moving From – below median herd, above median cash income
3) Staying With – above median herd, below median cash income
4) Combining- above median herd, above median cash income.

So sorting cases/ individual observations into bins is part of the art, or our first step from complicated reality to our artificial constructed version of reality used for analysis.

Discrete Choice Experiment on Access to PrEP. Variables and attributes.

Table 2				
PrEP Access Variables	**Baseline Option**	**Alternative 1**	**Alternative 2**	**Alternative 3**
Total out of pocket costs	Free	$20 per month	$50 per month	$200 per month
PrEP insurance coverage	No insurance coverage is available	PrEP is covered by your current insurance	You need to get an extra insurance plan that covers PrEP	You need to get a new plan that covers PrEP
Insurance privacy	Your insurance information is completely private	Parents, spouse, or employer might know you are on PrEP;		
PrEP timeliness	Start PrEP at first appointment	Takes 1 week to be cleared before starting PrEP	Takes 4 weeks to be cleared before starting PrEP	
PrEP location	Lab tests every 3 months; see your provider in the office	Lab tests every 3 months; communicate with your provider online		

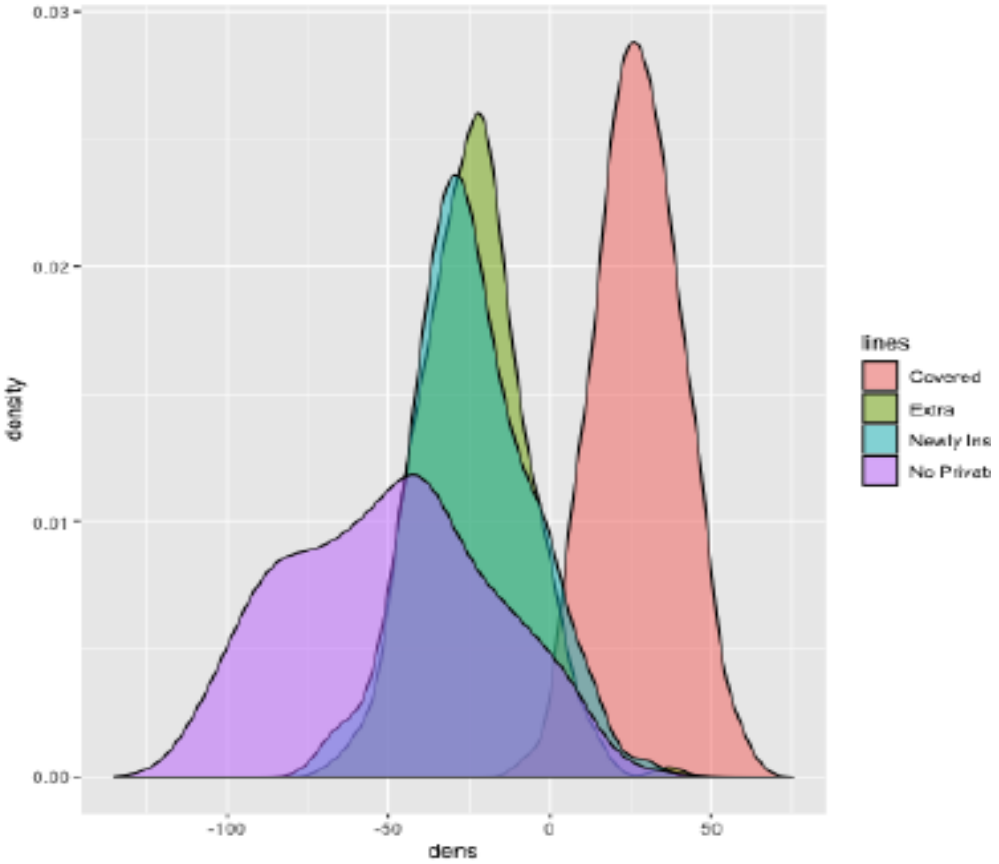


Figure 1: Densities of posterior distributions of mWTPs for insurance effects

Another issue we confront is the relationship between variables; some attribute values for one variable occur frequently with some attribute values for another variable. When looking at two variables, we might see some kind of pattern in correlation. The variables move together in some sense.

