Exploring Survey Data on Health Care
Prof. Pratap C. Mohanty
Department of Humanities and Social Sciences
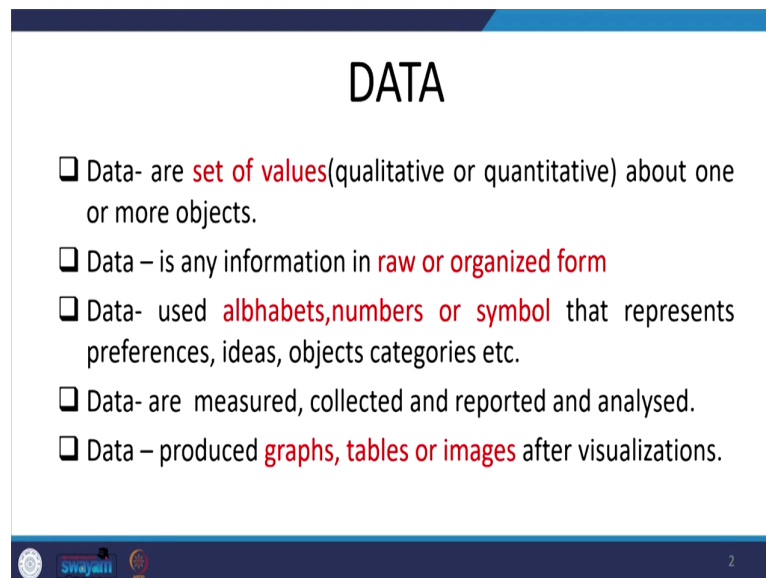**Indian Institute of Technology, Roorkee**

Lecture - 02
Data, Measurement and Scaling

Welcome friends to this module on health survey data. We are on the 2nd lecture, and it is on Data, Measurement and Scaling. We have already discussed about the research problems, the very foundation of research hypothesis building, how we start with a basic research, what are the different types of research problems, we already discussed in the last lecture.

In this lecture, we will clarify with you about different types of data, their measurement and their scaling. Why are we discussing this? Because this is very useful especially in filtering the data, which scale of measurement we require in the data.

That pre-filtering analysis is very essential before any sort of inferential statistics we derive from the data. Now, data, as we all know that these are a set of values, both qualitative and quantitative about one or more objects.

(Refer Slide Time: 01:46)

So, data is all about different sets of values, and this gives information in raw, or in an organised format. So, data use are usually alphabets, numbers, symbols and represents preferences, ideas, objects or their categories etc.

These are measured, collected, reported and analysed. So, data actually accordingly produces graphs, tables, useful images with their appropriate visualisation. So, this is where now the data which we are going to deal with which we have already dealt in my previous module on NPTEL. But this time, it is more applied based, and it is more health-related research and where we will be using unit level data IHDS data to some extent then will be use STATA software for the analysis. And regarding health data, these are all about observations related to health conditions reproductive outcomes different causes of death or quality of life of the individual or population.

(Refer Slide Time: 03:20)
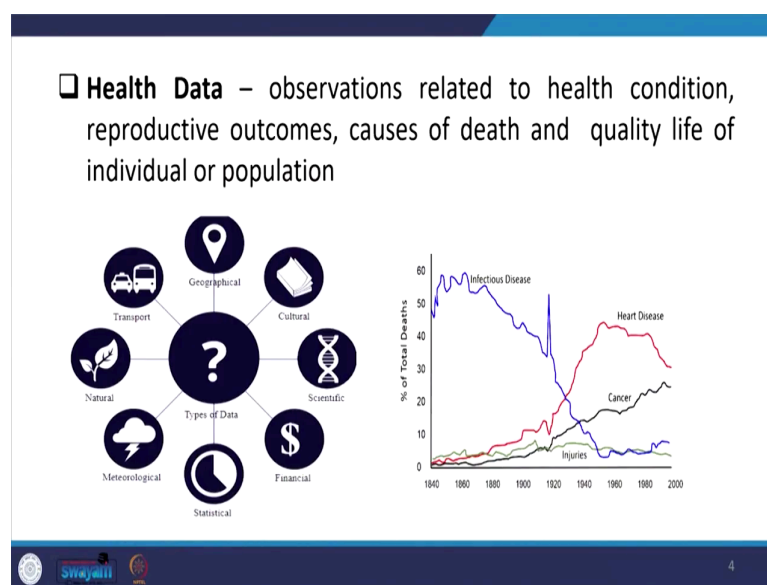
This ranges from health data include geographical factors, cultural, scientific, financial, and then statistical, meteorological, natural, transport related issues. So, there are various forms of data connecting to each of the factors. Now, from the data, we may draw certain directions, certain frequencies about the occurrence of heart diseases, cancer, injuries, infectious diseases etc.

(Refer Slide Time: 03:55)

Now, what do you mean by dataset? Dataset is an ordered collection of data. When the data is collected in a order and they are called dataset, the data point basically singling out the particular observation or a particular point. All right like here in that example about use of medicine traditional and modern medicine, we have different variables like modern medicine or traditional medicines. Within the dataset we have discussion about observations, data point and data point basically the exact entry maybe in 6.37 or 7.54 etc. are called data points. What do you mean by raw data? This is basically preprocessed data or uncleaned data.

Just in the information that is before proceed to analyse or clean and or before organise or analyse are called raw data. There are number of open-ended responses to a survey, some survey responses are there.

(Refer Slide Time: 05:18)

So, a particular value might have been entered if those are all put to in different forms are called raw data. On the screen I think we have already shown you earlier or in my previous module this is in fact, called the raw data. All the bytes and their codes are entered.

We are referring to NSS 75th round raw data, published by MOSPI, MOSPI Government of India; Ministry of Statistics and Programme Implementation. Ministry of Statistics and Programme and Implementation gives these forms of data. So, a sample of information is given like health expenditure of nine individuals per year and their health expenditure data is presented. So, at this moment these are complete in raw form. We cannot guess anything out of it.

(Refer Slide Time: 06:23)

Criteria for Describing and Evaluating Dataset

Criteria	Implementation Guidance
Purpose	Identify the purpose
Nature	Definition of each element, Unit of measurement, specification of variables categories
Specification & Methodology	Data collection method, Sampling techniques and size, Questionnaire design
Accuracy & Correctness	Assessing accuracy and correctness by comparing data from different sources
Sources	Necessary to reduce the risk of misinterpreting the set of data during use
Time period	Time lag between collection and publication, Frequency of updates e.g. NFHS

So, coming to some criteria that describe and evaluate the dataset or are like you have to define the criteria. Those criterions are purpose of the data, nature, specification, methodology, accuracy or correctness, sources and time period. So, the time period has to be specified, the sources of the data have to be specified and the exact methodology which is used or specified has to be very clearly understood.

