Lecture - 14
Introducing Quantitative Research Methods

Hello and welcome to the NPTEL MOOC's course on development research methods. So far, we are halfway through the course. With this class we will begin with module 3. And in this module, we will look at the different kinds of quantitative research methods. We will basically look at the philosophy of quantitative research methods with particular focus on survey research methods. We will also look at mixed methods research and the possible combinations. And also, how best we can communicate research.

In today's lesson, we will have a brief introduction to quantitative research methods. However, the focus in today's lesson will be on how we can best think with quantitative data. And I will begin with this idea of or with this notion of ideas versus evidence and how important ideas and evidence are with respect to substantiating research. In fact, by now as a serious student of research methods would understand that ideas and evidence always go hand in hand. And sometimes a big deal is made about evidences and sometimes a big deal is made about ideas and sometimes we are comfortable with an idea that we want to put forward as some kind of a justification for policy research and we do not give much importance to evidence. And at other times we closely hold evidence and do not give much importance to rival ideas that may be challenging the evidence that we have.

Now it is in this context that we need to learn how to think with data and also understand the fact that for being able to put forward an idea, there may be not just one, but various kinds of evidences. And all of these evidences may not go to justify the idea that we are holding on close to. They may be contrasting in nature as well as in the ideas maybe rival to each other. And in such a situation, what are the things to keep in mind, what are the best practices that we can follow with regard to substantiating our research.

(Refer Slide Time: 02:47)

What we will cover in today's lecture

1. The relationship between ideas and evidence
2. The role quantitative data play in substantiating ideas
3. Using quantitative data in advocacy

So, what we will cover in today's lesson are as follows. First is the relationship between ideas and evidence. Second is the role quantitative data play in substantiating ideas. And thirdly, when do we or how do we use quantitative data in advocacy and more particularly to remember things that we can hold on to when we are using quantitative data in advocacy.

Now I just mentioned about ideas and evidence and the extreme positions that we can hold on to with regard to evidence. Sometimes we want to hold on to an idea without sufficient evidence. And at other times we want to hold on to evidences, certain kinds of evidences and completely ignore the rival evidences that may be challenging the evidence or the justification based on the evidence that we are arriving at.

(Refer Slide Time: 03:40)

The relationship between ideas and evidence

- Twin dangers of upholding ideas without evidence or, more commonly, defending ideas only with evidence which supports them are prevalent in policy analysis.
- Ideas in turn support the nature and direction of public action.
- Evidence consists of both quantitative and qualitative data. Here, the focus is on the role quantitative data can play in choosing between rival ideas

So therefore, there is this twin danger of upholding ideas without evidence or defending ideas only with evidence which supports them. And this is something which is highly prevalent in policy analysis. Ideas in turn can support the nature and direction of public action and evidence mostly consists of both quantitative and qualitative data. However, in this lesson our focus will be on the role of quantitative data and the role that quantitative data can play in choosing between rival ideas.

(Refer Slide Time: 04:16)

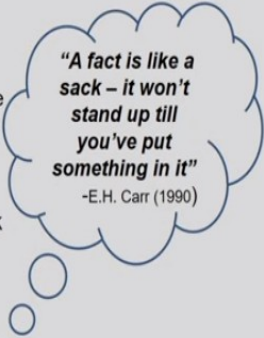

The role quantitative data play in substantiating ideas

- Quantitative data assume a dual role in analysis:
  - Testing ideas against data
  - Getting ideas from data

Both elements are important and tend to reinforce one another in the process of research

- It is important to acquire numerical literacy to make sense of numbers and be able to unpack quantitative arguments and subject them to scrutiny

- At the same time, quantitative skills which are not backed up by a sharp conceptualization of the problem at hand may look impressive, but often rest upon shaky foundations

"A fact is like a sack – it won't stand up till you've put something in it"
-E.H. Carr (1990)

So, what is the role of that quantitative data can play in substantiating ideas? First, we must understand that quantitative data assumes a dual role in analysis. It is often said that quantitative analysis deals with numbers while qualitative analysis deal with

words and this is not surprising because as evidences features prominently in any piece of quantitative analysis. Whereas qualitative analysis focuses on more of case studies, case specific studies, keeping the historical context and the time specificity in mind. So, it is in this context that we will need to look at the dual role that the quantitative data can play. One is quantitative data is used for testing ideas against data and also getting ideas from data and both of these elements are important and they tend to reinforce one another in the process of research.
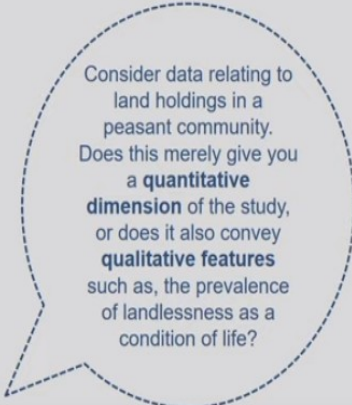
It is important to acquire numerical literacy to make sense of numbers and be able to unpack quantitative arguments and subject them to scrutiny. Many policy analysts and policymakers suffer from some kind of a data phobia and they give up or they submit to the data without putting up a decent amount of fight with how to analyze the data, how to interpret the data. And therefore, it is important to acquire some kind of numerical literacy to make sense of numbers and to be able to come up with arguments that can justify the data or what people say is to give a story to the data that we are trying to analyze.

Now quantitative skills which are not backed up by a sharp conceptualization of a problem at hand may look impressive, but often rest upon shaky foundations. Often research articles have very elaborate formulas and very elaborate sophisticated statistical tools and techniques for concluding, for providing a good interpretation of information. However, the theoretical foundation based upon which these analyses are being carried out may be shaky. And in such case, it may not qualify as good research. E. H. Carr is supposed to have said that 'a fact is like a sack. It would not stand up till you have put something in it. And it is always essential to understand how is it that we are going to put, how is it that we are enabling data to tell our story.

