2HSOE52 Introduction to Economics

Chapter 2 : Measurement Scales , Data Structure and Data sources

Samir K Mahajan, Ph.D. , UGC NET

A. MEASUREMENT SCALES

Two words "counting" and "measurement" are very frequently used by everybody

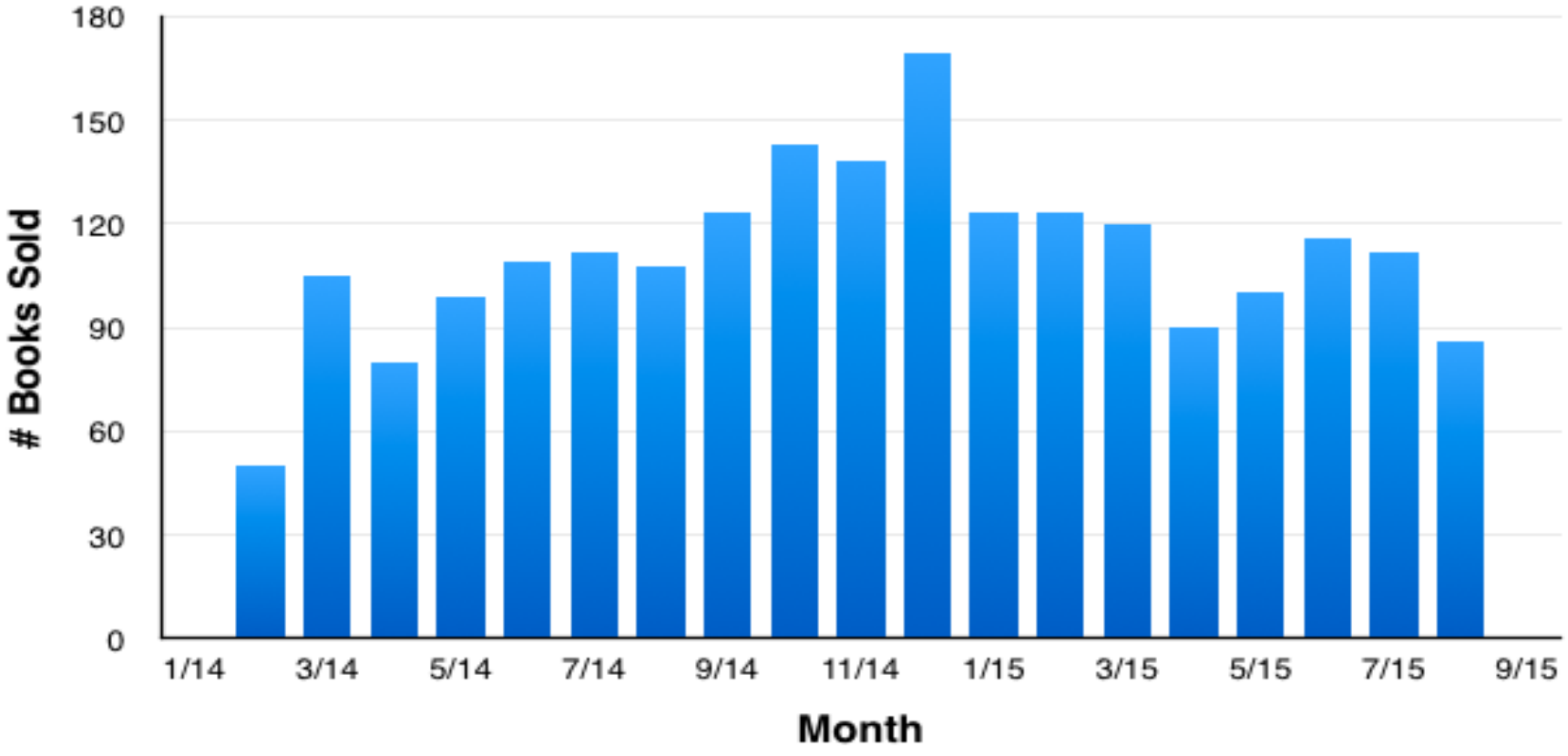
- **Items that we count?**
- **Items that we measure ?**

Count

HOW MANY MARBLES ARE IN THIS JAR?