So, the objectives of the study should have been clearly spelt out. So, coming to the nature of the data we may define each of the elements within the data. Like their unit of measurement their specification of variable and their categories. And in the questionnaire or in the methodology we are supposed to talk about the sampling techniques their size their questionnaire design.

And so, like in the sources we are supposed to find out to talk about the exact source of the data. It is necessary to reduce the risk of misinterpreting the state of data during use. So, in the time period I think all the database we always refer to the time period and in due course of time I will let you know about its time periods.

(Refer Slide Time: 08:02)
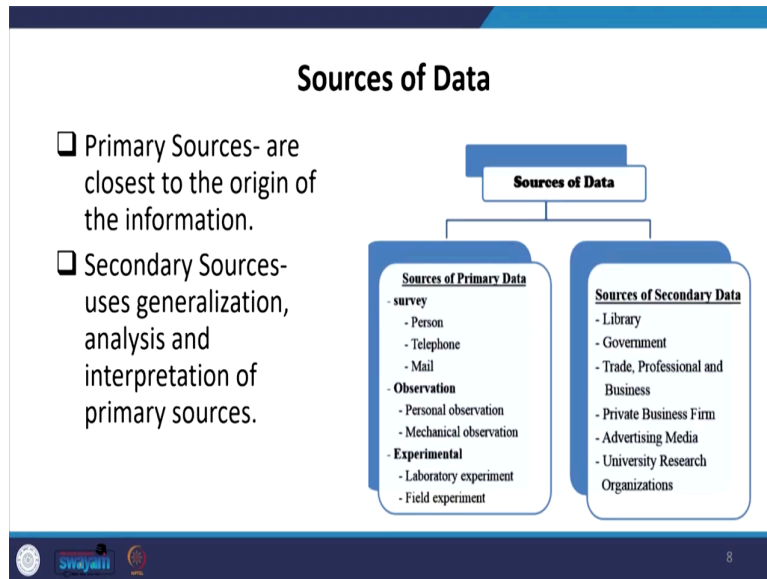
Coming to the sources of data. So, regarding sources of data, we have two types. One is called primary second one is secondary. So, the primary data I think we know that we usually follow survey methods or we directly with the individual with the person or by telephonic method or by E-mail approaches and sources of primary data could be your observations by personal observation, by mechanical observation, by laboratory testing or by laboratory experiment or field experiment.

And in the secondary data sources are like if you are collecting from the library or from the government sources or from the trade or professional bodies, private business firms or advertising media, university research, organisations etc.

So, basically, the primary sources are the closest ones so far as the information from the direct person or the respondent is concerned. Whereas in the case of secondary one these are usually for some forms of generalisation some forms of interpretation of the primary sources.

(Refer Slide Time: 09:29)

Sources of Some Health Database

- National Sample Survey Organization – Ministry of Statistics and Programme Implementation(MoSPI)
- The National Family Health Survey – Ministry of Health and Family Welfare(MOHFW)
- Longitudinal Ageing Study in India – International Institute for Population Sciences(IIPS)

And coming to the sources of some health database since we are going to use it in our research, in our discussion in our succeeding lectures. So, those are like National Sample Survey Organization, NSS published by Ministry of Statistics and Programme Implementation, it is called MOSPI.

The second one is called the National Family Health Survey (NFHS) published by Ministry of Health and Family Welfare (MOHFW). And the third one we are more interested to explain to you and since these databases have more dimensionalities. So, like we have another one is called Longitudinal Ageing Survey in India. These are also called LASI; LASI published by the International Institute of Population Sciences.

So, about the type of data and their nature, we already discussed that there may be primary or secondary. Now, we are discussing on the basis of their nature, which nature it may they may be categorical or qualitative or maybe numerical or quantitative. In case of qualitative one, they deal with the description, some attributes data can be observed, but not measured.

(Refer Slide Time: 10:59)
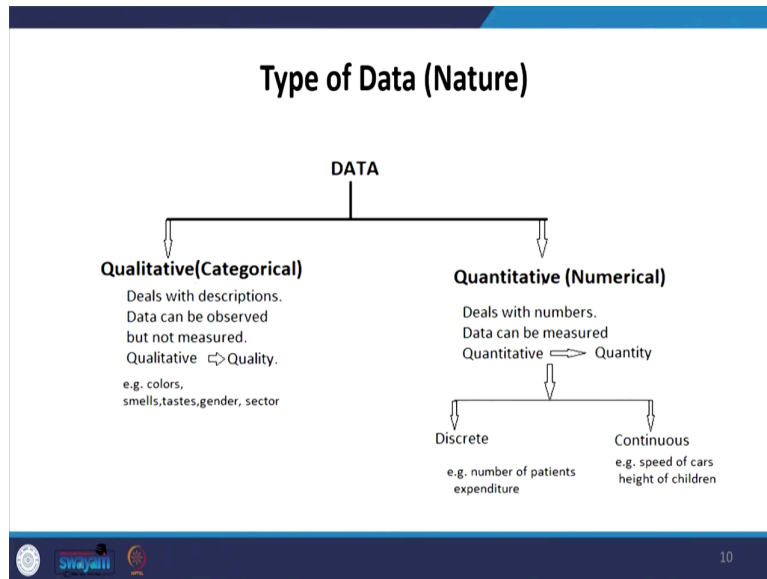
Those are called qualitative data or having certain qualities or colours, smell, taste, gender, sector etc. So, those are not quantified as such. We may code it, but codings are not called quantification. Whereas, coming to quantitative data we have particular numbers. They represent the context; data can be measured and can be scaled correctly.

So, again the quantitative data are of two types. One is called discrete and the second one is continuous. Discrete one, where the discrete choice function is defined, like the number of hospital consultations or the number of doctors consultations. It is not converted with certain enough decimals. It cannot have a clear cut what is called continuous series. So, it is clearly a discrete one. Either one consultation maybe five consultations, it cannot be a 5.2 consultation. Similarly, coming to continuous data speed of cars, height of the children; maybe in age not in round figure by its years rather age in any number may be of years, month, days, time, there are lots of possibilities. So, even income of the person it's possible to define in fraction. So, this is all about called quantitative data.