(Refer Slide Time: 06:51)

Nature of inference used in quantitative enquiry

- The fact that quantitative analysis deals with numbers, should not leave one to conclude – that quantitative analysis only deals with quantity and qualitative analysis only with quality.
- They are each **important in their own right**, and **also tend to interact**. Neither qualitative nor quantitative analysts can ignore taking account of both these **dimensions of social reality** – quality and quantity.

Consider data relating to land holdings in a peasant community. Does this merely give you a **quantitative dimension** of the study, or does it also convey **qualitative features** such as, the prevalence of landlessness as a condition of life?
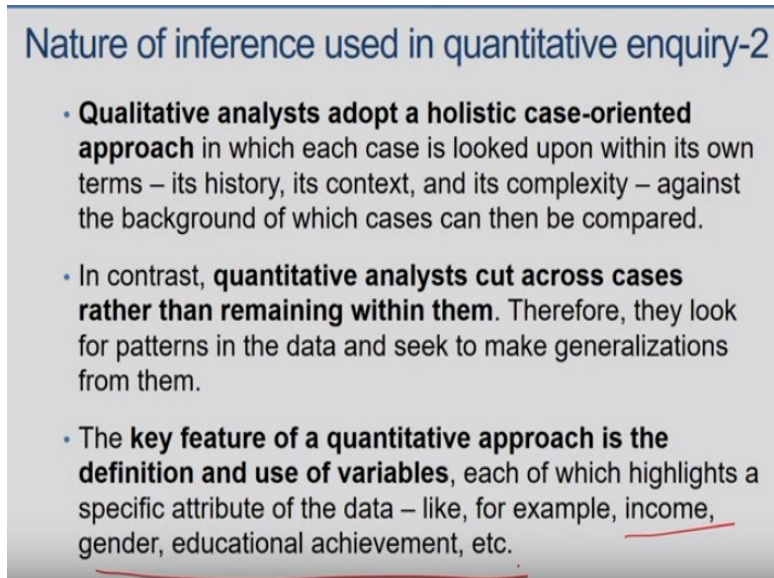
Now let us look at the nature of inference used in quantitative inquiry. The fact that quantitative analysis deals with numbers should not leave one to conclude that the quantitative analysis only deals with quantity and qualitative analysis only deals with quality because this is incorrect. For example, let us consider the issue of land holdings in a peasant community. Now does this description of landholdings in a peasant community, does this merely give us a quantitative dimension of the study or does it also convey qualitative features such as prevalence of landlessness is a condition of life?

Now another thing to remember is that a number can stand alone, but quantity always comes with units attached to it. In other words, a quantity is always a quantity of something. For example, we talk in terms of acres of land or we talk in terms of kilograms or age in years and so on and so forth. So, the key feature of quantitative methods is that we are always measuring something and it is always best to understand what is it that we are measuring. Whereas numbers may be standalone. Numbers, we can generate numbers based upon the qualitative information that we possess with us.

Another key feature of quantitative approach is the definition and use of variables which we will come to presently. But what is important to keep in mind here is that quantity and quality each are important in their own right and they also tend to interact. Neither quantitative nor qualitative analysts can ignore taking account of both of these dimensions of social reality. Qualitative analysts also have to take

account of quantity various times and quantitative analysts also have to take into account with regard to qualitative categories or qualitative case studies to bear in mind.

Now, what is the difference between qualitative and quantitative analysis is that qualitative analysts mostly adopt a holistic case-oriented approach in which each case is looked upon within its own terms; its history, its context, and its complexity against the background of which cases can then be compared. But quantitative analysts cut across rather than remain within these cases. Therefore, the distinctive feature of a variable as the name implies is that it varies across cases. Therefore, they look for patterns in the data and seek to make generalizations for them.

As I was mentioning the key feature of a quantitative approach is a definition and use of variables, each of which highlights a specific attribute of the data like for example, income, gender, educational achievement, etc. and so on. Now gender of being a male and a female is a qualitative attribute, but we can always provide numbers to them for bringing them into a quantitative analysis domain let us say for example. So, in doing so, what we are aware of is that chance is an ever-present rival explanation. In other words, they are aware that patterns they discern in the data may just be due to chance, variation, a fluke, a product of a particular sample and so on.

Nature of inference used in quantitative enquiry-3

- The fact that quantitative analysis is focused on making generalizations across cases does not mean that case-specificity is ignored altogether.
- **Good practice requires quantitative analysts to look for the exceptions or outliers**: cases that do not fit the general patterns and thus warrant special attention.
- Also, **not all quantitative analysis is variable-oriented** rather than case oriented.
- But, unlike in qualitative analysis, **case-specificity is always looked at from the perspective of patterns first identified across cases**.

An accountant who is looking at the financial data of an NGO will adopt a method of enquiry more akin to that of the qualitative researcher: case oriented, context-specific and with an eye for past history and performance.

One of the important features of quantitative analysis is that they seek generalizations, they are more focused on making generalizations across cases. But that does not mean that case specificity is ignored completely in quantitative analysis. In fact, a good practice requires quantitative analyst to look for exceptions or outliers and cases that do not fit the general patterns and therefore might want special attention from the researcher.

And also understand that not all quantitative analysis is variable oriented. It may be case oriented. For example, an accountant who is looking at the financial data of an NGO will adopt a method of inquiry that is most similar to that of a qualitative researcher because then the accountant will mostly look at the case history of the NGO whose financial statements are being checked for, the time span with which the NGO has been involved in a particular task and so on. So, in this case, an accountant who is looking at the NGO case is mostly looking for quantitative fact checking, but is approaching the case or is utilizing the approach of a qualitative analyst. Because you are looking at context specific and case oriented with an eye for past history and performance. Unlike in qualitative analysis, case specificity is always looked at from the perspective of patterns first identified across cases.

