

Research Methodology

Content

- Types & Sources of Data
- Classification and Tabulation of Data
- Measures of Central Tendency & Variability

What is Social Research

- ▶ Social Sciences research deals with the study of human behavior and societies whereas natural sciences are concerned with study of natural phenomena.
- ▶ Objective investigation of empirical phenomena in terms of regularity
- ▶ Does not refer to any general or particular body of knowledge but to a distinct methodology
- ▶ Acquisition of factual knowledge through empirical regularity

- ▶ Quantitative research, is based on measurement of quantity or amount. It is applicable to phenomena that can be expressed in terms of quantity.
- ▶ Qualitative research, on the other hand, is concerned with qualitative phenomena i.e. phenomena relating to or involving quality or kind.

› ***Quantitative Research Focuses at***

- Estimates
- Levels & Trends
- Determinants & Consequences
- Predictors & Response

Qualitative Research Methods helps in.....

- Understanding of processes
- Psycho-social and cultural constructs and deconstructs of various research issues
- Filling up the gaps in the knowledge and understanding of researchers

Quantitative Vs Qualitative methods

› ***Quantitative Research methods.....***

- *Questionnaire*
- *Schedule*
- *Interview*

› ***Qualitative Research methods.....***

- Observation
- In-depth interviews
- Focus Group discussion
- Case Study

Quantitative Vs Qualitative Data

- ▶ Data: *Information collected keeping in view the objective of the research.*
- ▶ **Quantitative Data:**
 - Deals with numbers
 - Can be measured
- ▶ **Qualitative Data:**
 - Deals with description that is in form of interviews, pictures
 - Can not be measured but can be observed

Sources of data

- ▶ Primary source
 - Information collected by researcher using different quantitative or qualitative methods.
Interviews, newspaper clippings,
- ▶ Secondary Source
 - Any published/ unpublished information collected already but used by some other person.

Sources of available Data

- ▶ The available data sources may be placed in five broad categories as follows:
 - Public documents and official records, including the extensive archives of the Census
 - Private documents(letters, diaries etc)
 - Mass media
 - Social science data archives (Survey data)
- ▶ Though these categories are not mutually exclusive more than one data source can be used in a study

Data Editing

- ▶ Generally done to control the quality of data
- ▶ Process of checking and correcting the data collected is called statistical data editing.
- ▶ Data editing done to : Minimize the measurement error, missing values, typing error

- ▶ Classification is a process of statistical analysis where as tabulation is a process of presenting the data in suitable form.
- ▶ Classify data by assigning arbitrary limits called class-limits.
- ▶ The group between any two class-limits is termed as class or class-interval.

Rules of Classification

- ▶ All the classes should preferably have equal width or length.
- ▶ Neither too large nor too small
- ▶ Classes should mutually exclusive
- ▶ No. of class-intervals= Range divided by Width.
- ▶ Range indicates the variation in data

Marks of 100 Students of a Class in Economics

72 61 63 65 62 68 69 64 65 67 69 (56) 60
 66 62 57 72 67 65 70 64 66 71 (73) 67 65
 64 63 61 58 64 62 69 66 65 63 63 59 61 64
 65 57 66 71 68 70 67 66 60 62 65 58 63 68
 64 61 62 65 66 59 62 65 65 60 64 61 64 69
 62 64 62 63 68 67 65 62 65 68 61 63 62 72
 62 66 66 65 63 67 66 63 63 66 65 63 62 62
 66 64 62 62

Frequency distribution

- ▶ Simple frequency distribution
- ▶ Grouped Frequency Distribution
 - $UL - LL = \text{Range} / \text{width} = \text{No. of class interval}$
 - E.g. $73 - 56 = 17 / 10 = 1.7 = 2$
 56-57, 58-59.....

Tabulation of data

- ▶ Tabulation may be defined as the systematic presentation of numerical data in rows or/and columns according to certain characteristics
- ▶ **Table number:** A number must be allotted to the table for identification, particularly when there are many tables in a study.
- ▶ **Title:** The title should explain what is contained in the table. It should be clear, brief and set in bold type on top of the table. It should also indicate the time and place to which the data refer.
- ▶ **Stubs, or, Row designations:** Each row of the table should be given a brief heading.
- ▶ **Column headings or, Captions:** Column designation is given on top of each column to explain to what the figures in the column refer

- ▶ **Body of the table:** The data should be arranged in such a way that any figure can be located easily. Column or row total should be given.
- ▶ **Unit of measurement:** Stated at the top right-hand corner of the table along with the title.
- ▶ **Source:** At the bottom of the table indicating the primary and secondary sources
- ▶ **Footnotes and references:** If any item has not been explained properly, a separate explanatory note should be added at the bottom of the table.