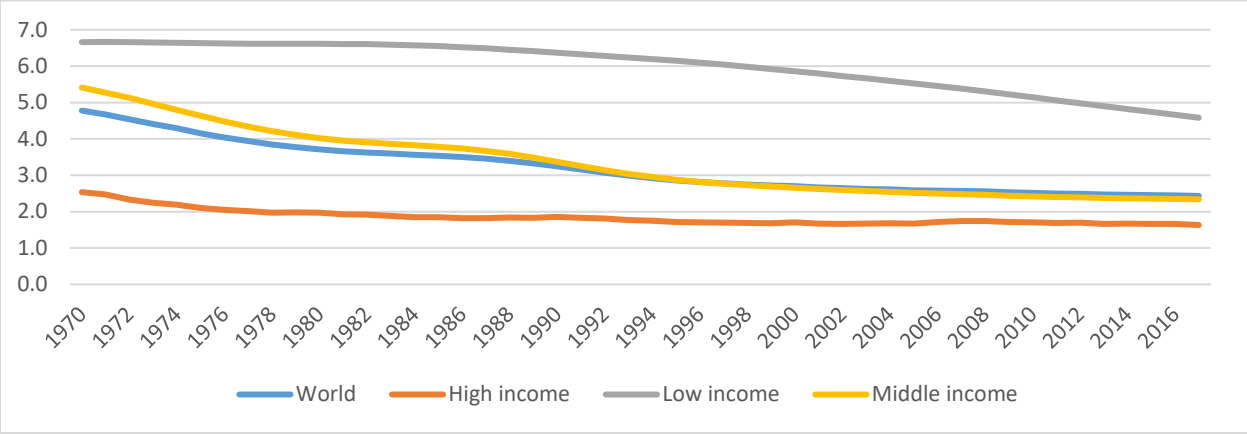
Correlation.

Rho=Expected value [((x – mean x)*(y-mean y))]/(standard deviation x*standard deviation y)

That computation is mechanical based on statistical theory.

The correlation allows us to see if things move together positively or negatively.

Total Fertility rate and wealth?



Income and rule of law?

Income and openness to world trade?

Income and geographic location?