(Refer Slide Time: 11:49)

Testing ideas against data Vs getting ideas from data

- Many researchers hold the view that proper scientific analysis consists of testing an idea (hypothesis) against its evidence, however, we shall argue that empirical analysis is essentially contrastive in nature.

- This way, the relative strength of different ideas, their loose ends (that is, evidence which works against them), and their reliance on ad hoc justifications to guard against such loose ends, become more apparent.

Contrastive empirical analysis consists of **assessing rival ideas in the light of the evidence each brings to bear on the problem** to arrive at an explanation which is most plausible in the light of the overall evidence.

Now let us come to this idea of testing ideas against data versus getting ideas from data. Often, we are not just testing ideas against data but we may also get some ideas from data, which needs further probing and investigation. Many researchers hold the view that proper scientific analysis consists of testing an idea or a hypothesis against its evidence. However, we will see that empirical analysis is essentially contrastive in nature. And this is where we began with as far as this lesson is concerned, that contrastive approaches are more important to keep, to bear in mind when we are trying to think with quantitative data.

And what is contrastive approach? Contrastive empirical analysis consists of assessing rival ideas, in light of the evidence each brings to bear on the problem to arrive at an explanation which is most plausible in light of the overall evidence. This way, the relative strength of different ideas, their loose ends that is evidence which works against them and their reliance on ad hoc justifications to guard against such loose ends become more apparent.

Now let us take the example of, many of you might be aware of Amartya Sen's works on poverty and famines and the debates surrounding the explanations about why and how famines occur. And what are the most plausible reasons as to the situation of famine. Now before Sen's analysis took center stage on poverty and famines, the widely held dominant view was that famines are a situation that occurs due to natural shortages of food in a certain region. However, based upon evidences that Sen collected from various parts of the world, including the Irish famine, the Ethiopian

famine, Bangladesh famine, Indian famines and so on, he came up with this conclusion that although food availability decline appears in a certain region to be one of the temporary causes of famine but there are regions where famines have occurred in spite of food being in surplus. So therefore, food availability decline may not be posed as one of the immediate reasons of famine, but food accessibility becomes an important reason of why famines have occurred in the first place. And numerous debates arose surrounding this, this contrasting evidence that appeared and that could be posited against the dominant view that famines are a situation which occurs because of shortage of food.

So, what does one research analyst do in such cases where rival evidence appears? Do we ignore these rival evidences that appear? Or do we enter into further probing of the rival evidences and come up with a contrastive analysis or what is referred to as a contrastive empirical analysis by assessing the rival ideas in light of the evidence that comes up? Let us look at another example with regard to contrastive empirical analysis.

(Refer Slide Time: 15:00)

Average land size (acres)	Mean household size	% of households
0.8	3.5	6.6
1.8	3.5	12.3
2.7	4.4	20.5
3.7	5.1	15.0
5.2	5.6	22.5
7.1	6.7	10.9
10.0	7.6	9.4
16.7	9.9	2.8

Source: Collier et al. (1986), cited in Mukherjee and Wuyts (1998)

The table that shows on your slide deals with a case where evidence supports rival notions equally well. In this table the relation between average size of land holdings in acres the first column, and household size for a sample of 600 Tanzanian peasant households in 20 villages are shown. And the authors of this table, of this paper

basically are trying to say that land size is largely determined by household size. And they are trying to say that Tanzania is a land abundant country, and the amount of land that households possess is largely determined by the household size rather than the presence of various other variables. So, this table is listed as evidence for the hypothesis that land size is determined mainly by household size.

But if you look at this table closely you will see that this construction is somewhat awkward, since it clearly refers to grouped data. Here land size instead of being presented as intervals, class intervals, is grouped in terms of the average land size and without specifying class intervals, and it also lists group averages for each unspecified interval. So then do we agree with the authors that the evidence supports their claim? And if you look at the association of household size with land size, you would see that the there is a positive association. As the household size increases, land size also increases. So probably we will be quick to conclude that the hypothesis that the authors are forwarding is indeed plausible. The point however, is whether this is the only hypothesis that the data support and can the authors explore further possible explanations or alternative explanations that may not hold the hypothesis true.
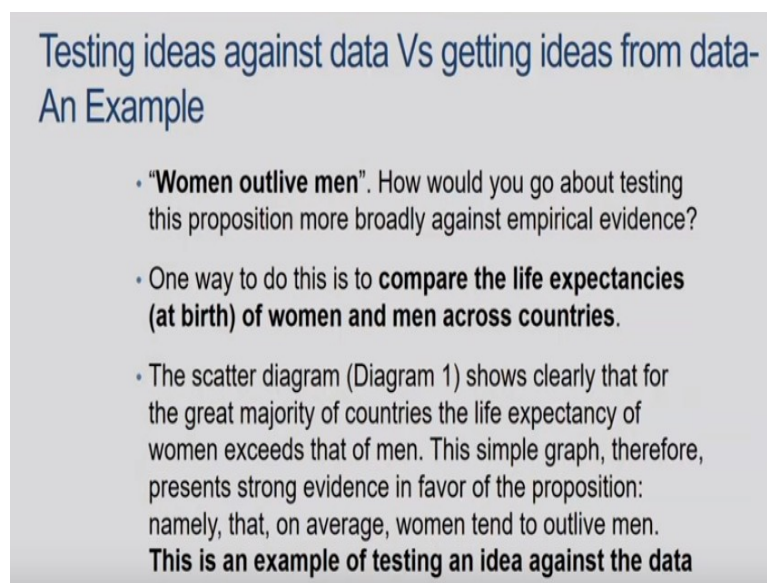
Now in this table the authors have considered household size as given and this table is mostly taken from Chandan Mukherjee and Wuyts' paper on thinking with quantitative data and scholars like them largely argue that there may be contrasting or alternative plausible explanations that may reject the hypothesis that they are trying to support here.

