

Handling Large-Scale Unit Level Data using STATA
Professor. Pratap C. Mohanty
Department of Humanities and Social Sciences
Indian Institute of Technology, Roorkee
Lecture No. 01
Introduction to Data

Welcome to the NPTEL module on handling large scale unit level data using STATA. It is one of the unique modules so far as the social science is concerned where we will be discussing the length and breadth of understanding unit level data in the Indian context. So, our title is defined accordingly and this is very much fitted to the MOOC schedule. MOOC program of NPTEL is part of the MHRD initiative. So, it is actually mentioned in the SWAYAM program.

So, let me give you a background to it, little background to it in addition to my own field on it. So, my name is Dr. Pratap Mohanty. I am a faculty member at the Department of Humanities and Social Sciences, IIT Roorkee. I used to deal with unit level data and particularly called large scale unit level data. I have already handled some of the professional modules on this aspect and particularly processing the unit level data or you may say extracting unit level data, processing, decoding, mining to some extent and with the final process we usually go to the extent of interpreting the data, otherwise it is completely meaningless.

Especially in a society where there are large number of you may say multicinity of problems, number of conflicts of interest, we are actually puzzled with the correct direction for any particular context to draw certain inferences or conclusion especially I am referring to the policy context. So, those who are in the Ministry or formulating policy, they do require inferences from the actual data.

Now question arises what do you mean by data? How should I start discussing about data? Whether any individual just saying any narratives or just interpreting narratives is called data? Or a structure or a systematic presentation is called data? So, there are large number of confusions on these particular aspects. So, this particular module is called handling large scale unit level data is very trivial in the present day's research especially for the Ph.D. students, for the professionals and also for the business classes or the corporate houses. let me give you an example in this context.

Like suppose one corporate, one company called FMCG one, Fast Moving Consumer Goods Segment, these simply wanted to understand their products to be marketed in a particular location. How do they know that their product has potential demand unless and otherwise they don't check the field reality especially the socio-economic and demographic background of the individuals in that particular locality they may not come to a conclusion of how much to be sold. Of course, it requires a number of extrapolation, interpolation, number of techniques to derive the exact results for interpretation like elasticities of demand or even supply. So, those are little advanced aspects of modulating the data.

So, we are actually targeting ourselves to the proper and systematic or structured order of a particular data so that some interpretation and conclusion can be derived. Our module is developed because our module is actually attracting, because of the fact that handling these unit level data in a single platform is actually missing. There is no holistic approach by any particular institution or by the ministry which gives a systematic direction to the management of this kind of data. So, even neither in the YouTube platform nor in a single digital platform where students are actually motivated or guided completely to understand large scale unit level data.

So, in this pretext I wish to introduce the complete direction of this particular course. This course is actually framed in understanding the unit level data particularly the National Sample Survey of some rounds and India Human Development Survey (called IHDS). There are two rounds, in fact three but we will be dealing with two. These are actually called internationally acclaimed data particularly surveyed in Indian context. Another one is in the context of health issues called NFHS (National Family Health Survey) as part of the demographic health system or demographic health survey data.

So, the module is actually structured into 8 weeks and in the 8 weeks will have different sections like, we will understand first, what is called data, what are different setups, what are the structure of data, then we will understand how to extract the data and then we will also understand how to mine those data which are the layouts and how those layouts actually interpreted, how we can filter, if there are some missing values also, how we can, simplify those missing values and convert to a right interpretation.

(Refer Slide Time: 07:20)



Then in addition to that we also added, certain sections called factor analysis and understanding regression tools like understanding the linear regression approach and also nonlinear regression approach. So, majorly it is confined around linear and within the linear we will first of all check certain basics before introducing the linear regression and its interpretation. We will identify multicollinearity, we will talk about heteroscedasticity. We will also discuss pooling of the data. We also discuss panel form of it. So, it is the fact in our last module is on understanding panel and pooled. So those are overview of our entire segment.

In the very first lecture we wanted to discuss about what we mean by unit level data. So, the first class is therefore defined as Introduction to Data then our second class will be on Unit Level Data. So, what do we mean by data? What is data? So, as you can see from this picture these are actually going to be used throughout our entire module. So, these are going to be used, like if you see it, here we have IHDS, here we have panel component, here we have extraction, here we have STATA use and we have NSS data. We have some APPEND issues, some merging techniques, normalizing data so that we can make more representative and can be discussed in different context.