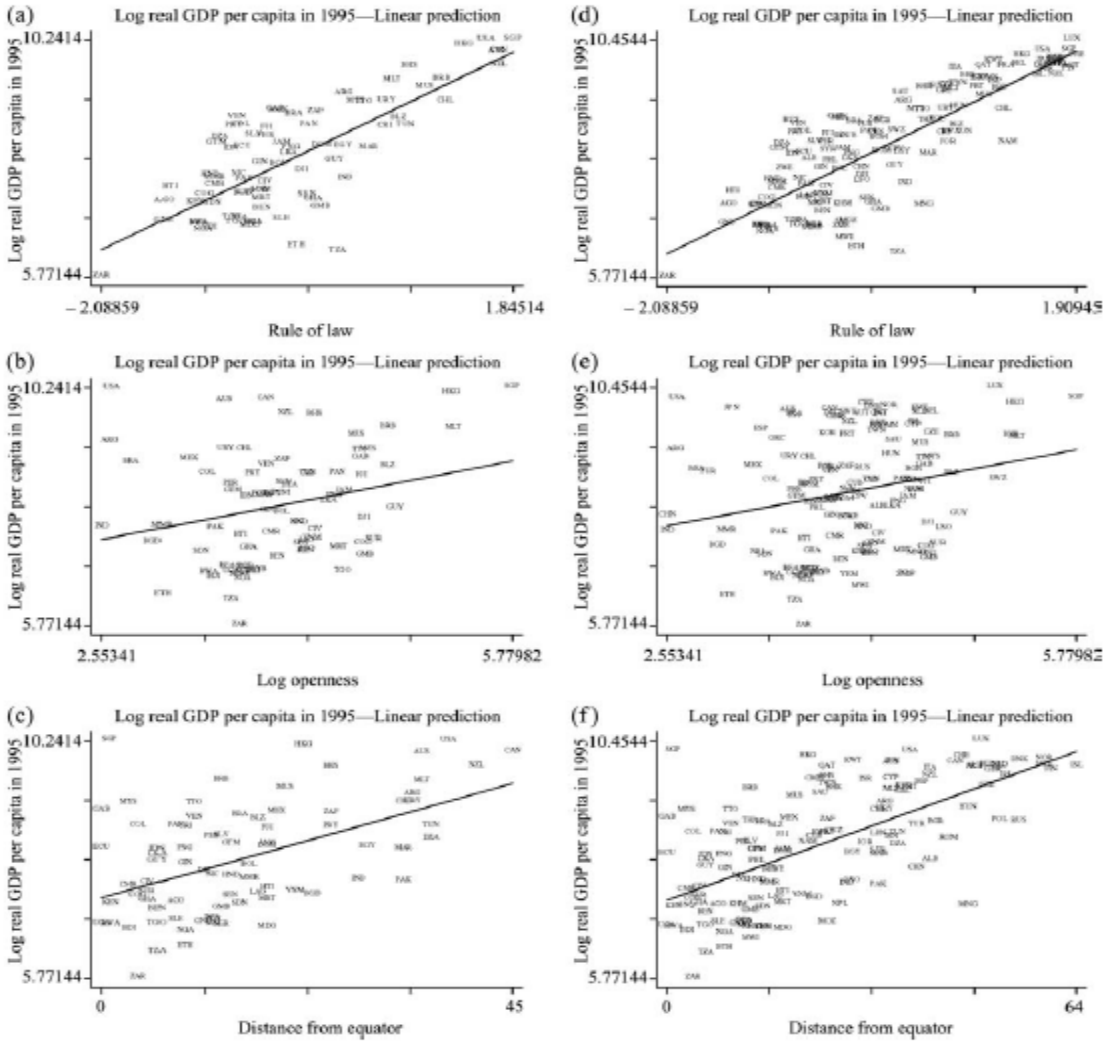
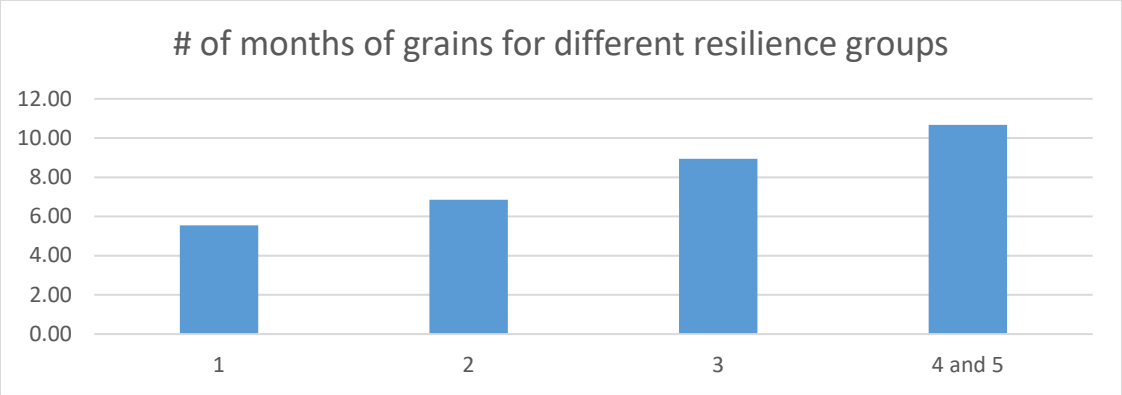


Figure 2. Simple correlations between income and its determinants (sample of 79 countries for (a)–(c); sample of 137 countries for (d)–(f)).