So here the other scholars argue that the authors assume that household size is given, and hence land size adapts to the number of hands available and mouths to feed. But is it correct to see household size as a given? Because household size is dependent upon a large number of variables. For example, studies have shown us that there are certain families in which household size becomes less because of fragmentation and certain families are more cohesive, particularly in the case of low-income developing countries. Poorer families seem to fragment more than richer families, because members of the poorer families have to move out in search of work and therefore, they may have a smaller size during the period of the survey. Richer families may have a larger family size, because of their members of the family not moving out in

search of better opportunities. And also, because they may be able to hire more labor hands from poorer relatives.

Therefore, family size itself becomes a determinant of asset holdings of the households. So, land size instead of being associated with family size may see an association with asset holdings of the peasant households here than the family size, and which is a rival evidence which is running against the hypothesis that the authors are trying to make here.
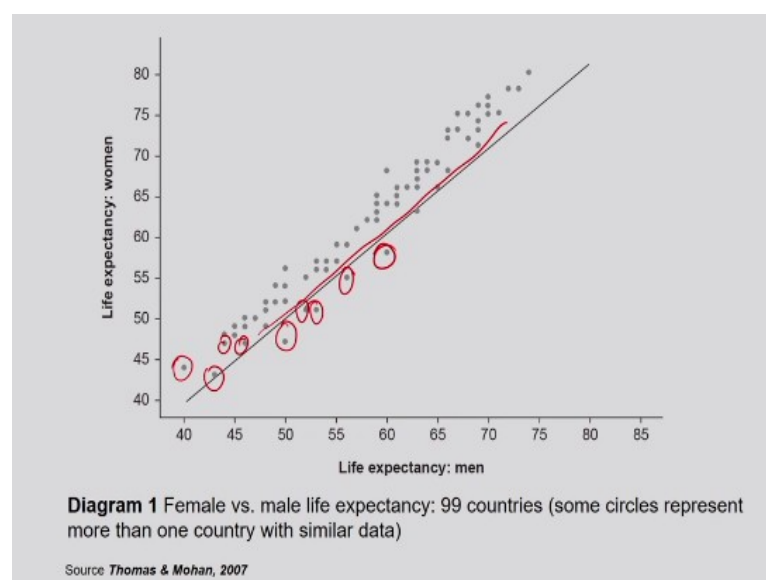
(Refer Slide Time: 19:09)

There can also be situation where we are testing ideas against data versus getting ideas from data as I was pointing out. Now until recently statistical texts or statistical textbooks usually focused more on confirmatory analysis. And the underlying idea was that scientific method should consist of testing ideas against data. But nowadays, modern texts recognize the importance of getting ideas from data and devote attention to techniques which allow us to use this better. The underlying principle is that we should never impose a story on the data but allow the data to tell a story by themselves.

Now let us look at this example where there is this general understanding that women outlive men. Most people will actually know from casual observation that women outlive men. But is this everywhere the case? Is this the case in every country or every region of the world? And how does one go about testing this proposition more broadly

against empirical evidence? Now one way to do this is to compare the life expectancies at birth of women and men across countries. Life expectancy is basically a demographic indicator that measures the amount of years that a person would live depending upon the mortality rates or age specific mortality rates that can be attributed to the person whose number of years we are following. So, it is a demographic indicator of mortality. Now how do we go about measuring it?

Now note that for example, the life expectancy of a certain country comes out to be 45 years. Now this does not mean that an average person in that country lives up to only 45 years. But it could also be the case that because infant mortality rates are very high in this country, therefore, the average lifespan is being pulled down because of very high infant mortality rates. So what are the different conclusions that we can come up with or what are the different ideas that we can come up with when we are empirically testing this observation. Now this can be done with the help of a scatter diagram. Very simply, we have to put across this idea. All the statisticians make use of various probability techniques to be able to come up with an inference to be able to come up with proper conclusions on this. However, a simple graphical representation in the form of a scatter diagram can also give us a lot of ideas.

(Refer Slide Time: 21:53)

Diagram 1 Female vs. male life expectancy: 99 countries (some circles represent more than one country with similar data)

Source *Thomas & Mohan, 2007*

Let us look at this figure here. Now in this figure female versus male life expectancy for 99 countries is shown from 1990 onwards. So, the x axis measures life expectancy

of men and the y axis measures life expectancy of women. And you would see that there is quite a bit of variation in this plot. So each of these points refers to countries. All of these points on the scatter refer to different countries.

The plot features a 45-degree diagonal. Why is this? We draw this line for a very simple reason to show that the locus of points with the life expectancy of women and men are equal. So, in general if there were no differences between life expectancies of women and men, the actual observed points would fall on the 45-degree line. If however, life expectancy of women generally exceeds that of men, most points will be situated above the line; so as in these cases above the line. And if the life expectancies of women are less than men, then it would fall below the line. So, this is a 1998 paper. And the focus here is on the data from 1990 onwards for 99 countries. These are basically taken from the World Bank tables.

Another thing to note in this plot is that life expectancies of men and women range from 40 to 44 years. If you look at these points here, the life expectancies range from 40 to 44 years and goes up to about 84 years, 74 to 81 years. So, with this means that there are significant differences between countries and therefore, we would expect the scatter of points slopes upwards within a fairly narrow range. And also we have seen that in countries where male life expectancy is lower, female life expectancies may also be relatively lower, although higher than the male life expectancy and vice versa.