(Refer Slide Time: 12:42)

Type of Data (Time and Entity)
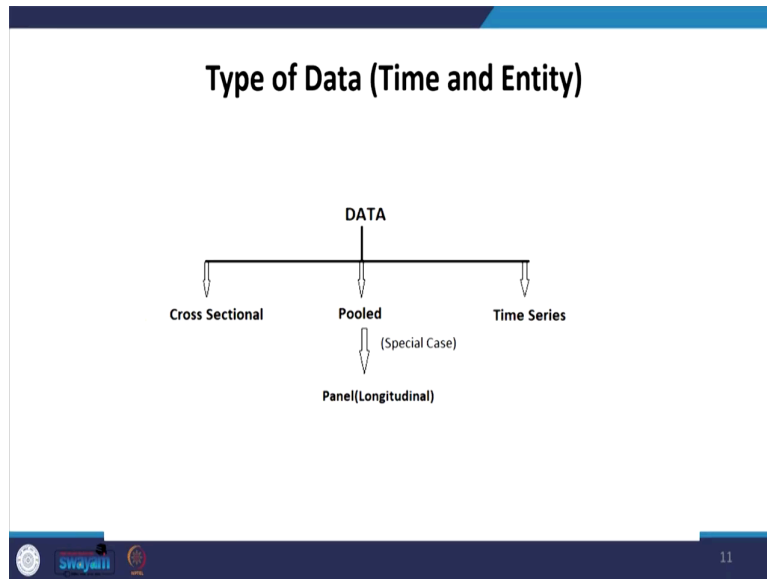
Now, based on time and entity data can also be divided into three broad category. One is called cross sectional, second one is called pooled, third one is called time series. So, cross sectional, time series and the middle one is called pooled where both the type of data are actually are entered like the cross-sectional observation, individual behavioral observations are presented at a particular period of time whereas, the context in case of time series is actually repeated over time.

So, we are basically doing what in the pooled, we are actually mixing these two together cross-sectional and those cross-sectional units are represented at different time periods. The special case of pooled data is actually called panel. So, all pooled are not necessarily panel. If the time component does not represent the individual person over time, then that cannot be called a panel. Rather it is simply called pooled because in the pooled the different cross-sections might be added in different time periods. But we are concerned in panel or the longitudinal data where the same person same observation is repeated in different time period.

(Refer Slide Time: 14:05)

> ❑ **Cross Sectional Data** : are data on one or more variables collected at the same point in time.
> Example
> ➢ NSSO: Social Consumption Health Survey Data etc.

Entity	Year	Nature of treatment (traditional=0, Modern=1)	Health expenditure(Rs.)
A	2021	1	2500
B	2021	0	1200
C	2021	1	2200
D	2021	0	1500

Coming to cross-sectional data, I think I have already mentioned that it is the same point in time when the variables are collected or the observations are taken. In case of our 75th data on social consumption of health or Social Consumption Health Survey data, we have entities like A, B, C, D or E, F, G, H those represents persons and the time period is the same that is 2021.

So, if it is 2021 or if it is of 1516 then or 1718 in fact, this if you are referring to 75th data you have to stick to the particular year. Now, this is for simplicity, any year can be taken. So, similarly about variable there could be any entries and we will find out what is the relevance of those entries.

(Refer Slide Time: 15:06)

About time series data is a set of observations on different values that a variable takes at different time. So, those time series data are collected at regular intervals such as daily, weekly, monthly or annually. Such examples are stock prices, weather report unemployment rate etc.

Whereas, the pooled one we have already said is the mix elements where both time series and cross-sectional units are present. In pooled data different cross-section entity is surveyed over time.

(Refer Slide Time: 16:03)

So, like we are saying, this is the same period and entities are studied in two time period. This is in the same year, two persons are studied and another year another two maybe A and C are the same person. If the same then it is completely called a panel data.

(Refer Slide Time: 16:31)

Pooled data this is what we are explaining in this example. Pooled data basically, you are adding all the cross-section units. A and C are not the same. So, B or D are not same, but time period we are simply pooling the two time period, 20 study and 21 study.

But in case of panel data A is present in 2020 and it is also available in 2021. Similarly B in two time periods. So, then one of the example dataset we are going to take is from LASI; Longitudinal Ageing Study in India.

(Refer Slide Time: 17:19)

So, different forms of data are there like alphanumeric data, text data, image data and audio data. So, examples tables from the stata file are here. We will also work with the exact dataset in the respective lecture where we will mention what is this alphanumeric?

Alphanumeric data, the famous example for it is our Aadhaar card, PAN card number. It has alphabets, it also has numeric numbers. And text data where only descriptions are given where only some sort of explanations are given on the text. Image data where picturisation or visualisations are made.

Audio data where you are supposed to listen and interpret, your voice signals are given based on their frequencies, some forms of interpretation could be done. On the screen, you have raw data from the 73rd round of a national sample survey on enterprises.

So, the point of importance here is that we deliberately want to show this figure to highlight this colour. Look at this color. These colors are in fact red. So, red indicates your data is in string form. So, black could have refer to your numeric data.

The string form while at the time of interpretation you we will convert it and then only any sort of numeric estimation would be possible. So, we will do some destring operations then our data will be refined for interpretation. I think in the startup window it is given this is called string data and it is shown, what is the type of data it is shown here.

(Refer Slide Time: 19:42)

Now, after giving the background about data, there are different types of data. Their nature by certain time by their coverage. Now we are moving to the explanation of measurement, what do you mean by measurement and how it is relevant in the module? So, measurement is basically the assignment of numbers or symbols to characteristics of objects according to certain pre-specified rules.

So, those assignments are especially numbers or symbols, which are simply called measurement. Quantifiable objects are easy to measure and highly accurate. Qualitative objects are quite difficult to measure and less accurate. Coming to one example called height, we measure it through certain scales, so that is part of our measurement.

(Refer Slide Time: 20:55)

Measurement Scales

❑ **Scaling** : The generation of a continuum upon which measured object are located.
❑ **Primary Scales of Measurement** :
➢ Discrete scales – Nominal, Ordinal
➢ Continuous scales – Interval, Ratio

Whereas in case of scaling, we are supposed to put them in a continuum format. The generation of a continuum upon which measured objects are located is called scaling. Primary scales of measurement are discrete scales or continuous scales. So, in discrete scales, we have already based on the discrete series of the data may be in categorical form may be in nominal form or in ordinal form.

So, nominal form has no orderings, whereas in case of ordinal form they have certain orderings. In the case of continuous scale, it is called interval data or it is also called ratio data. All sorts of measurement inferential statistics; mean, median, mode, standard deviation etc., we can do it on the continuous scale data, but it is not possible to do it in the discrete series.

Then when we define about the dataset, they are basic understanding about measurement. Now, we should understand the characteristics of measurement scales. What are the characteristics? The first characteristic is called identity. The identity means we need to refer to the assignment of numbers to the values of each value in the dataset.