Exercise

- ▶ Draw up a blank table to show the number of Individuals surveyed classified according to
- ▶ (i) Sex: Male and Female;
- ▶ (ii) two age-groups: below 40, 40 and above;
- ▶ and (iii) Four caste-groups: SC,ST, OBC Others

Distribution of Individuals classified by sex, age group and social group

Caste/ sex	0-40		40 & above	
	Male	Female	Male	Female
SC				
ST				
OBC				
Others				
Total				

Graphical Presentation of Data

- ▶ Line Diagram
- ▶ Bar diagram
- ▶ Pie diagram
- ▶ Histogram
- ▶ Frequency Polygon

- ▶ Graphs do not provide quantitative statements or do not allow easy comparison

Measures of Central Tendency

- ▶ An index or measure which represent the whole group.
- ▶ Averages are often referred as measures of central tendency.
- ▶ Tells about the nature of the distribution

Characteristics of Good Average

- Easy to understand & simple to calculate
- Representative of all item
- Not affected by extreme observations
- Rigidly defined (algebraic formula)
- Further algebraic treatment

- ▶ Measures of Central Tendency:
 - ▶ - Mean • The sum of all scores divided by the number of scores.
 - ▶ - Median • The score in the middle when the scores are ordered.
 - ▶ - Mode • The most frequent score.
- ▶ Each represent central value of the data

MEAN

- ▶ The mean is defined as the arithmetic average of a set of numerical scores, that is, the sum of all the numbers divided by the number of observations contributing to that sum.
- ▶ $\text{MEAN} = \frac{\text{sum of all data values}}{\text{number of data values}}$ or, more formally:
- ▶ Though it is a central value of distribution need not necessarily in the domain of the variable

Arithmetic Mean

- ▶ Sum of the values divided by number of values
- ▶ Sample: 5,4,3,5,4,7,6,5
- ▶ Mean =

$$\frac{5+4+3+5+4+7+6+5}{8} = 4.8$$

Advantages

- ▶ Based on all observations
- ▶ Simple to calculate
- ▶ Reliable in large data set

Disadvantages

- ▶ Influenced by extreme values
- ▶ Can yield decimal figure (5.1 person)
- ▶ Less reliable for small sample size

Contd...

Individual series:

▶ i. Direct: $\bar{X} = \frac{\sum X}{N}$

▶ ii. Short-cut: $\bar{X} = A + \frac{\sum d}{N}$

Contd...

▶ **Discrete series**

◦ i. Direct : $\bar{X} = \frac{\sum fX}{N}$

◦ ii. Short-cut: $\bar{X} = A + \frac{\sum fd}{N}$

▶ **Continuous Series**

◦ i. Direct : $\bar{X} = \frac{\sum fm}{N}$

◦ ii. Short-cut: $\bar{X} = A + \frac{\sum fd}{N}$

◦ iii. Step-deviation: $\bar{X} = A + \frac{\sum fd' \cdot n \cdot i}{N}$

Mean in grouped data

X	f	m	fm
10 - 12	4	11	44
13 - 15	12	14	168
16 - 18	20	17	340
19 - 21	14	20	280
	N=50		832

$$\bar{x} = \frac{832}{50} = 16.6$$

Combined Mean

- ▶ Applicable where more than one group information is provided.
- ▶ Ex: In a class of 10 students, 4 students average score is 30, 6 students average score is 20, then combined mean
- ▶ $((4 \times 30) + (6 \times 20)) / (6 + 4) = (120 + 120) / 10 = 24$
- ▶ Combined mean is 24.