Resilience and food security? Household self-assessment for resilience on a 1-5 scale and responses to how many months out of a year are the able to provide sufficient grains for the family rho=.52.



To put structure on the correlation, we can reach for the idea of causation. That moves beyond statistical manipulation and forces us to reach for a behavioral theory.

Much of what we do in social science research is a search through correlations to try to identify causation.

What is the independent variable, the thing that has an impact on the outcome, and what is the dependent variable, the outcome?

Why does Total Fertility Rate decrease as expected income increases?

Why is higher income positively related to a measure of the rule of law?

Y=m*x+b

We are trying to identify independent variables, the things on the right hand side(x), the predictors, from the dependent variables, the things on the left hand side, the outcomes(y).

In classic y=f(x) notation, the y is the outcome, the dependent variable, the x is the independent variable(s).

We spend time arguing about what is the outcome and what is the input. What is the dependent variable? What is jointly determined?

Higher income countries have better governance.

Countries with better governance have higher income.

Q=286-20*p is the demand, Q = 88+40*p is the supply

We are contending that a change in p leads to a change in Q. But it could be that a change in Q leads to a change in p.

We also have to worry about the value you have for the outcome can be related to what you are trying to find out.

Our *survey monkey* results indicate **75% of students are satisfied or highly satisfied** with the IR curriculum. 50% of students filled out the survey.

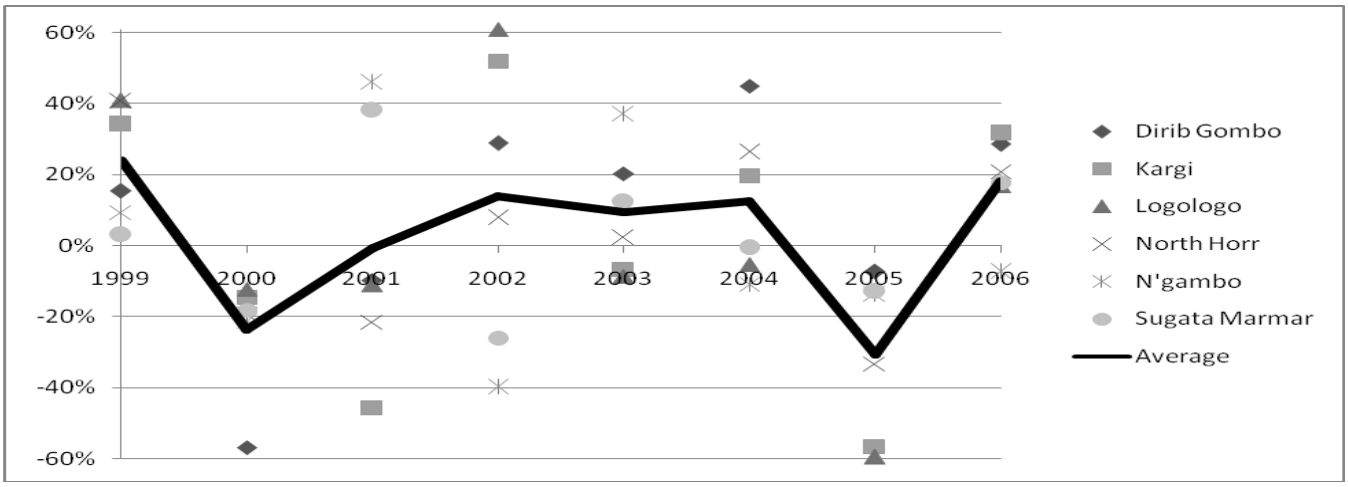
- What if who filled out the survey is related to the satisfaction level of the student?
    - Lower extreme- Assume the other 50% non-responses are all dissatisfied or highly dissatisfied. 75% of 50% or **37.5% of students are satisfied or highly satisfied** and 62.5% are not.
    - Middle – Assume the other 50% of non-responses are the same as the responses so **75% of students are satisfied or highly satisfied** and 25% dissatisfied.
    - Upper extreme – Assume the 50% who did not respond are all satisfied or highly satisfied so **87.5% of students are satisfied or highly satisfied**, 12.5% are not.