The particular number which is assigned is important so far as identity is concerned. A serial number is important as well. Let's consider a questionnaire that asks for a respondent's gender with the options about male or female. The values are entered as 1 or 2 or at the reverse.

(Refer Slide Time: 22:55)

Characteristics of a Measurement Scale

- **Identity**
  - Identity refers to the assignment of numbers to the values of each variable in a data set
  - Consider a questionnaire that asks for a respondent's gender with the options Male and Female for instance. The values 1 and 2 can be assigned to Male and Female respectively.
  - Arithmetic operations can not be performed on these values because they are just for identification purposes. This is a characteristic of a nominal scale.
- **Magnitude**
  - The magnitude is the size of a measurement scale, where numbers (the identity) have an inherent order from least to highest.
  - They are usually represented on the scale in ascending or descending order.
  - The position in a race, for example, is arranged from the 1st, 2nd, 3rd to the least.
  - This example is measured on an ordinal scale because it has both identity and magnitude.
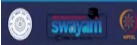
1, may stand for male, and 2, stands for female. The arithmetic operation cannot be performed on these values because they are just for identification purposes, not for measurement, and this is a characteristic of a nominal scale that we have just mentioned. Now, coming to another characteristic called magnitude, this is basically discussing about the size of the scale where the numbers have an inherent order from least to the highest. They are usually represented on the scale in ascending or descending order and the position in a race for example, is arranged from the 1st, 2nd or 3rd to the least. This example is measured on an ordinal scale because it has both identity as well as magnitude. So, the ordinal scale has identity basically of their position and their level where they lie. So, their magnitude is also defined in case of the original scale.

(Refer Slide Time: 24:11)

- **Equal intervals**
  - Equal Intervals means that the scale has a standardized order.
  - the difference between each level on the scale is the same.
  - This is not the case for the ordinal scale example highlighted above.
  - Each position does not have an equal interval difference. In a race, the 1st position may complete the race in 20 secs, 2nd position in 20.8 seconds while the 3rd in 30 seconds.
  - A variable that has an identity, magnitude, and the equal interval is measured on an interval scale.
- **Absolute zero**
  - there is an existence of zero on the scale, and is defined by the absence of the variable being measured (e.g. no qualification, no money, does not identify as any gender, etc.
  - Absolute zero is a feature that is unique to a ratio scale.
- **Levels of Data Measurement**
  - The level of measurement of a given data set is determined by the relationship between the values assigned to the attributes of a data variable.
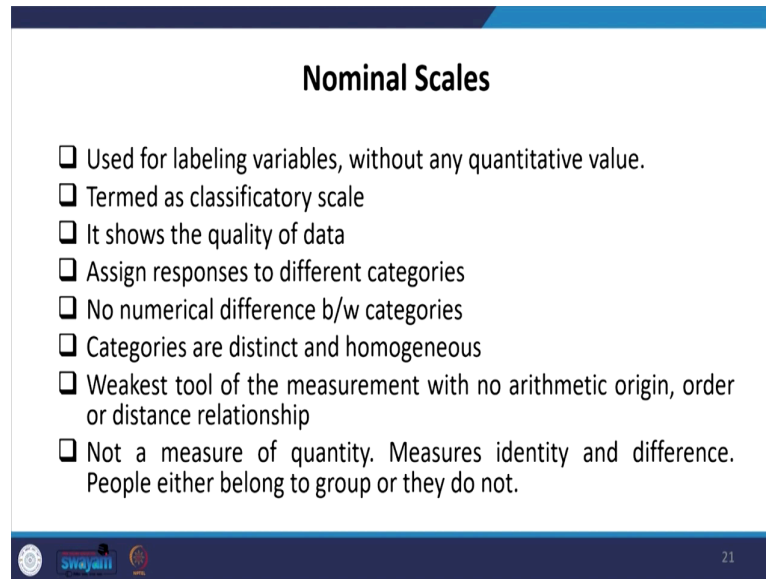
There are some scales called interval scales. What about equal intervals? Equal intervals mean the scale has a standard order. The difference between each level on the scale is the same. So, this is not the case for the ordinal scale as for the example highlighted before. You know where in case of positioning 1st, 2nd or 3rd, it may not be the fact that it has equal distance.

Like each position does not have an equal interval difference. In the race 1st position may complete the race in 20 seconds, the 2nd may be in 20.8 seconds, while the 3rd might have completed in 30 seconds. So, it is not necessarily the fact that these frequency points are not are of equal scale. A variable that has an identity magnitude and equal interval are measured on an interval scale.

Another characteristic is called absolute zero, which kind of data has the possibility of absolute zero and it has a certain value. There is an existence of zero on the scale. It is defined by the absence of the variable being measured, and that is no qualification, no money does not identify any gender etc. So, absolute zero is a feature that is unique to a ratio scale.

So, zero is not defined in some of the previous examples which we have cited. Like in the temperature scale for Fahrenheit or Celsius, we are not referring to zero values. It has a certain range and certain intervals. Now coming to the level of data measurement the level of measurement of a given dataset is determined by the relationship between the values assigned and the attributes of that particular variable.

(Refer Slide Time: 26:18)

So, what about nominal scales as against ordinal scales? Nominal scales are used for labelling variables without any quantitative value. These are termed as classificatory scales, and this shows the quality of the data. This assigns responses to different categories; no numerical difference between the categories is defined in case of nominal values or nominal scales.

Categories are distinct and homogeneous, but that is very clear in this sort of data. So, the weakest tool of the measurement with no arithmetic origin or an order or distance relationship. Basically, in nominal scale, this is, in fact, the weakest tool of measurement since it does not entertain any forms of arithmetic origin. And this is not a measure of quantity, we have already mentioned.

This measures identity and difference in only people either belong to a group, or they do not this kind of explanation is important. Examples are like nature of treatment in our 75th round, and we are going to discuss about nature of treatment for traditional or modern medicine; hair colour may be brown or maybe black, maybe blonde, maybe grey and others.

Sectors those who stay, you know they might be in rural areas and urban areas. Here the numbers do not have numeric implications; they simply talk about a label, not a level. It's simply a label with a number. This makes no difference if the order of categories is changed. If we are simply flipping the order of the categories, it is not going to change any sort of value.

(Refer Slide Time: 28:12)

So, now, in the chart gives the number of males and females in a given class. Male number of students is 14 on the first example, and the female they are 17, and their frequency table is mentioned. Similarly, you know, if we change their code, you know female comes first, and male lies second in that case their frequency is simply flipped not by their number.