Tipping Point Math

Count


180
150
120
90
60
30
0
Books Sold
1/14
3/14
5/14
7/14
9/14
11/14
1/15
3/15
5/15
7/15
9/15
Month

Measure

Infant & Toddler Size Chart

Sizes	Height (in.)	Weight (lb.)	Chest (in.)	Waist (in.)	Hip (in.)
3 mos.	18-24	7-13	17	18	17
6 mos.	24-27	13-17	18	18½	18
12 mos.	27-29	17-22	19	19	19
18 mos.	29-31	22-25	20	19½	20
24 mos.	31-33	25-28	20½	20	20½
2T	33-34	28-29	21	20	22
3T	35-36	29-32	22	20½	23
4T	37-39	33-36	23	21	24

A.1. Levels of Measurement Scale /Data Accuracy

In Statistics, act of counting and measurement is divided into 4 levels of measurement scales known as

(1) Nominal Scale
(2) Ordinal Scale
3) Interval Scale
(4) Ratio Scale

A.1.1 Nominal Scale

In Latin, ‘Nomen’ means name.

- **Objects are divided into two or more categories by giving them unique names. We can code the objects.**
- **Each object either belongs to a category or not**
- **In statistical language categories must be mutually exclusive and exhaustive.**
- **Mathematically, we may use the symbols $=, \neq$ if an object falls in a category or not.**
- **We can not order /rank the categories. Sign of less than or greater than does not make any sense in nominal scale.**

(A.1.1 Nominal Scale : examples

Classification into Different Categories Based on Gender

Category	Name/Code	Binary Code
Female	F	1
Male	M	0

- male > female or female > male does not make any sense.

(A.1.1 Nominal Scale : examples

ii) Classification into Different Categories Based on Caste

Different categories	Name given	Codes allowed /
General	Gen	1
Scheduled caste	SC	2
Scheduled tribes	ST	3
Backward class	BC	4
Others	‘O	5

- **WE have coded Gen , SC by ‘1’ and ‘2’ respectively. But Gen > SC or SC > Hindu does not make any sense.**

A.1.2 Ordinal Scale

If in a measurement scale orders make sense then, this scale comes under the heading ordinal scale.

- **We do ranking of categories .**
- **Difference between different categories is not possible, and is meaningless.**
- **Ordinal scale is used when we want to measure the attitude scores towards the level of liking, satisfaction, preference, designation in jobs etc.**
- **Mathematically, ordinal scale includes symbols >, < in addition to those used for nominal scale, i.e. =, ≠**
- **Say in class of 5 students we can rank the students as 1^{st}, 2^{nd}, 3^{rd}, 4^{th}, 5^{th}**

We can not take the difference in ranking of the students.

A.1.2 Ordinal Scale: examples

- **Based on economic condition, generally , families of a society are divided into three categories:**

Higher class family
Middle class family
Lower class family

We can make a ordering here.

We can say economic condition of higher class family is better than middle class family and middle class family is in a better condition compare to lower class family.

2) Ordinal Scale: : examples

Opinion of persons about proposal of introducing co-education in a college in this Likert Scale are given as under:

Opinions	Strongly Agree	Agree	Indifferent	Disagree	Strongly Disagree
Score	5	4	3	2	1

Here, the order matters i.e i.e. 5 > 4 > 3 > 2 > 1 or 1 < 2< 3 < 4< 5 holds good .

But difference between 5 and 4 or 4 and 3 are meaningless.

A.1.3. Interval Scale

- Nominal scale can order , and take the difference between the order.

If I = [4, 9] , 9> 5 , and the difference between 4 and 9 is 5.

- Property of difference holds in case of intervals.
- Interval Scale can not take ratio.
- Interval scale does not have an absolute/true zero. Zero is arbitrary in interval scale.
- Mathematically, this scale includes +, – in addition to >, < and =,≠ .