So, what is the simple graph? What is the story that this graph is telling us? This graph is basically giving us a strong evidence in favor of our proposition that women outlive men, and the life expectancies of women are higher than that of men. On average women tend to outlive men. And this is an example of testing an idea against the data. We hold this idea and we have tested this idea against the data by plotting life expectancies of men against women. But do we stop our analysis here or we probe further? That is the question to keep in mind when we are trying to think with quantitative data?

Should we just be pleased with the fact that our data has confirmed the idea that we were holding on to or can the data also give us further ideas to probe about. If you look at the figures very closely again, you will see that there are these plots which are lying below the 45-degree line. And therefore, there is a question that we need to

ponder about with regard to which are these countries that are falling below the 45-degree line and what are the circumstances, socioeconomic conditions of these countries that are giving rise to such a situation.

(Refer Slide Time: 25:14)

Testing ideas against data Vs getting ideas from data- An Example-2

- The fact that we confirmed the initial proposition against the data does not mean that our analysis should stop here.
- There are five data points situated below the line. This means that, in these countries, women's life expectancy is less than that of men: a feature which contrasts sharply with general world-wide patterns.
  - Which countries these might be?
  - Whether or not there is a gender bias against women in these countries?
  - From left to right along the graph, the scatter of points becomes more distanced from the 45° line. Why is this so?

The countries corresponding to these points below the line are, respectively, Bhutan, Bangladesh, Nepal, Pakistan and India. There is quite a debate in the literature whether or not there is a gender bias against women in these countries. From left to right along the graph, the scatter of points becomes more distanced from the 45° line. This means is that the discrepancy between the life expectancies of women and men increases as life expectancy of both women and men increases.

So, as I was saying there are five data points that are situated below the line. This means that, in these countries women's life expectancy is less than that of men, which is a feature which sharply contrasts with general worldwide patterns. So, the questions that we need to ask are which countries these might be. Whether or not there is a gender bias against women in these countries.

If you look at from left to right along the graph, the scatter of points becomes more distanced from the 45-degree line. Now these countries correspond to the points below the line are Bhutan, Bangladesh, Nepal, Pakistan and India. And there is quite a debate in this literature whether or not there is a gender bias against women in these countries. From left to right along the graph the scatter of points becomes more distanced from the 45-degree line. This means that the discrepancy between life expectancies of women and men increases as life expectancy of both women and men increase. And this is how we are confronted with the debates surrounding the socioeconomic conditions of men and women across countries.

When we were debating about qualitative analysis and quantitative analysis, now this is a prime example of how a quantitative analysis brings us in confrontation with the

whole qualitative debate regarding quality of life and gender bias that is prevalent in specific countries vis-a-vis those of the rest. So and this is how qualitative and quantitative data go hand in hand.

So, what are some of the points to remember with regard to getting ideas from data, we should always be on the lookout for clues and hints which point in a different direction or which require us to deepen our analysis. Data Analysis is similar to an open-ended dialogue, if for example when we are asking questions in an interview method, if we do not ask the relevant questions to the respondent, we will not get the right kind of answers. Similarly, in the case of data analysis, if we do not know what is it that we are measuring and how is it that we need to go about measuring, we will not get the right kind of answers. So, we must make sure that our questions are not worded as our own pet answers which we merely seek to confirm with a simple yes or no. And most importantly, we should never impose a story on the data.

But equally we should not expect the data to tell a story by themselves. And this is where data needs to be theory inspired. The theory that we are working on should be one of the stepping stones to moving on to data analysis. That brings me to the next slide on factuality of data.

The Factuality of Data

- Data always yield a selective view of an aspect of reality and, therefore, what is considered to be a fact depends in part on the criterion which underlies the selection of data.
- Data need to be put into context before they become useful knowledge.
- Facts, therefore, are always theory inspired: **it is theory which renders a piece of information relevant and, hence, accounts for it to be selected as a fact**.

Even though two data sets may be related, this does not mean they are causally related – just because children get taller as they get older and also progressively develop language skills does not mean that getting taller improves language skills or vice versa!

To say that national accounts data are theory inspired means that theory determines which data are deemed to be relevant and informs the way these data are collected and structured into meaningful policy-relevant macro-aggregates such as the gross domestic product, national income, consumption, investment, savings, and so on.

Now data always yield a selective view of an aspect of reality and therefore, what is considered to be a fact depends in part on the criterion which underlines the selection of data and data need to be put into context before they become useful knowledge. So, facts are therefore always theory inspired. It is theory which renders a piece of information relevant and therefore accounts for it to be selected as a fact.

What is fact? What is a factual piece of evidence is something that also has to be theory inspired. How do we look at that factual piece of evidence? Look at this example here. Even though two data sets may be related, this does not mean that they are causally related. We are often looking at association and causal relationship between factors. For example, just because children get taller as they get older and also progressively develop language skills does not mean that getting taller improves language skills or vice versa. Getting taller and getting improved language skills show some kind of a positive association. And often this positive association will be highly significant because this is what our general observation also tells us. But this does not necessarily mean that they have a causal relationship, there is a causal relationship between the two.