Our survey indicates 95% of people who went through our treatment program are employed and no longer homeless. However, our follow up survey could only find 60% of the people who went through the program. We did not find 40% of the sample who went through treatment.

Because we could not contact 40% we have 95% of 60% we could locate and had the intended outcome. .95*.60+0*.40 =57% is a worst case possibility, 95% is assuming the located are the same as the non-located, and 97% is the best case scenario if the missing 40% are all employed and are not homeless.

Purposes for conducting Social science research:

1) Exploratory. What do people around here mean by the word 'household'? What is a working definition of this concept that we can use to pose questions. Herd owning, common cooking pot, joint labor in fields, wives and co-wives, labor migration concepts….
2) Descriptive. How many households in this area are female headed? What is the average number of children? What is their experience with formal education? What is the mid upper arm circumference for the kids?
3) Explanatory. Do the kids residing in female headed households have higher or lower educational attainment than kids from non-female headed households? How about nutrition as measured by Mid Upper Arm Circumference (MUAC)? How can we explain differences in childrens' nutrition with differences in the share of household income controlled by women?



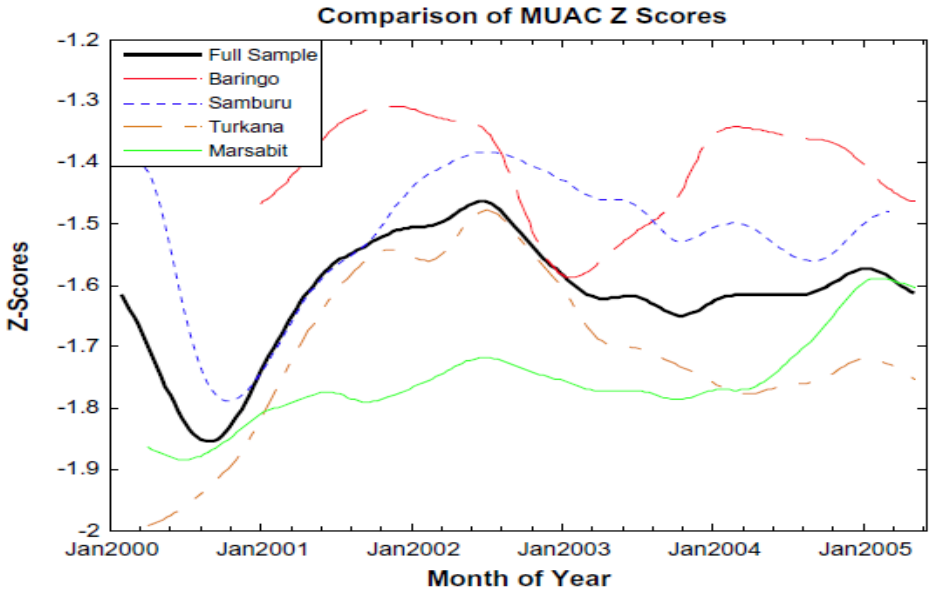


Fig. 2. Comparison of MUAC Z-scores.

Z-scored data points. Top panel is deviation of annual rainfall from the annual mean Bottom panel is mid-upper arm circumference data from children in different communities in northern Kenya ALRMP World Bank data.

Different methodologies and skills are called for when trying to address each purpose.

In research teams, different people tend to have comparative advantages in the different purposes, but we need to do all kinds at some point.

Explanations.

Ideographic explanations. This is less common in the research we will be focusing on in this class. The full listing of all the individual influences that lead to one specific unique outcome. Explain by exhaustive listing the multiple reasons a given unique outcome was realized. Fully explaining the single case / event in question. What are all the factors that led to the decision to delay the start of the spring 2021 term at Syracuse University? What are all the factors that explain the outcome of the Brexit vote?