A.1.3. Interval Scale : Examples

Interval scale is used when we want to measure

- **Years/historical time/calendar time**
- **Temperature (except in the Kelvin scale),**
- **Sea level**
- **Marks in the tests where there is negative marking, etc.**

1.2.3 Interval Scale : examples

Temperature in degree Celsius (^{0}C)

- Measurement of temperature in degree Celsius (^{0}C) assumes 0 ^{0}C when water starts freezing to ice and it becomes ice at – 4^{0}C .
- In degree Celsius , origin is arbitrary . Hence, No absolute zero.
- That's why measurement of temperature in degree Celsius comes in interval scale.
- Because in degree Celsius origin is arbitrary, so we cannot say that 30 ^{0}C is twice as hot as15 ^{0}C .
- I f it is so then can we say that 4 ^{0}C is –1 times –4^{0}C? But it is meaningless.

1.1.4 Ratio Scale

Recap

- **Nominal scale gives only names to the different categories**
- **Ordinal scale provides orders between categories other than names,**
- **Interval scale provides the facility of difference between categories other than names and orders**

1.1.4 Ratio Scale

- Ratio scale other than names, orders and characteristic of difference also provides natural zero (absolute zero) and allow division
- In ratio measurement scale values of characteristic cannot be negative.
- Ratio scale includes $\times$, $\div$ in addition to +, –, >, <, =, $\neq$. But be careful never take '0' in denominator while finding ratios.
- Ratio scale is the highest level of measurement