(Refer Slide Time: 09:02)

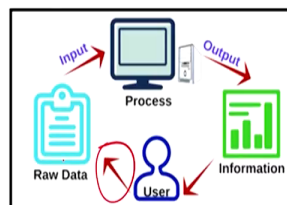
- ❑ Data is a gathered body of facts, figures, statistics used for reference or analysis.



So, let us talk about data in different forms. So, data is actually called a gathered of information of figures, facts, statistics used for references or analysis.

(Refer Slide Time: 9:18)

- ❑ Information in numerical form that can be digitally transmitted or processed.
- ❑ Data is the central thread of any activity.
- ❑ Numbers, characters, symbols, images etc., which can be processed by a computer.



so far as data is concerned we interpret as information in numerical form which are digitally transferred or processed. Data are generally centrally thread and leading with certain forms of information. So, it is called as central thread of any activity; and data is actually interpreted as numbers, characters, symbols, images and which can be processed by a computer or with a digital platform.

So, we have mentioned here that it is in fact a circular process. It starts with the individual. It ends with the individual. We start with certain information or extract information from the particular individual, their characteristics of the individual and we get the raw form. So, this arrow is important. We find the raw information then we structure it through certain inputs to the computer and with certain techniques, with certain tools the computer generates relevant outputs.

From the output we again interpret and narrate the story or narrate the information for a certain purpose, and those purposes are actually useful for the consumer, or here we defined as users. So, it is a circular process. So, data starts with the person, ends with the person. So, this is very very important so far as data is concerned.

(Refer Slide Time: 10:57)

Data must be interpreted, by a human or machine to derive meaning, else data is **meaningless**.

Examples:

Yes, yes, no, no, yes, no
75, 82, 33, 59, 98, 12, 22
111124, 101010

None of the above datasets have any meaning until they are given a context and processed into a useable form.

Understanding the nature of data is most fundamental for proper and effective use of statistical skills.

5

What else are important on the very first or beginning of our lectures, so data must be interpreted by a human or a machine to derive meaning else data is simply meaningless. If you don't derive any meaning out of it, may be through the human or through the machine, if no systematic meaning or information is derived, it is simply meaningless. So, it is not called data. So, basically what do we mean here? We actually say it is interpreted as meaningful information.

Now for example, we have some data, may be structured here. You can look at, like yes, yes, no, no, yes, no, again, 75, 82, any number, digits are given. If it is not systematically presented, there are number of question marks. We do not derive any interpretation out of it. So, none of these

datasets actually explain meaning or give any meaning until they are given in the context or processed into a useful form. Or if they are not actually discussed correctly these are not usable. So, understanding the nature of data is most fundamental for proper and effective use of statistical skills.

(Refer Slide Time: 12:23)

VH1104Q1V1101210110101101105274211020821	6349	300000420170823	135	1	2	5081274
VH1104Q1V1101210110101101105274212010811	7212	130000120170824	120	1	2	3387514
VH1104Q1V1101210110101101105274212020811	5212	140000220170823	120	1	2	3387514
VH1104Q1V1101210110101101105274212030831	6149	150000320170823	130	1	2	3387514
VH1104Q1V1101210110101101105274212040841	6512	150000320170823	125	1	2	3387514
VH1104Q1V1101210110101101105274213010831	5212	150000220170824	135	1	2	18631314
VH1104Q1V1101210110101101105274213020821	5312	200000220170823	125	1	2	18631314
VH1104Q1V110121011010110110528711010821	6319	150000420170818	125	1	2	10701584
VH1104Q1V110121011010110110528711020822	6199	80000120170818	125	1	2	10701584
VH1104Q1V1101210110101101105287112010821	4319	100000120170818	112	1	2	3132174
VH1104Q1V1101210110101101105287112020821	6219	140000220170818	120	1	2	3132174
VH1104Q1V1101210110101101105287112030821	2319	350000220170819	80	1	2	3132174
VH1104Q1V1101210110101101105287112040821	4119	60000220170819	115	1	2	3132174
VH1104Q1V1101210110101101105287113010821	1119	20000120170819	65	1	2	3915214
VH1104Q1V1101210110101101105287113020811	4312	100000220170819	125	1	2	3915214
VH1104Q1V11010701101021011050959110108122	6512	150000120170829	100	1	2	10599764
VH1104Q1V1101070110102101105095911020811	5512	150000420170828	100	1	2	10599764
VH1104Q1V1101070110102101105095912010811	3912	120000120170828	60	1	2	3499924
VH1104Q1V1101070110102101105095912020811	4912	150000120170828	70	1	2	3499924
VH1104Q1V1101070110102101105095912030821	4329	150000220170829	80	1	2	3499924
VH1104Q1V1101070110102101105095912040841	2519	80000220170829	60	1	2	3499924
VH1104Q1V1101070110102101105095913010821	2512	70000120170829	100	1	2	15199664
VH1104Q1V1101070110102101105095913020811	3112	50000120170829	80	1	2	15199664
VH1104Q1V1101210110102201105646011010811	5319	100000320170821	100	1	2	10972794
VH1104Q1V1101210110102201105646011020811	4219	100000120170821	90	1	2	10972794
VH1104Q1V1101210110102201105646012010811	3319	60000120170821	70	1	2	3383284