(Refer Slide Time: 28:44)
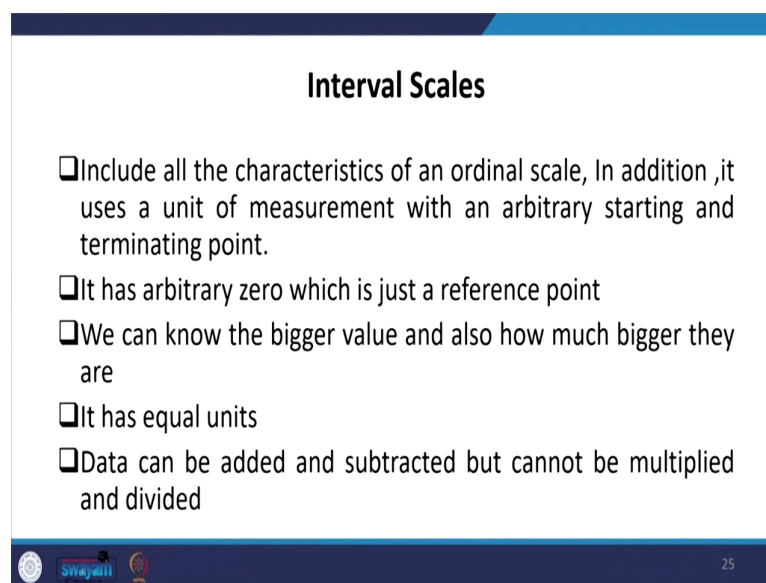
In case of the original scale, you know some ranks are given. The data has certain order like the standard of education we used to mention. So, these ordinal scales are used to interpret the relative positions. It has unequal units that do not know the numerical distance from each

category to the next. So, the numerical distance is not possible, and the ordinal scale divides objects into categories and arranges them in order.

More precise comparisons are not possible in case of ordinal scales. So, academic performance we said school level, high school level, bachelor and masters etc., we can at best get their frequencies. Coming to the interval scales, we include all characteristics of an ordinal scale; in addition, it gives a unit of measurement with an arbitrary starting and terminating point in case of interval scale.

(Refer Slide Time: 29:52)

It has arbitrary zero, which is just a reference point, but it does not have a true zero value that is zero does not have any meaning in the case of interval scales. We only compare the gap (the interval) between the points. We can know the bigger value or also how much bigger they are. These kind of things we derive have equal units. These data can be added and subtracted but cannot be multiplied and divided.

(Refer Slide Time: 30:24)
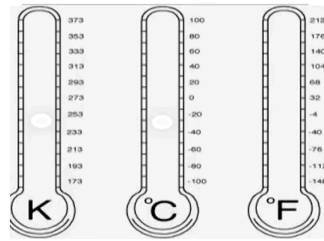
Like temperature does not have zero degree Celsius or temperature. So, IQ level like, we test for IQ level in the example it does not have a zero IQ level. There is some level of IQ. So, no IQ is hardly possible.

(Refer Slide Time: 30:39)
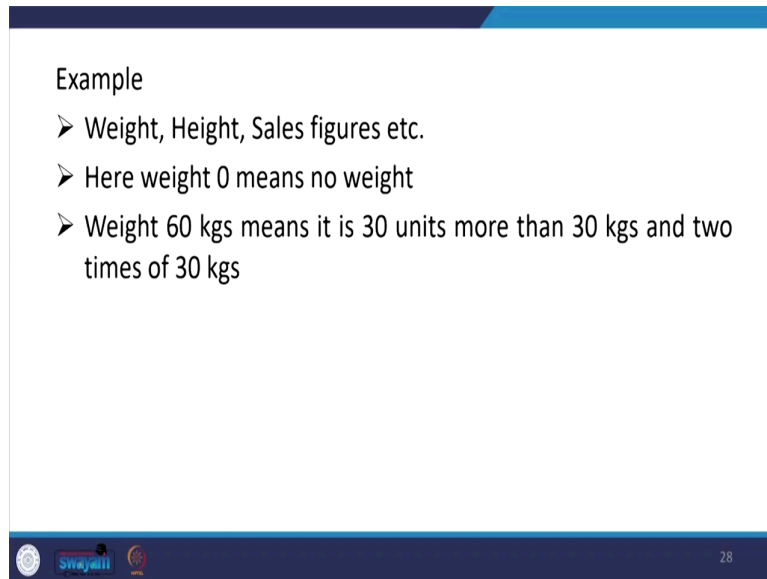
Coming to ratio scales, where we already mentioned that all sorts of numeric applications are possible. This gives equal units, absolute zero value, interval size is known, equal or constant. This measures only one side of zero. All the mathematical operations are applicable to this scale. So, top-level measurement is possible with the help of the numeric scale or the ratio scale.

(Refer Slide Time: 31:25)
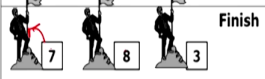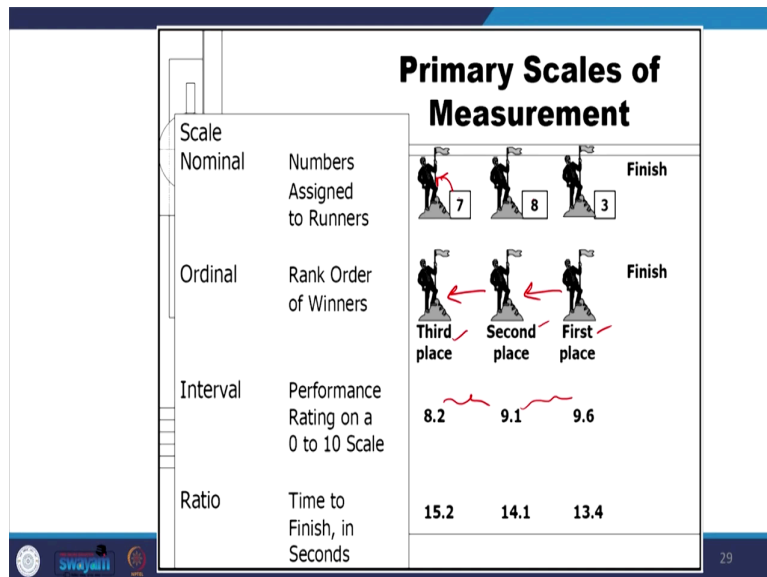
Now weight, height, scales, etc., where weight zero means no weight, it has a certain true value. Similarly, another example could be given.

(Refer Slide Time: 31:31)

So, the primary scales of measurement we have already said that nominal, ordinal, interval and ratio. So, we have given examples. This number 7, 8 or 3 only indicates a position (a number). It gives identification to this particular person. Whereas, 3rd position, 2nd position or 1st position gives certain ordering, ordering of finish; 1st position of finish and their time matters.