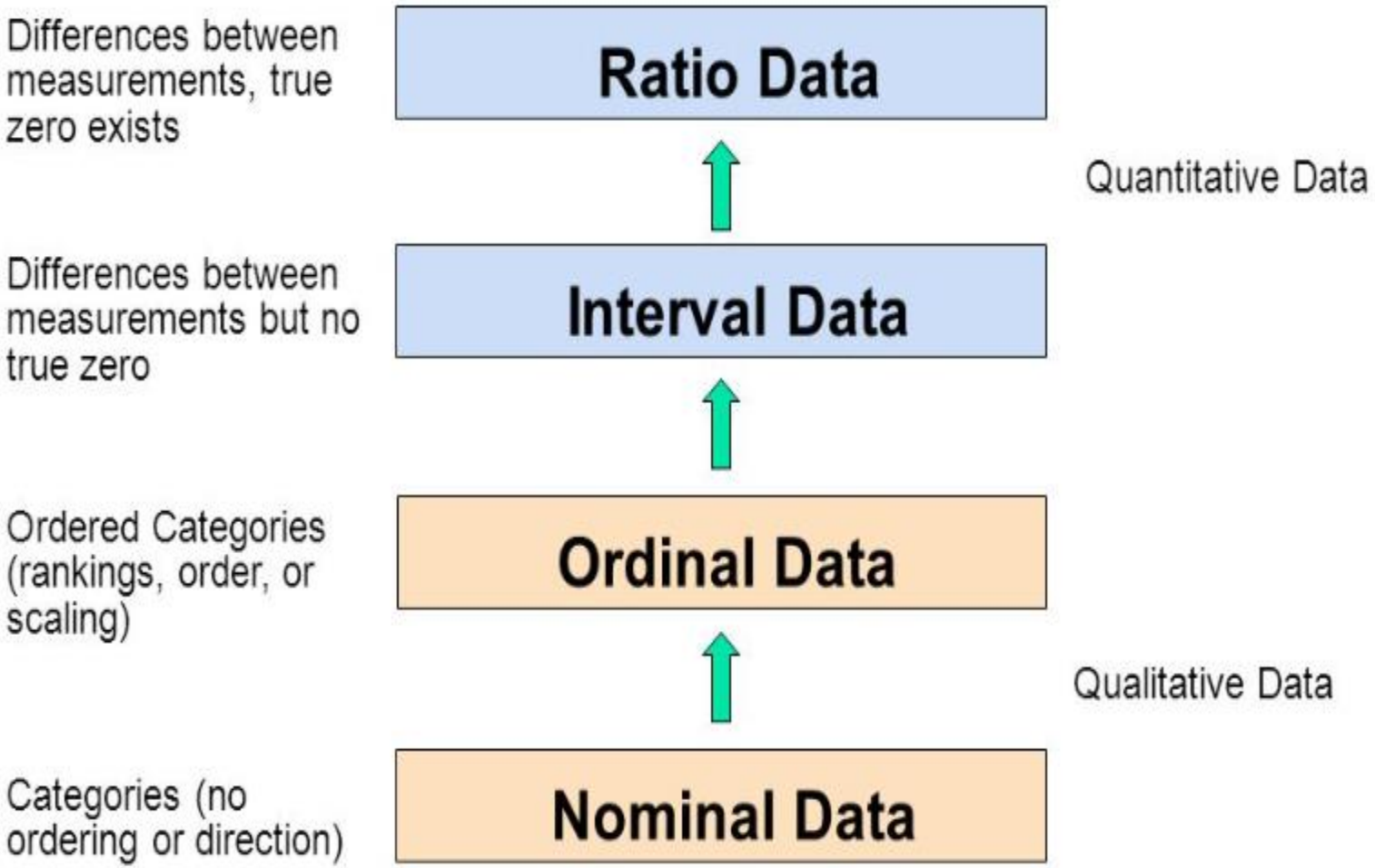
1.1.4 Ratio Scale: Examples

1. Measurement of money also comes under ratio scale because it satisfies all the requirement of interval scale and has a natural zero.

2. Measurement of temperature in Kelvin scale comes under ratio scale because it has an absolute zero which is equivalent to – 273.15^{0}C . This characteristic of origin allows us to make the statement like 50K ('50K' read as 50 degree Kelvin) is 5 time hot compare to 10K.

3. **Both height (in cm.) and age (in days) of students of ITNU B.Tech students . Statistics of a particular university satisfy all the requirements of a ratio scale. Because height and age both cannot be negative (i.e have an absolute zero).**

1.2 Levels of Measurement Scale in Nutshell



1.3 Measurement Scale and Permissible Statistical Tools

- **Measurement scale help us to decide which statistical tool should be used in a given situation.**
- **Statistical tools applicable on the lower scale will automatically be applicable on the next level scale.**

1.3 Measurement Scale and Permissible Statistical Tools

MEASUREMENT SCALE	PERMISSIBLE STATISTICAL TOOLS	LOGIC/REASON
NOMINAL SCALE	Mode, chi-square test and run test : A	Here counting is only permissible operation.
ORDINAL SCALE	A, Median all positional averages like quartile, Decile, percentile, Spearman's Rank correlation : B	Here other than counting, order relation (less than or greater than) also exists.
INTERVAL SCALE	B, Mean , S.D., t-test, F-test, ANOVA, sample multiple and moment correlations, regression : C	Here counting, order and difference operations hold.
RATIO SCALE	C, Geometric mean (G.M.), Harmonic mean (H.M.), Coefficient of variation	Here counting, order, difference and natural zero exist.

B. Variables

Variable is an element, feature, or factor that is liable to vary or change.

- **Variables can be measured on a numeric or quantitative scale. Interval and ratio scales are quantitative. A country's population, a person's shoe size, or a car's speed are all quantitative variables.**

- **Variables that are not quantitative are known qualitative or categorical variables . Nominal and ordinal are qualitative scales .**

B. 1. Quantitative Variables

Quantitative variables are of two types:

- **Discreet**
- **Continuous**

B. 1.1 *Discrete Variable*

A discrete variable is a variable whose value is obtained by counting.

A variable is said to be discrete if it takes countable number of distinct value. The values that a discrete random variable takes is whole number such as 0, 1, 2, 3,.....

For instance, the number of children (or adults, or pets) in your family is discrete data, because you are counting whole, indivisible entities: you can't have 2.5 kids, or 1.3 pets.