Median

- ▶ Middle number when measurements are arranged ascending(or descending) order
- ▶ For even number calculate the average of two central values
- ▶ Sample:
 - ▶ 6,4,7,5, 8,3,9,10
 - = 3,4,5,6,7,8,9,10
 - = Median = $\frac{4^{\text{th}} + 5^{\text{th}} \text{ value}}{2}$

Advantage

- ▶ Not affected by extreme value
- ▶ Useful for qualitative data

Disadvantage

- ▶ Arrangement of data
- ▶ Mathematically less accurate

- ▶ Median= 6.5

Median in grouped data

- ▶ Step 1: Construct the cumulative frequency distribution.
- ▶ Step 2: Decide the class that contain the median.
 - Class Median is the first class with the value of cumulative*
 - frequency equal at least $n/2$.
- ▶ Step 3: Find the median by using the following formula:

Median in grouped data

‣ Median = $L_m + ((n/2 - F) / f_m) * i$

- N = total frequency
- F = c.f before class median
- F_m = frequency of class median
- i = class width
- L_m = lower boundary of the class

Contd...

X	f	CF
1 - 10	8	8
11 - 20	14	22
21 - 30	12	34
31 - 40	9	43
41 - 50	7	50

$n/2 = 25$ --- the 3rd median class
 $F = 22$, $f_m = 12$, $L_m = 21.5$,
 Median = 24

Mode

- ▶ Value that occur most frequently in the data
- ▶ Sample: 5,4,3,5,4,7,6,5
- ▶ If the data set have one peak called uni-modal and two peak called bi-modal
- ▶ Mode: 5

Advantages

- ▶ Describes overall shape of the distribution
- ▶ Easy and no calculation

Disadvantages

- ▶ Not capable of algebraic manipulation
- ▶ Does not take into account all observations

Choosing an appropriate measure

- ▶ Mean is the preferred method but median is preferred over mean When distribution is skewed and used to combat the effect of outlier
- ▶ The mode is not usually used because the largest frequency of scores might not be at the center
- ▶ The only situation in which the mode may be preferred over the other two measures of central tendency is when describing discrete categorical data.
- ▶ The greatest frequency of responses is important for describing categorical data.

Assignment

- Expenditure of 20 households given below. Categorise the sample into 3 parts as high(400–500), medium(200–300) and low(100–200) and calculate the mean & median using appropriate formula.

500	250	100	340	430	270	460	390	120	350
190	340	450	270	190	360	470	490	500	320

Reference

- Follow some basic statistics book
- S. C. Gupta & V.K.Kapoor: **Fundamentals of Mathematical Statistics**
- S.C. Gupta – Fundamentals of Statistics

Poverty, Inequality & Social Exclusion

Unit-2

Introduction

- ▶ A country is poor because it is poor- Nurkse, R- Vicious circle of poverty.
- ▶ Low income- low saving-low investment-low income.
- ▶ Low income and inequality results in poverty
- ▶ Income inequality - poverty- can be understand through income distribution

Contd...

- ▶ Classical economics- Income distribution measured through factor share.
- ▶ Ricardo & Marx- land & capital- profit(land lord & capitalist) while labourer has no means of production except labour(wage).
- ▶ Concentration of income in few hands
- ▶ In advanced economies- employee possess both tangible & intangible asset(knowledge, skill, human capital).
- ▶ Inequality & poverty- distribution of income among households.

Contd...

- ▶ Economic growth & Inequality
- ▶ Least benefit goes to marginalized section.
- ▶ Poverty: A minimum level of living necessary for physical and social development of a person.
- ▶ Poverty estimates are vital input to design, monitor and implement appropriate anti-poverty policies

Causes of Poverty

- ▶ Economic factors:
 - ▶ Low agricultural productivity, fragmentation & subdivision, unemployment, underutilization of resources, price rise, decline in village industries
- ▶ Demographic factors: Household size, Population growth
- ▶ Social causes: Illiteracy, Indebtedness, social custom
- ▶ Environmental causes: Climate change, drought, flood
- ▶ Government expenditure

Types & Measurement of Poverty

- ▶ Poverty: Absolute vs. Relative
- ▶ Absolute PL: Persons below a pre-defined threshold income (consumption) are called poor.
- ▶ Relative PL defined in relative terms with reference to level of living of another person; or, in relation to an income distribution parameter.