Source: Original Database from PLFS 2017-18

Here we carry one sample screenshot of the very latest dataset. As I just discussed couple of minutes back that our data why we meant for introducing such a module for MOOC program because of the case that there are lots of hue and cry related to the recent evidences on employment and the number of questions targeted to the Government relating to policy failures specially on employment issues. It has been argued and debated that the Government is going by 45 years low employment.

Why should I believe as a researcher? If any information is coming and criticizing the Government, it is not necessarily as a researcher I will simply believe. I have to find a scientific route to it, scientific understanding to it. So, what do you mean by scientific here? Scientific is through the original information collected from the data, collected from the persons. Now what do you mean by original information? There are two ways of defining originality. One is, as if the person is directly observing. Another is, the person is not directly observing but there are set of institutions which observe the data and refine the data, interpreted the data so that we can

believe their interpretation because the methodology they adopted is actually very systematic. And they have minimized all possible errors.

So errors, I am referring to sampling and non-sampling error. The statistical division of our country mainly referred to as MOSPI which is responsible for collecting the unit level data at a large scale. MOSPI called Ministry of Statistics and Program Implementation. This is one of the best database in the world so far as coverage, by its variable, by its number is concerned and this has been very correctly set and observed by our different statistical division of the country and they validate the data by various forms. There are various reasons by which we may believe that this is one of the authentic data source for the individuals.

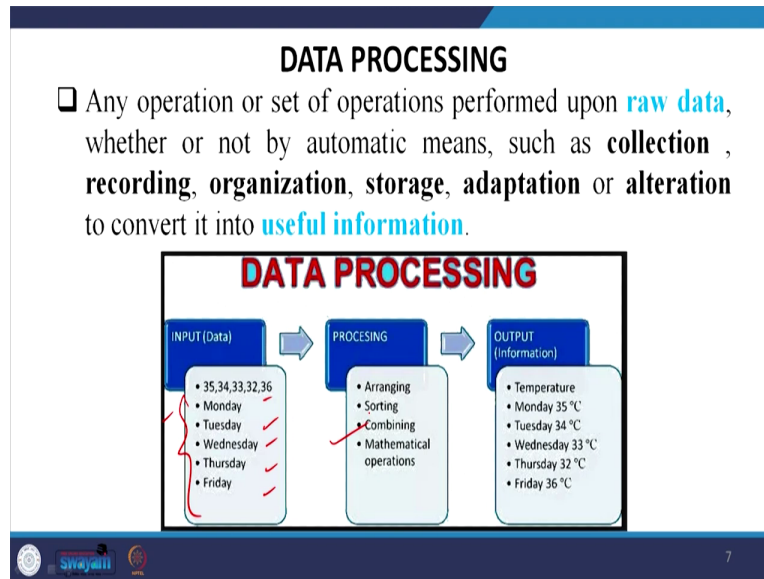
I am referring here the latest dataset called PLFS (periodic labor force survey), studied in the year 2017-18 and this database is called the most important source for understanding employment scenario in India. And usually the earlier rounds of the dataset used to be defined as Employment Unemployment Survey and they have changed the name to PLFS. So, we will clarify those details while will have separate sections on NSS, at this moment I am just introducing why we go for this kind of digits and what those digits represent?

So, you can see, this is the snapshot from the original data surveyed by Government of India in the year 2017-18 and mentioned in a note file or in ASCII format. This has different byte space. I am just mentioning in dot form different byte space, with certain blanks in-between and the byte space and the byte space actually represents certain characters of the particular person.

These are actually particular individual. On the row different individuals, on the column usually we consider the information regarding the different variables. And also added here as, the beginning characters are defined as unique characters which is similar in all the individuals or all the blocks. In the later parts, we will explain all those things in detail, but at this moment I am just introducing.