***Examples*:** **number of students present**
number of red marbles in a jar
number of heads when flipping three coins
students' grade level
number of head from toss of a coin

B. 1.1 *Continuous Variable*

- **A continuous variable is a variable whose value is obtained by measuring.**
- **Continuous variables are numeric variables that have an infinite number of possible values between any two values. It can take integers , fraction.**
- **A continuous random variable is not defined at specific values. Instead, it is defined over an *interval* of values, and is represented by the *area under a curve* (in advanced mathematics, this is known as an *integral*).**

***Examples*:** **height of students in class**
weight of students in class
time it takes to get to school
distance traveled between classes

B. 2 *Qualitative Variable*

- ***Binary /dichotomus viz.*** **right/wrong, true/false, or accept/reject.**

- **Multichotomous : more than tow catagories viz. Post graduate, undergraduate, higher secondary passed,**

- ***Ordered*** **or** ***ordinal*** **data such as (a) "Short, Medium, or Tall.**

(b) survey question that asks us to rate an item on a 1 to 10 scale, with 10 being the best. This implies that 10 is better than 9, which is better than 8, and so on.

C. CLASSIFICATION OF DATA

- **Data can be classified with respect to some characteristics.**

Broadly, data can be classified under following categories:

(i) Geographical classification (say location, region)

(ii) Chronological classification (over a period)

(iii) Qualitative classification (characteristics that can be measured numerically such as height, weight, income, age, sales)

(iv) Quantitative classification (attributes or qualitative characteristics such as sex, colour of hair, literacy, religion)

D. Data Structures in Economics

As far data structure is concerned, there are two basic types of economic data:

- **Cross-sectional data**
- **Time series data**

- **combine features of cross-sectional and time series data sets**

- **Panel or longitudinal data**
- **Pooled cross-section time-series data,**
- **.**

D.1 Cross Section Data

- **Cross-section data are data on one or more variables collected *at the single point in time***

- **A cross-sectional data set consists of a sample of observations on individual economic agents or other units taken at a single point in time or over a single period of time.**

D.1 Cross-sectional data : Example

Human Development Index and its components, 2018

HDI rank	Country	Human development index (HDI) (index value)	Life expectancy at birth (years)	Expected years of schooling (years)		Mean years of schooling (years)		Gross national income (GNI) per capita (2011 PPP $)
1	Norway	0.954	82.3	18.1	b	12.6		68,059
2	Switzerland	0.946	83.6	16.2		13.4		59,375
3	Ireland	0.942	82.1	18.8	b	12.5	c	55,660
4	Germany	0.939	81.2	17.1		14.1		46,946
4	Hong Kong, China (SAR)	0.939	84.7	16.5		12.0		60,221
6	Australia	0.938	83.3	22.1	b	12.7	c	44,097
6	Iceland	0.938	82.9	19.2	b	12.5	c	47,566

D.2 Time series data

- **A *time series* is a set of observations on the values that a variable takes at different times.**