Measurement of Poverty

- ▶ Two basic items for measuring poverty:
- ▶ (1) Poverty Line: definition of threshold income or consumption level
- ▶ (2) Data on size distribution of income or consumption (collected by a sample survey representative of the population)

Poverty Line

- ▶ The cut-off value of the consumption that determine the poor and non-poor is poverty line.
- ▶ Usually measured by the cost of purchasing bundle of goods & services to satisfy the basic needs
- ▶ Basic needs: Food & non-food

Expert groups on measuring Incidence of poverty

- ▶ Dandekar & Rath
- ▶ M.S. Ahluwalia
- ▶ Planning Commission
- ▶ Rangrajan, C- Rs. 32 & Rs. 47
 - ▶ Rs. 972 & Rs.1407
 - ▶ 29.5% poor
- ▶ Tendulkar report- Rs.27 & Rs. 33.
 - ▶ Rs. 816 & Rs.1000
 - ▶ 22%

An Example of Size Distribution of Consumption Expenditure

▶ MPCE	%Population
▶ 0-150	3.2
▶ 150-200	4.0
▶ 200-250	6.5
▶ 250-300	8.6
▶ 300-340	10.0 (half of 10% are below poverty line 320)
▶ 340-400	11.3
▶ 400-450	8.6
▶ 450-500	9.2
▶ 500-550	9.3
▶ 550-650	11.4
▶ 650-800	8.9
▶ 800-1000	5.0
▶ Above 1000	4.0

Measures of Poverty

- ▶ **Head count Index:** Ratio of population living in poor households with consumption expenditure / income below poverty line to total population.
- ▶ **Poverty Gap Index:** (how poor they are)
 - ▶ Mean differences in poor people's consumption expenditure or income from poverty line
- ▶ **Squared poverty gap Index:** Adds the dimension of inequality among the poor to the poverty gap index. For a given value of the PGI, population with greater dispersion of income among poor indicates a higher value for the SPG.

Measures of Inequality

- ▶ **Lorenz curve:** Represents relationship between cumulative proportion of income and cumulative proportion of population in income distribution by size, beginning with the lowest income group. If perfect income equality, Lorenz curve coincides with 45-degree line.
- ▶ **Gini- Coefficient:** Ratio of area between Lorenz curve and 45-degree line, expressed as a percentage of area under 45-degree line.
 - ▶ 0- perfect equality
 - ▶ 1- Perfect inequality

Multi-dimensional poverty

- ▶ Criticism of money-metric poverty
- ▶ Poverty is not uni-dimensional
- ▶ A person experience poverty beyond income
- ▶ Welfare depends: monetary & non-monetary
- ▶ Income/ expenditure- sole indicator inappropriate
- ▶ Multi-dimensional approach:
 - ▶ size of landholding, type of house, availability of clothing, food security, sanitation, ownership of consumer durables, literacy status, status of labour, means of livelihood, status of children, type of indebtedness, empowerment

Social Exclusion

- ▶ Developed by R.Lenoir & Amartya Sen developed it through poverty & deprivation
- ▶ social exclusion is the denial of the basic welfare rights which provide citizens positive freedom to participate in the social and economic life and which thereby render meaningful their fundamental negative freedoms"

Contd...

- ▶ A set of processes by which individuals, households or communities are pushed towards or kept at margin of society.
- ▶ Not only material deprivation but also denial of opportunities to participate in social or civil life.

Indicators of Social Exclusion

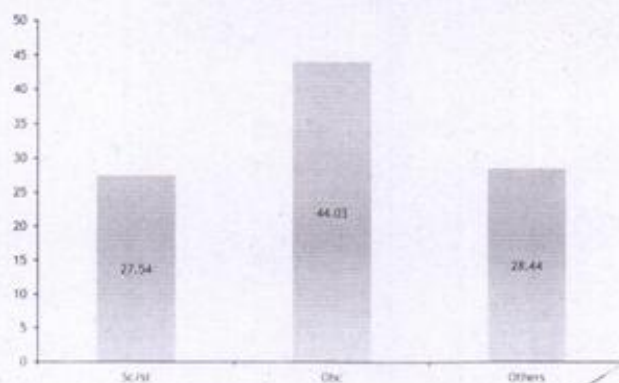
- ▶ Socially :the exclusion is based on caste untouchability such as the excluded member of a community do not have the right to entry to public places.
- ▶ In economic spheres : denial of equal rights and opportunities to the low caste groups, assets like agricultural lands and non land assets employments social needs like education, health, housing and others which leads to lower income and high poverty among the Dalits.