If this is the dataset given by the Ministry and if I am a student of PhD, I wanted to understand the nature of the labor force in our country, so how should I go for it? So, our module is going to give you a systematic presentation of all those digits and their byte spaces. In our due course of time we will certainly emphasize everything in detail.

(Refer Slide Time: 17:13)



In our third lecture, we are going to discuss the first part of this Notepad file. Now let us understand how to process it. Usually which form we do process data? these are actually called raw data which I have shown you are called raw data, and any operational set of operations performed using raw data whether or not by automatic means such as collections, recoding, organizatios, adaptation, storage or alteration, to convert it to a meaningful information is defined as a useful information.

So, usually there are 3 pillars of processing. First is input, then processing, then output. Some examples are there in the inputs, maybe of different informations are there for our understanding. These are the inputs, we may sort it, we may combine, we may enter certain mathematical relationship, mathematical equation to combine or arrange and accordingly we interpret the data with our outputs.

(Refer Slide Time: 18:17)

DATA NEED TO BE TURNED INTO MEANINGFUL INFORMATION

- ❑ Information is different from data; **information is the result of processing.**
- ❑ For example: **22081992**, there is no meaning of these numbers without processing, it can be interpreted as a birthday, an account number, an Aadhar card number, a mobile number, ATM card number etc.

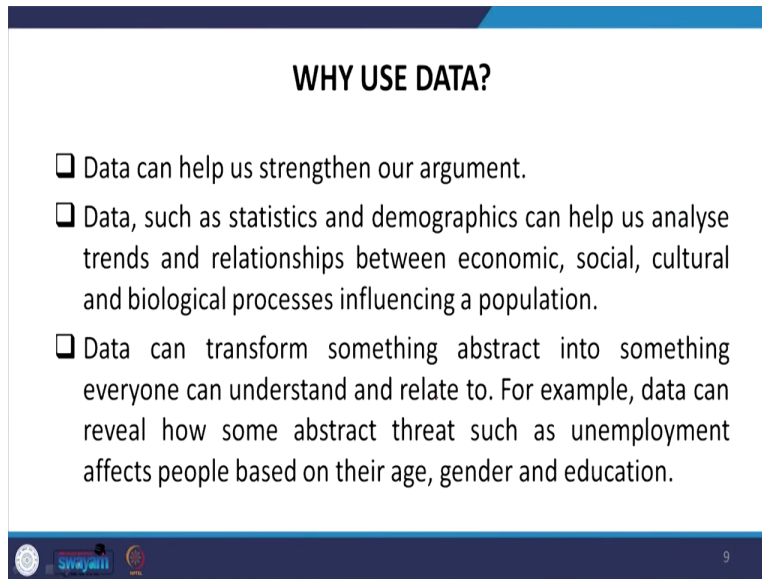
```
graph LR; A[Raw data (4.9, 5.2, 6.0, 5.8, 6.2...)] --> B[Context (Height of employees in an industry)]; B -- Processing --> C[Information (average height of employees in the industry 5.6)];
```

8

So, information actually is important. Systematic information is important. So, data needs to be turned into a meaningful information and the information is the result of processing. Now for example, if we have a number called 22001992. So, does it mean anything unless we divide it and interpret in different order? Actually there is no meaning of these numbers without processing. So, it may be a birthday, it may be a composing of an account number, maybe an Aadhar Card number, may be an ATM Card number, so systematic presentation is required.

Let us have an example. If the raw data compose of 4.9, 5.2, like this and so on, if we put it in certain interpretation with variable called height of the employee in a particular industry, then it gives certain meaning. Just 4.9, 5.2 actually meaningless unless we don't name it, unless we do not code it with particular information like height. Average height may be meaningful, median height might be also meaningful and if we make it in a group which height has highest frequency that is also. If you go by certain modal value to this, that will also be important. So, processing and generating certain output is actually converting to a meaningful information. Now how we use data?

(Refer Slide Time: 20:00)



WHY USE DATA?

- ❑ Data can help us strengthen our argument.
- ❑ Data, such as statistics and demographics can help us analyse trends and relationships between economic, social, cultural and biological processes influencing a population.
- ❑ Data can transform something abstract into something everyone can understand and relate to. For example, data can reveal how some abstract threat such as unemployment affects people based on their age, gender and education.

9

So, data can help us strengthen our argument and data such as statistics, demographics can help us analyze trends, relationship specially economic, social, cultural and biological processes influencing population. Usually these are important for representing our population and a large segment can be interpreted and discussed. So, data actually transfers the information into somewhere meaningful information and those we have already discussed. So, in the context of unemployment we are interested to understand their age, their gender, their education, their blah-blah other information.