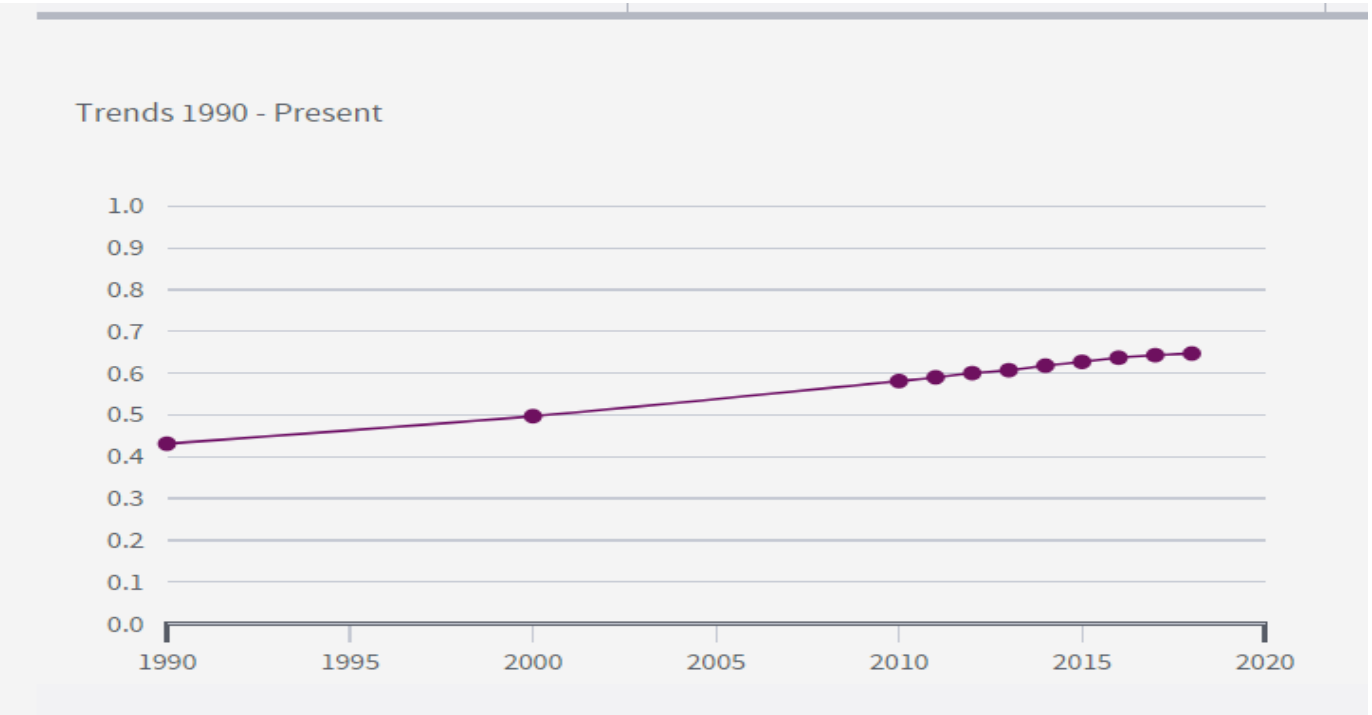
- **Such data may be collected at regular time intervals, such as:**

  - **daily (e.g., stock prices, weather reports), weekly (e.g., money supply figures),**
  - **monthly [e.g., the unemployment rate, the Consumer Price Index (CPI)],**
  - **quarterly (e.g., GDP), annually (e.g., government budgets),**
  - **Quinquennially, that is, every 5 years (e.g., the census of manufactures),**
  - **or decennially (e.g., the census of population).**

Data Structures in Economics

D.2 Time series data : Example

HDI : India



Source : UNDP

D.2 Time series data : Example

Prices the months

Month	Price ($)	Month	Price ($)	Month	Price ($)
1	2.27	13	2.84	25	3.91
2	2.63	14	2.73	26	3.68
3	2.53	15	2.73	27	3.65
4	2.62	16	2.73	28	3.64
5	2.55	17	2.71	29	3.61
6	2.55	18	2.80	30	3.45
7	2.65	19	2.86	31	3.38
8	2.61	20	2.99	32	3.27
9	2.72	21	3.10	33	3.38
10	2.64	22	3.21	34	3.58
11	2.77	23	3.56	35	3.85
12	2.85	24	3.80	36	3.90

Data Structures in Economics

D.3 Panel Data/Micro Panel Data/Longitudinal data

- A panel or longitudinal data set consists of two or more sets observations on the *same* sample of cross-sectional units at two or more points in time.

- A panel data set consists of *repeated observations* over time on the *same* set of cross-sectional units of a sample.

A panel-data observation has two dimensions: X_{it}

where i = 1, 2, ..., N cross sectional unit

t= 1, 2, ... T time period

Say X indicate income . Then X_{it} indicate income of individual i at time t .