Similarly, to say that national accounts data are theory inspired means that theory determines which data are deemed to be relevant and informs the way these data are collected and structured into meaningful policy relevant macro aggregates such as GDP, national income, consumption, investment, savings and so on.
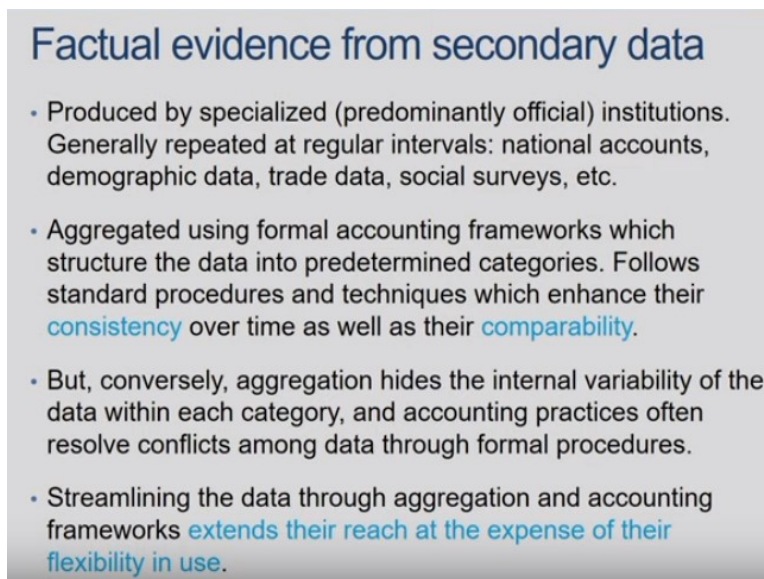
Now another thing that needs to be kept in mind is factual evidence is something that can be collected both from primary data and secondary data. However, there are certain characteristic features of assessing primary data and that of assessing secondary data. Now primary data is usually collected by a single researcher or a team of researchers in policy analysis. And so, it is the researcher who selects them, who assesses them, who interprets the data and so on and shapes them as facts with analysis and therefore, this is a very highly time-consuming process. And also, it tends to be case and time specific. Often it is impossible, with regard to primary data it is often impossible to verify data or to correct inherent biases in data collection. And hence it is important to explain the nature of the data and the procedures which went into the collection of the data.

Now some primary data analysis involves fieldwork and here researchers generally develop strong local knowledge. They acquire a feel of the specificity of the location in question and as producers of data, they tend to be quite aware of complexity, variability, and uncertainty of social data. So, they are therefore less inclined to generalize too quickly. When we are focusing on survey data there is a risk of generalization and this is something that is a matter of continuous debate among data analysts with regard to generalizations. And one of the things to keep in mind is whether generalization is being attempted through the study or not also needs to be kept in mind.

We might want to tell the story of a certain region with the help of quantitative information that may have policy relevance. However, given the historicity and the context of the region that we are studying, it may not be proper to generalize the findings of that region to rest of the world or to rest of the country or the rest of the state. Because studying that particular region also has policy relevance. And these are also things that needs to be kept in mind with regard to generalization.

Other primary data analysis is based on surveys that are carried out as I said by the researchers themselves and such surveys are similar to but generally smaller in size than the large-scale surveys carried out by official institutions for the purposes of publishing secondary data. And survey analysts unlike field workers tend to be more distant from local circumstances and knowledge. One of the classes in this module will be devoted to survey research methods in which we will discuss each of these features that I am pointing out to now.

(Refer Slide Time: 32:23)

Now as I was saying, factual evidence may be collected from primary data or secondary data. And large sample surveys are usually carried out by official institutions and they are generally repeated at regular intervals- national accounts, demographic data, trade data, social service etc. For example, we have the Indian census operations carried out by the Registrar General of India, which is carried out every 10 years. Or we have large sample surveys carried out by National Sample

Survey Organization every five years. Similarly, we have enterprise surveys, livestock surveys and so on and so forth. Agricultural census which are carried every five years and so on. So, these are also survey data. However, these are carried out by official institutions on a large-scale basis and then put up on various domains to be used by individual researchers.

But there is a problem of aggregation here and one needs to look at these aggregation issues very minutely. Now aggregation using formal accounting frameworks, which structure the data into predetermined categories. Usually most of these data, that qualify as secondary data sources based upon surveys. They come in predetermined categories and they follow certain standard procedures and techniques which enhance their consistency over time as well as their comparability. For example, the recent national Family Health Survey NFHS-4 that comes out or the recent NSSO round surveys which has been in the news with regard to employment statistics and so on. There are issues of comparability and therefore methodological exercises becomes central to these surveys.

But aggregation also hides internal variability of the data within each category and accounting practices often resolve conflicts among data through formal procedures. One of the basic difference between survey analysts that are carrying out small sample surveys, single researchers or team based researcher sample surveys vis-à-vis those carried out by established institutions is that with single researchers or team researchers carrying out survey on a small scale, you are still closer to the data, still closer to the cases that you are studying. You know the cases on a one on one basis- may or may not but largely you know the cases on a one on one basis. Whereas, the survey data that the specialized institutions collect, they are completely separated from the cases as far as the data is concerned. So, streamlining data through aggregation and accounting frameworks extends their reach at the expense of their flexibility in use.

(Refer Slide Time: 35:10)

Some basic principles of using secondary data

Questioning secondary data

- Use the data systematically and creatively
- Take time to look carefully at the structure of the table, its various categories, and the way they are cross-tabulated
- Reflect upon and note down any important conclusions you can draw about the limitations inherent in the data
- Fire questions at the data. Each question approaches the data from a different angle. These questions need to be essentially theory-inspired and motivated by the key components
- Each question gives you a different "handle" on the data. Each handle allows you to look at your data and gain specific insights
- Follow up the questions that look promising
- Collapse the data so that it emphasizes the question we want to pursue. Do not carry around any extra baggage in terms of data overload
- Transform the data appropriately

Now some of the basic principles of using data that must be kept in mind are as follows. The questioning, we must question secondary data. And what are the things to keep in mind when we are questioning secondary data? Now some basic principles on how you might go about questioning secondary data are as follows. How should you go about using this table more systematically and creatively? And this first principle is very simple yet unfortunately often forgotten. It consists of taking time to look carefully at the structure of the table, its various categories, and the way they are cross tabulated. This is something which usually researchers downplay the importance of it. When they take time to look carefully at the structure of the table like in the case of the Tanzanian peasants example that we saw on land size and family size, if we look very closely at the structure of the table and its various categories and the way they are cross tabulated, that will also give us a sense of how we want to interpret our data. And whether we want to question the secondary data that we are using.