Nomothetic – Most of what we will focus on. We identify a set of causal factors that lead to general classes of outcomes. A general explanation of what leads to a class of conditions and outcomes; an 'on average' perspective. What variables tend to lead to success or failure based on patterns seen in a sample of 30 nuclear arms deals over time? What economic, political, and historical factors have influenced referendum votes on political and economic integration in the post war era? What are the most important variables in the characteristics of students in Maxwell programs that predict they will be enrolled in this class?

Induction and Deduction.

- Induction; moving from the particular to the general.
    - Take observations of individual cases and make statements about general patterns.
    - Begins with whether (I wonder whether this factor matters for that outcome) and moves to why (Hmm, it seems to matter, I am going to try to explain why it matters).
    - I start with data and move towards a theory.
        - Individual household wealth seems to influence whether children attend secondary school. How to interpret the pattern of answers on the survey form.

- Deduction. Moving from the general to the specific.
    - Begins with why and moves to whether.
        - In other studies, we have found household wealth influences school enrollment rates. Is this the case in southern Ethiopia?
    - We write questions on the survey form to see if patterns seen elsewhere hold up here as well. Based on theory I think this will matter. I gather data to test this hypothesis.

Figure 1-3.

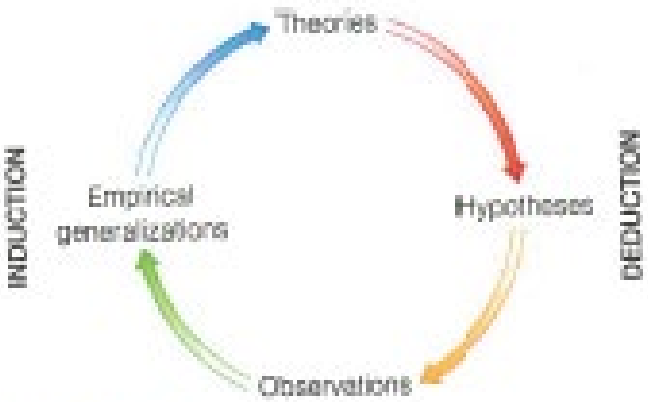
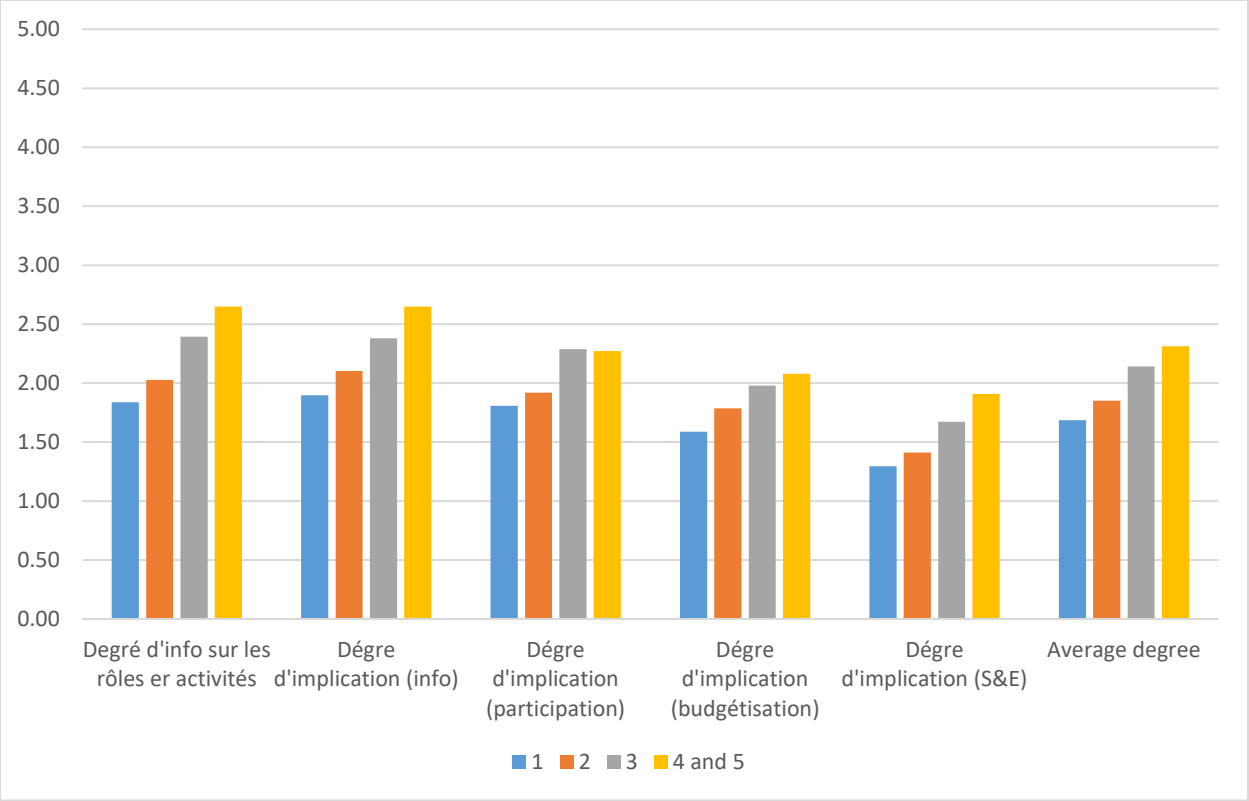