(Refer Slide Time: 20:39)

DIFFERENT FORMS OF DATA

- Alphanumeric Data (combination of numbers and letters, also called String data).

Examples of Alphanumeric
• "I have dogs"
• ABC123
• SR67JH8
• 100 miles
• madeup@gmail.com
• 13 Cairo Street Egypt
• 50 pounds
• 20°C

Data are actually sometimes or often the case converted in string form. In the raw data or even after processing also, we will see certain string form. What do you mean by string form? It is a combination of numbers as well as letters. So, we are given certain example of string form. These are also called alphanumeric data. Suppose I say I have a dog and ABC, 123 and so on. These are in string form.

(Refer Slide Time: 21:15)

The screenshot displays the STATA Data Editor interface. The main window shows a dataset with columns: ENTPID, PPRcode-d, v111_010_Step, Round, Sample, Sector, State_Region, State, State_District, and State_Region. The ENTPID column is circled in red. The Properties window on the right shows the variable ENTPID with a label 'Primary key - v111' and a format of '%30s'. The Properties window also shows the variable's type as 'string', its format as '%30s', and its label as 'Primary key - v111'. The Properties window also shows the variable's format as '%30s' and its label as 'Primary key - v111'. The Properties window also shows the variable's format as '%30s' and its label as 'Primary key - v111'.

I will show in the exact shape in STATA software. This is our snapshot from the original data of PLFS, original data of one of the rounds of the NSS where the dataset gives certain information.

If you click somewhere here, in a box it is clicked. if you want to check on your own, I suggest that you install STATA. You request a demo version from the STATA software. You get the user for may be one week, may be for one month, whichever way they cooperate, you go through that and try to install from the beginning.

You will find out this type of layout, this type of pattern and this is the data storage pattern in STATA where if you open that box, you will have these informations. I have clicked here. the arrow mark shows that it is written as STR 30. STR 30 stands for string 30. That means 30 characters space is actually occupied in that particular box in that particular column. So, beyond 30 it is not allowed. So, we can see next to that, a format is written called 30 percent. That is, out of the total space, 30 percent is occupied with our information as per the data of NSS. So, other details we will understand and go through it in our successive module.

(Refer Slide Time: 23:01)

The slide contains the following text:

- Text Data (sentences and paragraphs used in a written communication).
- <http://www.pressure.to/qda/>

Text Data

+ Meta data at each level!

The diagram is a pyramid with the following levels from top to bottom:

- Multi-corpora
- Document Collection
- Document Clusters
- Document
- Chapter
- Section
- Paragraph
- Sentence
- Multi-word collocates
- Word
- Letters

Red circles highlight 'Multi-corpora', 'Document Collection', and 'Letters'.

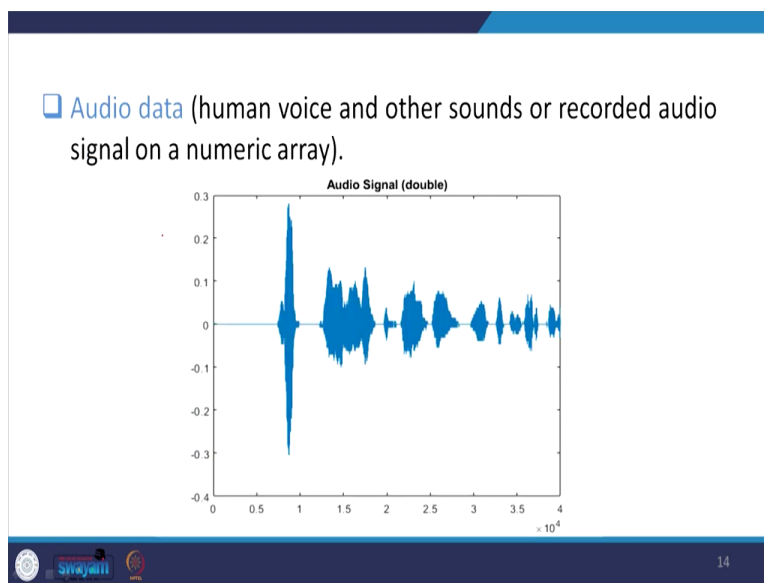
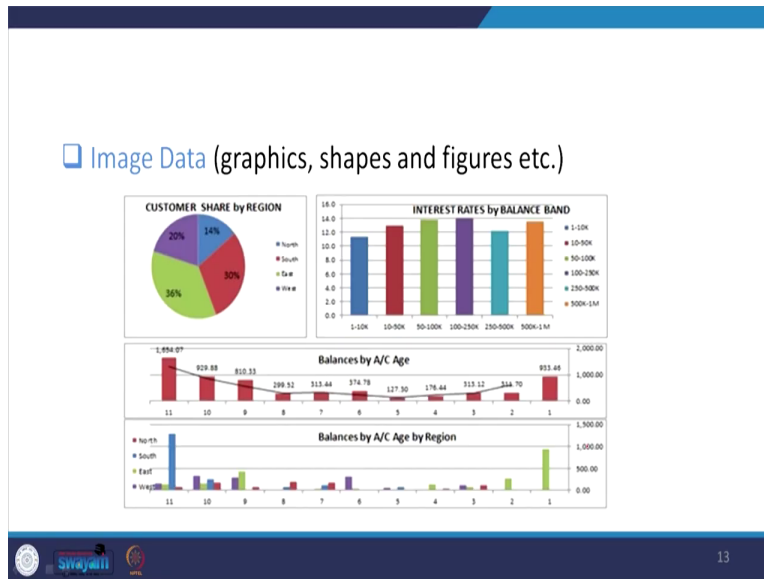
At the bottom of the slide, there are logos for 'Swayamii' and '12'.