We must also reflect upon and note down any important conclusions that we are drawing about or limitations inherent in the data. So first, we also need to fire questions at the data. Each question approaches the data from a different angle. And these questions need to be essentially theory inspired and motivated by the key components. It is also important that each question gives us a different handle on the data. Each handle allows us to look at the data and gain specific insights. And we must follow up the questions that look promising. Collapse the data so that it emphasizes the question we want to pursue and we should not carry around any extra baggage in terms of data overload.

And final step is that of transforming the data appropriately. Data transformation, we have collected raw data, but the raw data will not be able to tell the story all by itself. So, it all depends upon how best we transform the data into appropriate, creatively designed tables. And that is where the use of statistical tools and techniques of being able to transform the data comes in.

So, one is about thinking with the data, thinking about theory inspired data, thinking about stories that we want to test against the data, thinking about contrastive empirical evidences that may help us to justify the information that we have on hand. And the other is to be able to transform the data with the help of appropriate statistical tools and techniques so that we can tell the story better.

(Refer Slide Time: 37:57)

So, this brings me to the last and the concluding part of this lesson where we want to use quantitative data in advocacy. When we have already carried out empirical studies on fact finding based upon quantitative data, how is it that we use quantitative data for advocacy purposes, public policy purposes and so on. In this last part I will pose, mostly discuss this part in the form of questions that need to be kept in mind. And finally end this with an example which I will read out from certain article that I am referring to, the reference of which is given at the end of this slide and I will be showing that in this lesson.

(Refer Slide Time: 38:42)

Making the most of data

- Practical ways of maximizing the impact of your data. There are three golden rules:
  - Understand the positions of all the actors in the debate and how they interact with each other.
  - Constantly reflect critically on your own position and the way to achieve your goal.
  - Seek new ways to influence the debate and define the problem in ways that integrate your views and the data.
- It is important to be careful to focus on where new information is most needed or will have most effect.
  - What do you hope the research will show?
  - What effect will they have on the debate?
  - Is there a better, easier or cheaper way to achieve the same effect?
  - Are you looking at the best parameters?
  - What are the weak points?
  - How will the new evidence lead the debate forward?

Now the thing to be kept in mind is how to make the most of data. So, there are practical ways of maximizing the impact of our data and there are three golden rules for this. One is to understand the positions of all the actors in the debate and how they interact with each other. I was taking the example of famines. How famines were defined before Sen and Sen's analysis on poverty and famines and entitlement analysis came on scene and what were the debates that arose after.

So when one is approaching the study of famines or food security and so on and so forth, one needs to be completely aware of what are all of these actors telling us about or what are the positions of all of these actors in this whole debate of how to explain famines or what are the causes behind famines and how famine situations or famine like situation should be dealt with. And all of these positions of the different actors should not be standalone, they should be interacting with each other. And keep in mind here that we have already done a class on systematic review of literature, where we are talking about, we are talking precisely about this, that how the different arguments that we are trying to collate with the help of systematic review of literature should be talking to each other. There is an interaction between all of these actors for the arguments that are being made.

The second golden rule to remember is to constantly reflect critically on our own position and the way to achieve this goal. So, if you are coming with a certain position or a certain idea, we should also be critical of our own position so that we can find the best ways to achieve the goal. And the third golden rule is to seek new ways to

influence the debate and define the problem in ways that integrate our views and the data.

Now it is important to be careful to focus on where new information is most needed or will have most effect. So, the questions to keep in mind here as follows. What do you hope the research will show? What effect will they have on the debate? Is there a better, easier or cheaper way to achieve the same effect? Are you looking at the best parameters? What are the weak points? How will the new evidence lead the debate forward? These are the questions that must be kept in mind for being able to focus on where information is most needed or what effect with that information have, once we have collected it and presented it for our story.

(Refer Slide Time: 41:09)

How good is the evidence? Has the evidence been collected and analyzed properly? Are the conclusions being drawn from it justifiable? What other interpretations of the data are there? What are the weak points? Should the issue be reconceptualized? And this is where we connect it with the research process of going back and forth on the different steps of research that we are carrying out. Once the data has been collected and analyzed, do we want to reconceptualize the research question that we began and accordingly change the conclusions or accordingly work on the conclusions that we want to come up with through our research project.

(Refer Slide Time: 41:47)

Where data do not support your arguments

- Don't announce in advance what your research will show or prove.
- In the light of your data, critically evaluate your policy position.
- Reassess whether you have used the correct methodology and analyzed the data properly.
- Question whether you have collected the most appropriate data.
- Ask whether your data is relevant to the policy debate.

Now where data does not support our arguments, there are certain key things that needs to be kept in mind and, this is something that frequently occurs in the world of research where we have began with a certain position, with a certain argument, but the data that we have collected runs completely opposite to our arguments. So then what are the things that need to be kept in mind and therefore, it is better to be a cautious researcher, than be a very bold researcher and make big announcements in the beginning of the research.

So, things to keep in mind is one do not announce in advance what your research will show or prove. And therefore, remember the wordings of the research outlines and we say that this research will try to explore, this research will attempt, this research will inquire into these points or will investigate into these points rather than announcing that this is what we are trying to test. Because your data or the evidence might completely be contrastive in nature and run against the argument that we are trying to make.

In light of your data critically evaluate your policy position, reassess whether you have used the correct methodology and analyze the data properly, question whether you have collected the most appropriate data and ask whether your data is relevant to the policy debate.