FIGURE 1-3

The Wheel of Science. The theory and research cycle for the social sciences can be compared to a relay race; although all participants do not necessarily start or stop at the same point, they share a common goal—to examine all levels of social life.

Source: Adapted from Walter Wallace, The Logic of Science in Sociology (New York: Aldine deGruyter, 1971). Copyright © 1971 by Walter L. Wallace. Used by permission.

Figure 11: Degree of involvement in community development activities (y axis) grouped by self-scored resilience on 1-5 scale (x axis)



Based on this, we did some follow up work to try to understand this result. This is an inductive finding, and led to some theorizing. Why might this occur? Gender? Caste / Ethnic / Lineage? Wealth?

Determinism versus agency; $y=f(x)$ or $y=f(x)+u$.

By setting up structure of average behavior we have not described individual decision making or motivation, but a description of the aggregate.

Quantification. Putting numbers on things. What year were you born? But why years? Months? Days? A choice is made in how you define intervals.

Resilience 1, 2, 3, 4, 5? -2, -1, 0, 1, 2?

Qualitative data. Are you young or old? What is the threshold that divides? We make choices to define categories.

We also have quantification of qualitative concepts if we agree to measures and constructs.

- Strongly disagree (-2), disagree (-1), neutral (0), agree (1), strongly agree (2).
- Gender binary (female=1, male=0).

Qualitative tends to work better for ideographic and quantitative for nomothetic explanations though this does not have to be the case.

For the term paper in this course, you will be developing a research proposal based on the material covered in this class. This is a brief introduction to the idea of research proposals.

Overview of Research Proposals (but also of research projects):

- Abstract / Summary.
- Introduction.
    - Review of the Literature. What is already known?
    - Specify the Problem / Question / Topic in this research project.
        - Explicitly note why this study is needed and what is being added to knowledge.
        - Sometimes also add in why you or the research team is well placed to conduct this research if proposals are competitive.
        - In some applied contexts a Theory of Change / Logframe, and possibly a set of indicators associated to this TOC.
- Research Design.
    - Where is this research going to take place?
    - Data gathering methods?
    - Sampling frame and population to whom this research is applicable
- A discussion of research ethics and measures to address ethical concerns in some contexts.
- Plan for analysis of data, noting methods and standards of evidence
- Timeline
- In some contexts a budget and a budget justification.

- Discussion and conclusion; what will we know if you do this research that we don't know now (reiterate now that it is all detailed as described above).

  Tell them what you are going to tell them, tell them, then tell them what you told them. Aristotle.