Now let us understand what other forms of data are there. One we have already discussed called string. Second one is called text data. Now text data are defined as sentences, paragraphs used in a written communication. And these days, in a different format of research, people call it as metadata after simplifying the data, processing the data they express in a form called metadata.

Metadata is also called data of data. And these may be, though expressed here in a pyramid, say because of their information, if it is text we have to actually make it in a certain shape so that we

will say that very less percentage of them who talks about this variable, some of them may talk about document collection. There may be larger set related to the last variable. There are various aspects. Some information of the text data can also be extracted through, a web link is given here called, you can download the software called WEFT, where some basic keys are very very useful in getting information related to text data and their research.

(Refer Slide Time: 24:23)

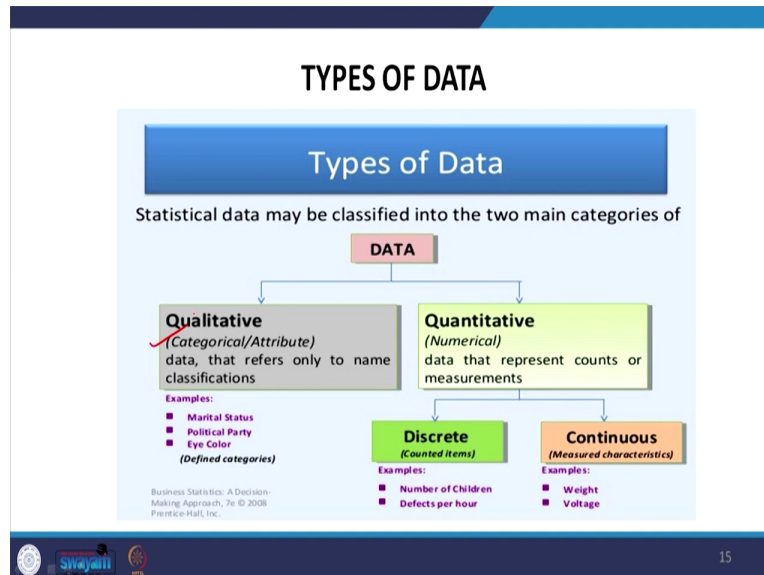


Other forms of data are called image data, now it is quite understood to everyone. I need not spend much time. From various form, may be from pie chart and bar chart, line diagram, we have lots of information related to image-related discussion. So, accordingly many interpretation could be made.

Another set is called audio data. I think you might have seen in different songs, in different volumes. It has different amplifying effect and those frequencies are actually important. And

they do control these frequencies. If the frequency is in the balanced format that means your voice is actually recorded very correctly. So, I am just giving an interpretation here.

(Refer Slide Time: 25:17)



Let us come to the discussion on types of data. After understanding different forms of data and its background to data and why we try to relate to data of large scale in nature. We are discussing different types and their structure. Broadly, it is divided into quantitative and qualitative. Qualitative data is largely used by the researcher from the field of social sciences, especially the sociologists and anthropologists because many informations are not coded or codified. We cannot have numbers of it. Some phrase of the analysis like some person is narrating his or her story, the story itself, the 2-3 lines are actually the code, not just a single number can be interpreted from that particular phrase.