(Refer Slide Time: 43:06)

Problems with using evidence
- Data cannot prove anything. The nature of data and their limitations means that any conclusions always carry an element of uncertainty.
- Research findings may also raise more questions or may not give a complete picture.
- The boundaries implied by a question also make a difference to which conclusions can be drawn.
- The very methodology one uses will be open to challenge. If survey methods are used, did these involve use of a representative sample? Were the questions asked openly and how were the data analyzed?
- Even if the framework of data analysis is agreed upon, there is also the issue of uncertainty in the raw data.
- Issue of ignorance. E.g. Before the discovery in 1985 of the ozone hole above Antartica, it was considered rather eccentric to ask questions about the potential effects of ozone depletion on human health.

What are the problems with using evidence? One is that data cannot prove anything. The nature of data and their limitations means that any conclusions always carry an element of uncertainty. We may always be tempted by the fact when we are initiating some kind of a research, there is this temptation that we will come up with, we will collect data and through our findings we will come up with some straightforward conclusions and that rarely happens. And therefore, this element of uncertainty is always lurking and one should be aware of that.

Research findings may also raise more questions or may not give a complete picture. And the boundaries implied by a question also makes a difference to which conclusions can be drawn. For example, some of the nutrition studies have shown that if family income is put in the hands of the mothers or the women, then it is spent better on children than if the money is put in the hands of men. And also, studies have shown us that there is a siphoning of our family incomes from being spent on children and the male adults utilizing them for their own social consumption purposes.

And this has brought in this conclusion somewhat erroneously, that cash transfer schemes or providing cash in the hands of mothers, will look after this issue of nutrition completely. Whereas counter studies and counter evidences are also quite a lot now and it has shown us that it has led to unnecessary and unimagined problems. These kinds of interventions, government interventions have created unimagined problems. For example, it has led to a monetarized economy, where the money is not being spent on what it was supposed to be spent in the first place.

It has also led to various other kinds of changes in the social institutions or the social structure in which these economies, rural economies were based on. It has led to some kind of a commoditization of the rural economy without it giving a proper shape to where it is going. So, these are some of the changes that have come up because of these kinds of cash interventions. And cash interventions programs increasingly are being shown to have not given rise to the desired results. Therefore, suggesting cash interventions or providing cash in the hands of the mothers or the women may not seem to be a sound nutrition policy.

Sometimes the very methodology one uses will also be open to challenge. If survey methods are used, the questions to ask is did these involve use of a representative sample, were the questions asked openly and how were the data analyzed. Even if the framework of data analysis is agreed upon, there is also the issue of uncertainty in the raw data. And there are always issues of ignorance. For example, before the discovery in 1985 of the ozone hole above Antarctica, it was considered rather eccentric to ask questions about the potential effects of ozone depletion on human health.

Now let me end this lesson with a small example of contrastive evidence. Now in the early 1990s, a dispute arose in connection with global warming debate between two NGOs, the World Resources Institute and the Center for Science and Environment in India. The WRI or the World Resources Institute based in Washington DC produced data detailing each country's emissions of carbon dioxide and other important greenhouse gases. The WRI calculated each country's emissions and then reduced the total by the amount that would be absorbed by natural processes. And calculating their emissions in this way, they came up with the ranking of countries from highest downwards. So, the ranking of countries were USA, USSR, Brazil, China, India, European Union, Indonesia, Canada and Mexico.

Now what was surprising about this ranking is a suggestion that less developed countries are major polluters. Now these results were challenged by the Center for Science and Environment in India. And on the face of it, the calculations may seem straightforward, and the conclusions although uncomfortable apparently correct. But the CSE made two criticisms of the data analysis. Number one, they said that the natural processes which absorb carbon dioxide should not be apportioned by country but rather according to population size and doing this gives recognition to the very

low greenhouse impact of average Indian compared to someone from USA. Calculating things this way leads to India and China no longer being amongst the top five polluters but replaced by Canada and Germany.

Number two, the criticism that the CSE made was that all types of emissions in the WRI study was treated as equal. And thus, luxury emissions associated with driving to shops were seen as being no different from carbon emissions, survival emissions of using fuel for cooking for example. And the policy implications arising from both of these criticisms are immense. And this came into the center stage of debate with regard to carbon emissions and the widening gulf between the so-called Global North, Global South developed developing countries and so on. I will leave you with this example of how contrastive evidence and interpretation of the same bit of evidence can lead to completely different outcomes.

(Refer Slide Time: 49:04)

References used for this lecture

- Mukherjee, C., & Wuyts, M. (1998). Thinking with quantitative data. Finding out Fast: Investigative Skills for Policy and Development, London: SAGE Publications.
- Mayer, S. (2007).Using evidence in advocacy.Research Skills for Policy and Development: How to find out fast, London: SAGE Publications.
- Thomas, A., & Mohan, G. (2007). Research skills for policy and development: How to find out fast. Sage.

For a comprehensive literature on the topics covered in this lecture it is also suggested that students go through the reference list of the above cited papers.

So, these are the texts that I have used for this lecture. Mukherjee and Wuyts 1998 'Thinking with quantitative data. Finding out Fast. Investigative Skills for Policy and Development.' This is an edited paper which is taken from this book, Thomas and Mohan. 'Research Skills for Policy and Development, How to find out fast'. I have also used the paper on Mayer, the book Using evidence in 'Advocacy Research skills for policy and development, How to find out fast'. Now there is also for a more serious student who may want to investigate further into development questions by

following quantitative research methods and who would like to focus on data from the developing countries I would highly recommend the book by Chandan Mukherjee on Analysis on Econometrics for The Developing Countries. If you want to use econometric analysis when dealing with developing country's data, that is a very good book to begin with.

So, I will end with this in today's class. See you in the next class. Thank you.