D.3 Panel Data : example

Personal ID	Age	Sex	Year	Income
1	27	M	2007	15000
1	28	M	2008	17000
1	29	M	2009	18000
2	43	F	2007	24000
2	44	F	2008	28000
2	46	F	2009	30000
3	35	M	2007	22000
3	36	M	2008	22000
3	37	M	2009	24000

D.3 Panel Data: Example

Human development index (HDI) during 1990-1996

HDI Rank (2018)	Country	1990	1991	1992	1993	1994	1995	1996
170	Afghanistan	0.298	0.304	0.312	0.308	0.303	0.327	0.331
69	Albania	0.644	0.625	0.608	0.611	0.617	0.629	0.639
82	Algeria	0.578	0.582	0.589	0.593	0.597	0.602	0.61
36	Andorra	..	..	..	..	..	..	..
149	Angola	..	..	..	..	..	..	..
74	Antigua and Barbuda	..	..	..	..	..	..	..
48	Argentina	0.707	0.714	0.719	0.725	0.729	0.731	0.738
81	Armenia	0.633	0.629	0.585	0.59	0.6	0.604	0.614
6	Australia	0.866	0.867	0.868	0.872	0.875	0.883	0.886

D.4 Pooled Cross-Section Time-Series Data

- **A pooled cross-section time-series data set consists of two or more different samples of cross-sectional observations from the same population taken at two or more points in time.**

A pooled -data observation has minimum m three dimensions: X_{ij}^{g}

where i = 1, 2, ..., N cross-sectional unit

t= 1, 2, ... T time period
g= 1, 2, ... G districts in a State (say)

Let X indicate income then

X_{ij}^{g} indicate income of individual 'i' in village 'g' at time t.

D.4 Pooled Cross-Section Time-Series Data

Personal ID	Village	Age	Sex	Year	Income
1	I	27	M	2007	15000
1	I	28	M	2008	17000
1	I	29	M	2009	18000
2	II	32	F	2007	24000
2	II	33	F	2008	28000
2	II	34	F	2009	30000
3	II	35	M	2007	22000
3	II	36	M	2008	23000
3	II	37	M	2009	26000
4	I	32	M	2007	26000
4	I	33	M	2008	27000
4	I	34	M	2009	29000

E. Data Source

As far as sources of data is concerned , there are mainly two types of data namely

- Primary Data and Secondary Data.
- Both have their own importance.

E.1 Primary Data

- Primary Data are collected by an investigator or agency or institution for a specific purpose and these people are first to use these data,

- That is, these data are originally collected by these people and they are first to use these data.

- Suppose, a researcher collects information (economics wage, social say education , demographic say age) residential private security guards in the city of Ahmedabad to study from the private security guards to make a study their socio-economic status.

E.1 Primary Data: Method of Collection

Here we will discuss only following commonly used methods.

(1) Direct Personal Investigation Method
(2) Telephone Method
(3) Indirect Oral Interviews Method
(4) Local Correspondents Method
(5) Mailed Questionnaires Method
(6) Schedules Method

E.2. Secondary Data

- **Secondary data are the data obtained/gathered by an investigator or agency or institution from a source which already exists.**

- **These data were originally collected by an investigator or agency or institution and have been used by them at least once. And now, these data are going to be used at least second time.**

- **For example, in the same study of socio-economic status of private security guard , the researcher used published data (FICCI) related number of private security guards over the years in India**

E.2. Secondary Data : Sources

(1) Published Sources

- World Health Organisation (WHO)
- World Bank
- Office of Registrar General of India
- (Central Statistical Organisation (CSO)
- National Sample Survey Organisation (NSSO)
- Reserve Bank of India Bulletin
- Directorate of Economics and Statistics (DES), etc.
- GOI Budget Documents
- Sarkaria Commission
- International Journal of Probability and Statistics (Journal)
- data.gov.in

(2) Unpublished Sources

- Unpublished projects works, field works or some other research related
- works submitted by students in their corresponding institutes
- Records of Central Bureau of Investigation
- Personal diaries, etc.

Reference

1. IGNOU Books

2.http://qed.econ.queensu.ca/pub/faculty/abbott/econ481/481note01_f07.pdf

3. https://blog.minitab.com/blog/understanding-statistics/understanding-qualitative-quantitative-attribute-discrete-and-continuous-data-types

Inputs in these slides are exclusively collected from above sources.