In interval scale, the gap is more important between two points, and so, in case of ratio scale any fraction of points are possible. So, usually, they are called scaled ratio data, and all numeric operations are possible. And some of the illustrations are given about primary scales of measurement.

(Refer Slide Time: 32:33)

I think I need not spend much time. Like these are the order of the numbers and their identifications by their serial number is given by their name. At this level, we will simply called as nominal scale. Once a 10 point ordering is made, whose ranking is highest or lowest, that can be interpreted in case of ordinal scale.

Similarly, in interval scale we are supposed to give preference or the rating in a scale maybe 1 to 7 rating or 11 to 17 ranking or any ranking, but it has to be in a range. Ratio scale gives all sorts of data points. So, we have already discussed.

(Refer Slide Time: 33:19)

Comparisons of Four Scales of Measurement

Properties	Nominal	Ordinal	Interval	Ratio
Order of value is known	no	yes	yes	yes
Frequency of distribution	yes	yes	yes	yes
Mode	yes	yes	yes	yes
Median	no	yes	yes	yes
Mean	no	no	yes	yes
Quantify the difference between each value	no	no	yes	yes
Add or subtract value	no	no	yes	yes
Multiple and divide values	no	no	no	yes
Absolute(true) Zero	no	no	no	yes

Let us compare those four scales of measurement. So, like all four scales of measurement, the order of value is known. In the case of ordinal, interval and ratio, whereas nominal, it is not possible frequency distribution all databases.

Mode frequency basically, we can get this information in all the data. The nominal does not have any median value, whereas the ordinal and other has .

Mean, I think only two. Mean value we can derive for ratio data and interval data and quantify the difference between each value which is again possible in case of these last two interval and ratio data. Similarly, subtraction and addition are also possible. Coming to multiply and divide, multiplication or division is not possible in any other database except ratio scale. So, absolute zero is only found, or it is defined in case of ratio data.

(Refer Slide Time: 34:32)
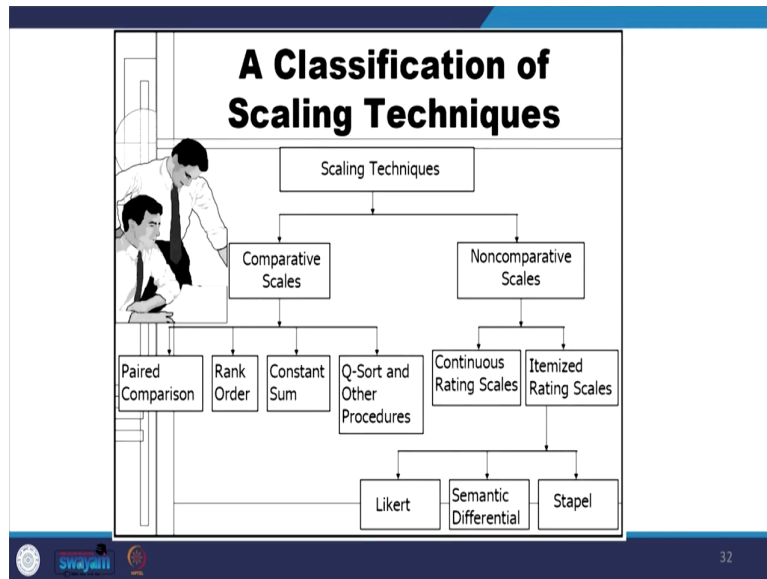
So, in short let us give a summary of our scaling techniques. Scaling techniques are broadly of two types. One is called comparative scales, and the second one is called non-comparative scales. So, comparative scales where we can able to compare the points of one estimation with another one.

In the case of non-comparative scales, it is not required to compare. In most of the marketing techniques non-comparative scales are used. So, comparative scales include paired comparison, rank order comparison, constant sum comparison, Q-sort and other procedures, whereas, in the case of non-comparative scale, we have a continuous and itemised rating scale. We will explain just now.

(Refer Slide Time: 35:27)

Within comparative scale, we have direct comparison of a given object. So, this results in in comparing the objects. The data are of nominal or ordinal and are usually non-metric; no calculations are usually made. And non-comparative scales, I think each given object scaled independently of each others resulting data is in case of non- comparative scale is of interval or ratio and these are in metric format.

(Refer Slide Time: 36:02)

A respondent is given only two options to select between Hyundai Kona Electric car and Mahindra E Verito car. So, only two comparisons if it is made: they buy by their price, their

mileage, their charging, or their power. So, if two comparisons are made, this is, in fact, called paired composition. We are simply pairing comparison.

(Refer Slide Time: 36:34)

Coming to rank order scaling, we have to keep the ranking. Out of so many options for a particular target, like, in the case of the battery, we get an MTEK, Su-kam, Exide, Amaron, Luminous etc. Out of all those things, we can ask the customers to rank those options. Each individual object is compared with the other. Some ranking could be given.

(Refer Slide Time: 37:04)

In case of constant sum though in the ranking we only give arbitrary number, but there is no caveat or boundary of their choice. A respondent in case of constant sum out of a certain limited number of entries one has to compare. Like a respondent assign constant sum of units out of total units of each individual attribute of an object. Like, suppose I just show it here.

Out of the total, let it be 150 every time. Out of the total, each group are supposed to evaluate out of 150. Your total points of feedback for a particular company could be 150. So, related to each indicator, one may give. So, your constant sum has to be constant; then, you can able to give different orderings.

(Refer Slide Time: 38:04)

And in the case of Q-sort, we are supposed to sort the responses. There are different questions and answers. So, those attributes could be sorted out. Like, if somebody is observing or understanding point of view of social license to offer it. So, what should be done in case of production unit for oil and gas? The job title is environmental advisor is supposed to understand the social license issues, how society is giving license to operate?

(Refer Slide Time: 38:29)

Now in that case there are different responses are taken. Out of those responses, 17 responses we have mentioned on this chart like further we will sort it to three important indicator which are those attributes comes under more like I think somewhat like how I think and less like what how I think.

(Refer Slide Time: 39:16)

So, I am just taking some of them to some of these entries. First, I have taken there as for my preferences then accordingly others have been taken. Finally, we can count which entries are important more like how I think which one are more in number or which categories comes under the weakness of the company or not.