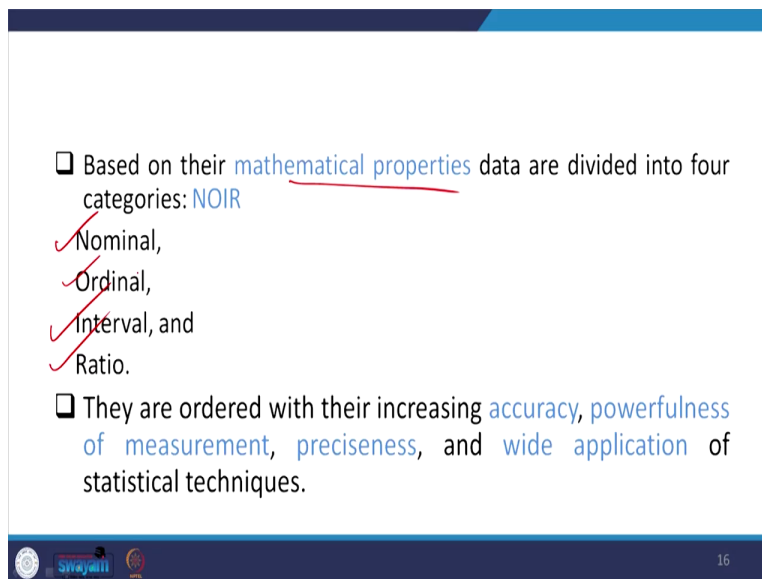
So, the story itself carry a qualitative information. So, for our simplicity we can make it different forms of data. Therefore we are discussing why qualitative and how these are useful in our analysis and how NSS or IHDS or NFHS have dealt those qualitative information. Though they are qualitative in nature but we will explain in terms of quantitative explanations. So, that is the beauty of the dataset.

Qualitative data are generally called categorical data or attribute-based data. And these are generally marital status, belong to a political party, eye color and gender of a person. Gender for

example if I say 1.2, it does indicate nothing. Gender is 1.2, either we say a complete number 1, 2 or even any number if you say but beyond that, 3, 4 I think it is not defined. But interestingly if you search on Facebook there are 17 varieties of gender defined. Now look at how qualitative informations are actually discussed, not just 2 or 3, transgender, beyond that various forms of gender-related information can be defined.

Now let us come to the discussion on quantitative aspects. Quantitative are generally of two types, discrete and continuous. So, discrete are generally counted. These are called counted items. And continuous, having certain major characteristics.

(Refer Slide Time: 27:58)



- ❑ Based on their mathematical properties data are divided into four categories: NOIR
- ✓ Nominal,
- ✓ Ordinal,
- ✓ Interval, and
- ✓ Ratio.
- ❑ They are ordered with their increasing accuracy, powerfulness of measurement, preciseness, and wide application of statistical techniques.

So, if you discuss those things in another approach, in mathematical, through their mathematical properties we say nominal, ordinal, interval and ratio. Broadly they called N O I R, nominal, ordinal, interval and ratio. And they are ordered with their increasing accuracy, powerfulness of their measurement, preciseness, wide applications especially in statistical techniques.

(Refer Slide Time: 28:25)

NOMINAL DATA

- ❑ Nominal means **name** and **count**; data are **alphabetic** or **numerical** in name only.
- ❑ They are categories **without order** or **direction**.
- ❑ Their use is restricted to keeping track of people, objects and events.
- ❑ They are **least powerful** in measurement with no arithmetic origin, order, direction or distance relationship.
- ❑ Hence nominal data is of restricted or limited use.

17

EXAMPLES OF NOMINAL DATA

18

Nominal data means names or count. Data are alphabetic or numerical in name only so far as nominal is concerned and they are categorized without order or direction. There is no such direction so far as nominal is concerned. Like as I told you for gender, there is no direction. Male, or female, whether male should be coded as 1 or female should be coded as 1, so there is no clear direction on it. If you are doing it for 1 every time for male, probably you are committing some mistakes while interpreting.

So, we must be very careful enough where to give, which label of weight. But if you are simply taking code by 1 and 2, but I have taken average of it, actually it will be biased towards 2. So, average is not at all meaningful. So, we must be very cautious of coding those data and their interpretations accordingly. And the central tendencies of the statistical technique used in those regard is very very important. So, they are least powerful in measurement with no arithmetic origin or order, direction or distance, relationship. This is example of nominal data. I have already discussed those. You may go through these.