- ▶ **Education:** Discriminations in schools take the form of denial of access to education and the skill development among the Dalits children. This reduces the quality of human resources and reduces the employability for quality jobs and force them to fall back on low earning manual wage labor in farming and non farming activities.
- ▶ Denial of education leads to high rate in illiteracy, low functionally literacy and high dropout rates and limited skill development, discrimination in education may cause high representation in menial jobs, low wages, low income and ultimately high poverty.(SThorat.et)

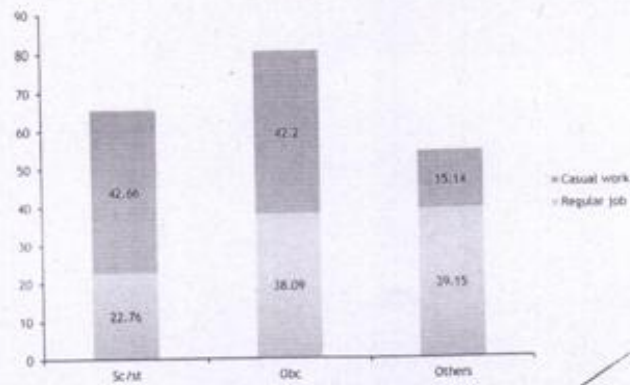
- ▶ **Health facilities:** Denial of visiting to Dalits home , denial of giving information about health facilities, , lack of care leading to requirement of private medical attention and loss of income, delay in complication delivery leading to private medical attention (S.Thorat.et)

- ▶ Political spheres: The Dalits are denied to practice they political rights such as rights and means to participate in the exercise of political powers and denial of justice, freedom of expression, rule of law.
- ▶ Due to this the Dalits can't elect their representative who can participate on their behalf in the policy making.

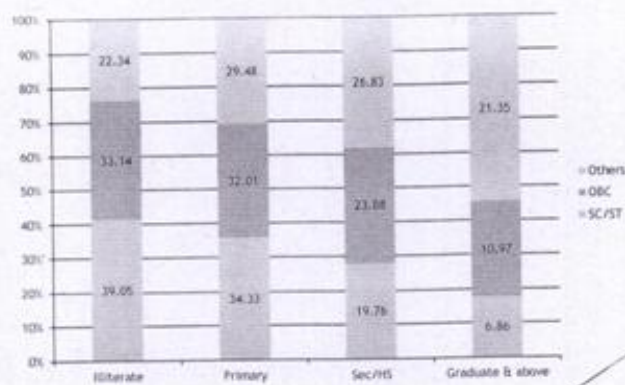
Distribution of population by caste(NSS,68th round)



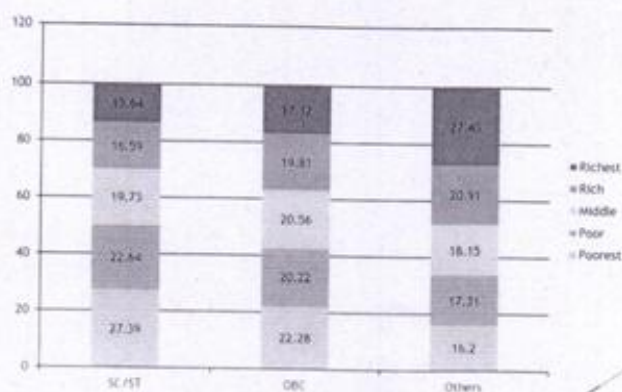
Distribution of population by caste & employment(NSS,68th round)



Distribution of population by caste & education(NSS,68th round)



Distribution of population by caste & Income(NSS,68th round)



Consequences of Exclusion

- ▶ Consequences of denial and discrimination in relation to all the public goods under scrutiny in the report are to further deepen and embed the poverty, exploitation and very low social power of vulnerable populations.
- ▶ Exclusion from one public good reflects, produces and reproduces exclusion from other public goods, and further entrenches the social and economic disadvantages of marginalized persons.
- ▶ For example, exclusion from schooling reduces chances of securing decent housing and decent work.

How to reduce discrimination

- ▶ Changes in law and policy for greater inclusion and justice;
- ▶ Improved implementation of existing laws and policies to secure greater inclusion;
- ▶ Measures to prevent discrimination
 - ▶ School education, decent housing, decent work in labour market

Assignment

- ▶ New Economic Policy, Social exclusion & Safety net
- ▶ Poverty and inequality in Bihar in the current Pandemic scenario.