(Refer Slide Time: 39:43)
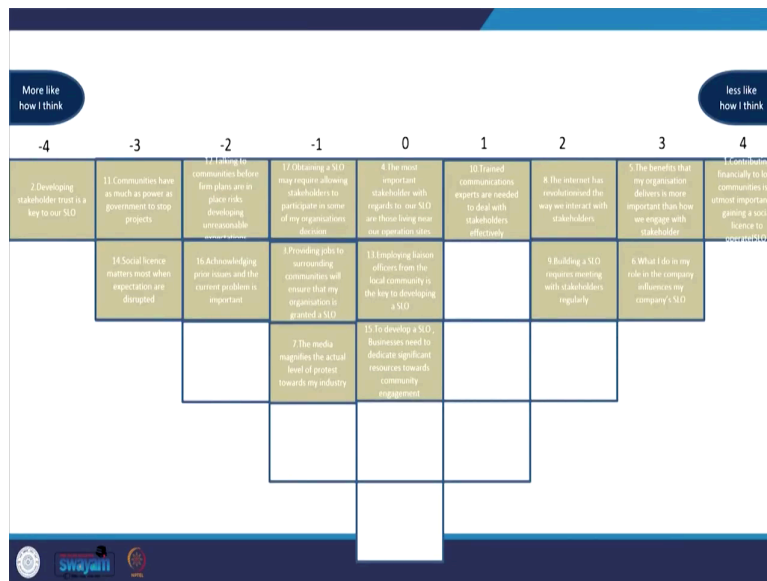
Out of that we can also put them in a scale to -4 to 4 or more likely to less likely and in between some numbers we can enter. So, these are the process called sorting Q-sort.

(Refer Slide Time: 40:01)

Now coming to the non-comparative scales these techniques measure objects independently of other objects. These techniques measure one article at a time. These are also called metric scales or monadic scale. So, some examples like customer may be asked to judge Colgate on a scale of 1 to 5. In this 1, may be least favorable and 5, may be the most favorable.

There are so many other such sort of assignment or assessment for maybe for Closeup may be for Pepsodent toothpaste. Non-comparative scales are of two types. One is called continuous scale and another is called itemised scale. The non-comparative scales is in fact, widely used in marketing research in since it is qualitatively built.

(Refer Slide Time: 40:52)

First item is first continuous scaling is explained here. The first one is called continuous rating scales where we are supposed to give a continuity in a scaling. So, all the entries from unfavorable till favorable a continuous scale is given one is supposed to put a tick mark.

So, respondents attitude towards an object is just by asking them to put a mark on the line from one extreme to another extreme. Otherwise, some discrete points may be defined and one has to put a tick mark. So, this is in fact, called continuous rating scale. Another is called itemised rating scale.

(Refer Slide Time: 41:40)

ITEMIZED RATING SCALES

- In an **itemized rating scale**, respondents are provided with a scale that has a number or brief description associated with each category.
- The categories are ordered in terms of scale position.
- The respondents are required to select the specified category that best describes the object being rated.
- Measures respondent's attitudes towards an event or an object such as product preferences and perception, purchase intentions and satisfaction.
- Three types of itemized rating scales that are commonly used in marketing research are:
  - Likert
  - Semantic Differential
  - Stapel

Itemized rating scales are provided with a scale that has a number of description or brief description associated with each category. The respondents are required to select the specified category that best describe about the object. So, this measures respondent attitude towards an event or an object such as product preferences, perception, purchase intention, satisfaction etc.

So, the itemised scales are of only three types Likert scale and semantic differential scale and Stapel scale. Most of the researchers use Likert scale. Semantic differential and Stapel scales are also used, but their use depends upon the exact requirement.

(Refer Slide Time: 42:28)

LIKERT SCALE
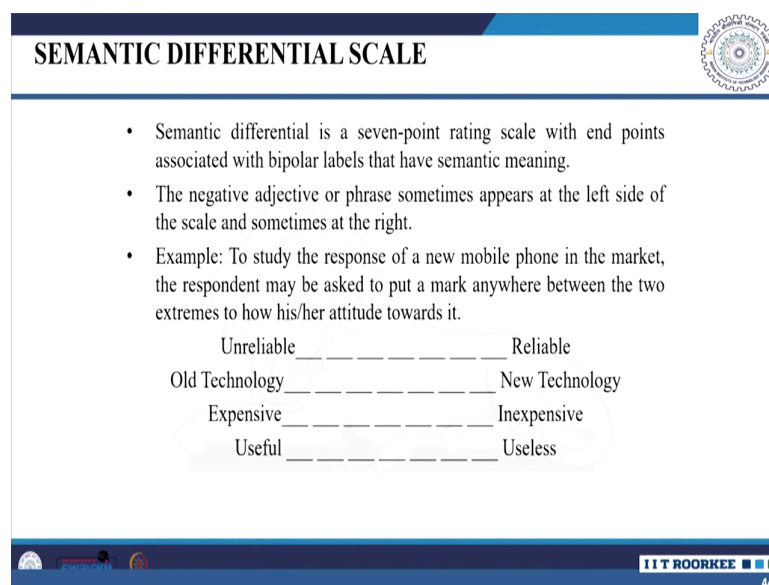
- **Likert scale** requires the respondents to indicate a degree of agreement or disagreement with each series of statements about the stimulus objects.
- Most commonly seen as a 5-point scale
- However, some practitioners advocate the use of 7 and 9-point scales which add additional granularity
- Each point of the scale has a corresponding score.
- The analysis can be conducted on an item-by- item basis (profile analysis), or a total (summated) score can be calculated for each respondent.
- It is important to use a consistent scoring procedure so that a high (or low) score consistently reflects a favourable response.
- Example: This restaurant has the best food quality.

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
(1)	(2)	(3)	(4)	(5)

These are usually put Likert scale now we are explaining as part of the itemised scale. Likert scales are usually put in 5-point scale with certain agreement to disagreement from the respondents about a particular objects that; however, some practitioner also advocate 7 to 9-point scaling techniques as well.

So, each point of the scale has a corresponding score. The analysis can be conducted on an item-by-item basis based on the total score. So, from high to low like here it is given strongly agree, disagree, undecided agree and strongly agree and its codes are given.

(Refer Slide Time: 43:22)

Coming to the semantic differential scale, the word semantic is important where the person who is responding is given the choice to respond in a contrasting manner either you from extreme best to extreme worst.

The negative adjective or phrase sometimes appears at the left side of the scale and sometimes at the right. In this scale suppose basically in the semantic scale the responses has to be bipolar.