(Refer Slide Time: 29:52)


ORDINAL DATA

- Ordinal means **rank** or **order**.
- Ordinal data places events in order; they are ordered categories like ranking or scaling.
- Ordinal data allows for setting up **inequalities** and nothing much.
- Adjacent rank needs to be equal in their differences.
- Has no absolute value. (only **relative position** in the inequality).
- More precise comparisons are not possible.

19

intellspot.com

EXAMPLES OF ORDINAL DATA



The first, second and third person in a competition	Economic status: low, medium and high	Letter grades: A, B, C, and etc.	Education level - elementary, high school, college graduate	Customer level of satisfaction such as satisfied, neutral, dissatisfied
---	---------------------------------------	----------------------------------	---	---

20

So, ordinal data are usually of rank and order. It has a clear ordering of the particular context for example if we put it, the ordinal data, this is very very useful for inequalities measurement because it has a clear ordering. The ranking of a person, it is an ordered data. Wealth of a person, if it is indicated in a ranking called ordinal data. And so relative position of those data is very very important. And it has no absolute value because it is an ordering, no question of absolute value is defined. So, example of ordinal data is like first, second, third position of a person, economic status as I discussed, so on and so forth for further understanding.


(Refer Slide Time: 30:40)

INTERVAL DATA

- Interval data in addition to ranking (setting up inequalities), further allow for **forming differences**.
- For interval data there is **no absolute zero**; unique origin does not exist.
- Interval data are more powerful than ordinal scale due to equality of intervals.

21

Interval Data



22

Other two categories are interval and ratio data. Interval data is important because there are some data defined to be in interval form. And no absolute zero is actually defined so far as interval is concerned since we are sticking to interval, so no question of absolute 0 is defined. Interval are more powerful than ordinal scale due to equality of intervals. So, I think you might have seen these temperature scale, and especially interval in that scale is very very important.


(Refer Slide Time: 31:15)

RATIO DATA

- Ratio data allows for forming **quotients** in addition to setting up inequalities and forming differences.
- All mathematical operations (manipulation with real numbers) are possible on ratio data.
- It can have an absolute or **true zero** and represent the actual amount/ value.
- The most precise data and allow for application of all statistical techniques.

23

AGE STRATIFICATION



The slide contains three illustrations. On the left, a woman and a young child are standing against a blue background, with the child reaching up. In the center, a hand is shown holding a silver kitchen scale. On the right, a group of cartoon characters of various ages, from a small child to an elderly man with a cane, are standing together. The text 'AGE STRATIFICATION' is written above the group of characters.

24

And last one to be discussed in this particular first lecture is on ratio data. Ratio data is generally called widest application in statistical techniques. It has absolute 0 value and 0 has certain

meaning to it. And let us combine example. If we talk about their height, their weight of the person and the age of the person it is clearly given and the age of the person is highlighted here very clearly.

(Refer Slide Time: 31:53)

SCALES OF MEASUREMENT

Operation	Nominal	Ordinal	Interval	Ratio
Equality	✓	✓	✓	✓
Order		✓	✓	✓
Add / subtract			✓	✓
Multiply / divide				✓
Mode	✓	✓	✓	✓
Median		✓	✓	✓
Arithmetic mean			✓	✓
Geometric mean				✓

The table is annotated with red circles around the 'Interval' and 'Ratio' columns, and a red line connecting the 'Arithmetic mean' row to the 'Ratio' column.

- ### META DATA
- Commonly Defined as **'Data about data'**
 - Summarized data that leads to detailed data.
 - Identifies the parts of the digital object and documents their content, location, relations, structure and functionality.

- These data are:
 - a. Date of survey
 - b. Time of survey
 - c. Investigator's detail
 - d. Editor's details
 - e. Response code
 - f. Reasons for non-response
 - g. **Data type and data attributes** etc.
- Sometimes meta data offers valid explanations for unexpected and unusual answers.

What kind of scales are important for measurement? It is very indicative, very clearly mentioned which methods of measurement are very appropriate. ratio scale data is concerned all the methods are important, whereas interval is little better than that of others. Ordinal is even better than that of the nominal. So, accordingly we can understand.

Let me give you one example here. Average as I told you, so arithmetic mean is applied in two cases. Similarly mean and mode, you can find out on your own, we will discuss these details in our other explanation. So, metadata we already discussed, data about data. We will take it forward in one of the sections on this, so let me stop here. Thank you so much.