It has to be clearly attaching the contrasting differences not like in Likert scale. Bipolar labels has to be there. The study like in an example about new mobile phones in the market the respondent may be asked to put a mark anywhere between the two extremes to show how his or her attitude varies. If the extreme views are taken that helps in taking better decision about the company.
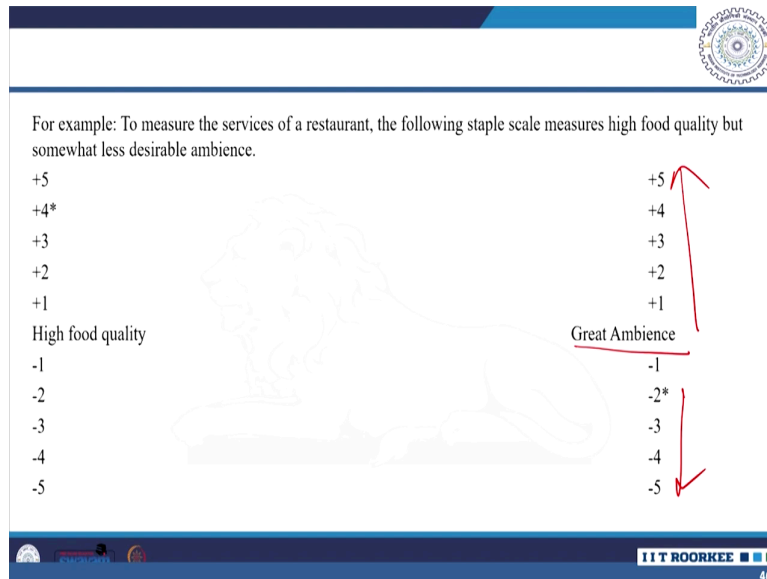
The Stapel scale initially developed by Jan Stapel. Stapel scale is a unipolar rating scale. So, now basically in it, is different than that of semantic because in semantic one we have bipolar responses, extreme references whereas, in case of Stapel we are supposed to evaluate within one direction. It is therefore, called unipolar rating scale.

When we know that our Honda showroom for automobile they know that our performance is better, it is not towards the worst side. So, it has to be from you know 5 to 10 onwards not be below 5. There should not be any negative evaluation, but from 5 to 10 what is the best indicator.

So, the scale is made of even numbers of responses that is another feature. We should note it down, like 10 classes ranging from minus 5 to plus 5 without having any neutral point. There is no neutral point in this Stapel scale. Respondents are asked to indicate by selecting an appropriate numerical response category which best suits to them.

Stapel scale is usually defined as a rating scale that is close ended with a single adjective that is unipolar adjective. This helps to gather better insights about a particular subject or event. And this is similar to semantic differential scale with the sole difference is that it has one side of the adjective.

(Refer Slide Time: 46:29)

So, this is what we said there is no zero value regarding ambience of a company you can ask from plus to minus, but there is no question of 0 value; 0 is not differentiating its responses.
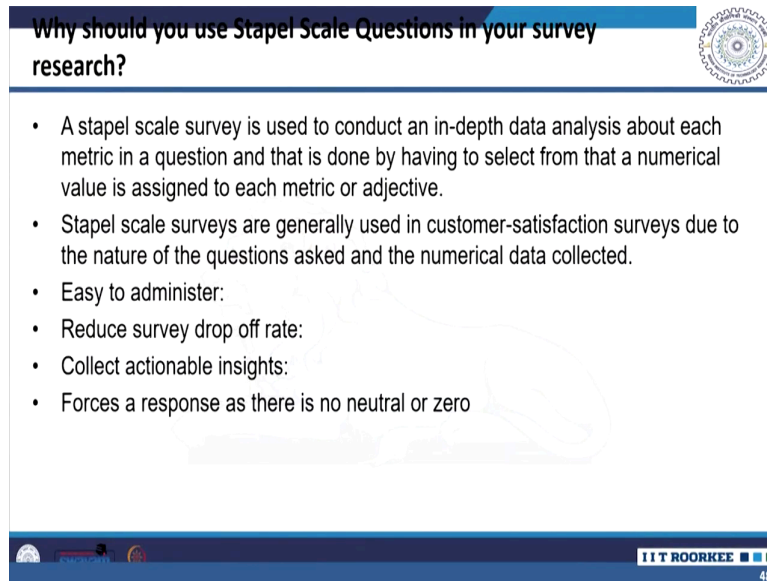
(Refer Slide Time: 46:46)

So, there are certain characteristic of Stapel scale. I think some of them I have already mentioned and some of are you may clarify from the slides. We will also upload for your reference.

(Refer Slide Time: 46:58)

Why should you use stapel scale questions in your survey research? You know the stapel scale as I already told you that it gives very intricated result give more insights and this is very easy to administer and this reduce drops many person do not respond since you gave a bigger scale.

If it is very specific scale they respond. These collect actionable insights forces responses there is no neutral value. Usually, in most of our responses we give neutral value. So, it forces the respondent to give non neutral value to your data or to your questions. So, stapel scale is usually used to conduct in-depth data analysis in your survey.

(Refer Slide Time: 47:47)

Stapel Scale Questions	Semantic Differential Questions
In a stapel scale survey, the options are **unipolar.** The rating given to each question is on the basis of one objective.	In a semantic differential question, the options are **bipolar.** The rating provided in a question is on the basis of two polar opposite objectives.
There is **no neutral point** in a stapel scale question which means each option can describe the objective.	The semantic differential **allows for a neutral point** which means that a survey respondent can leave a question unanswered if the the bipolar objectives do not appropriately define the experience.
There is no capped number of responses as long as the number of **responses are even**.	There are generally **seven responses** from highly unlikely to highly likely.
Each stapel scale survey question has **forced responses** due to the lack of a neutral zero option.	Semantic differential survey questions provide the flexibility for **unforced responses** due to the presence of a neutral option.
The options are represented **numerically.**	The options are represented **textually.**

So, these are the difference. I think I have already mentioned difference between stapel scale and semantic differential questions. So, all those things I have said, for your clarity I have put it in a box and these are referred to some of the publicly available documents.

(Refer Slide Time: 48:07)

Some Readings

❑ Dr. J K Nayak(IIT Roorkee), Marketing Research, NPTEL video lectures, 2018
❑ Dr. C.R. Kothari, Research Methodology: Methods and Techniques - 2nd Edition, New Age International, 2004
❑ Matt Anticole ,Why the metric system matters - Matt Anticole, July 21,2016, TED-Ed video
❑ http://www.fao.org/
❑ https://businessjargons.com/scaling-techniques.html
❑ https://pdfs.semanticscholar.org
❑ https://www.slideshare.net/
❑ https://themarketingresearch.com/
❑ https://www.cram.com/flashcards
❑ http://en.wikipedia.org/wiki/Likert_scale

So, for your interest some of the readings you may follow to further prepare for your research work. I think I need not read it out. It is very clear. With this I think I should end here. If any questions please be free to write down and raise your questions in the live sessions.

Thank you.