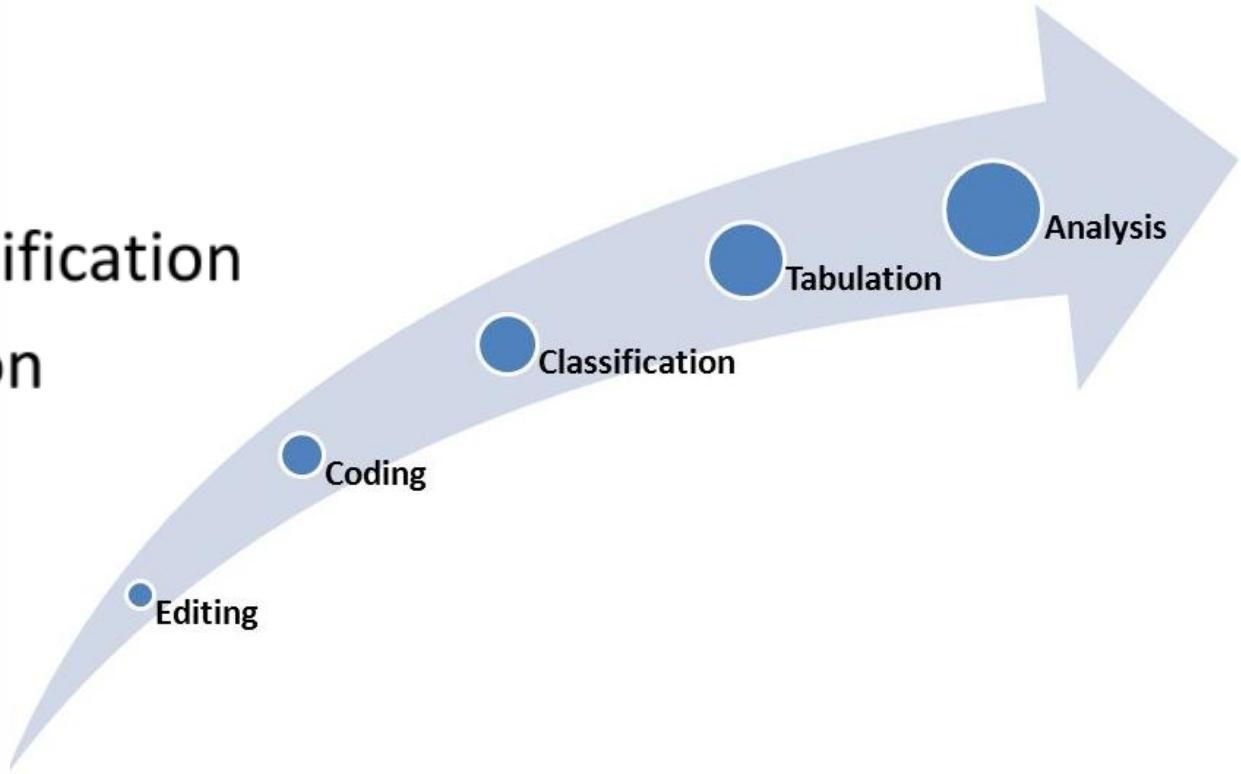
Introduction

- Data analysis and processing are two different process
- DA is a process of inspecting, cleansing, transforming and modelling data with the goal of discovering useful information and conclusion
- DP refers to the process of regrouping or rearranging the sorted data which was analysed previously.

Steps in data processing

- Identifying data
- Editing
- Coding and classification
- Data transcription
- Tabulation
- Summarization



Steps in data processing

1. Identifying data structures

- data format can be identified and arranged by using different software or manually

Eg. SAS (statistical analysis system) or

SPSS(statistical package for social science)

Microsoft Excel

Minitab

- 2. Editing of data: it is done at two levels
  - at the time of recording (Field editing)
  - at the time of data analysis(Central editing)
- Data editing ensure the following things
- Completeness
- Accuracy
- uniformity

Editing

- Incomplete Answers
  - What to do?
  - General rule
- Wrong Answers
  - Dog food example
  - Bookstore example
- Answers that Reflect Lack of Interest
  - Examples

1) EDITING

- Editing of data is a process of examining the collected raw data (especially in surveys) to detect errors and omissions and to correct these when possible.

- Editing is done to assure that the data are accurate, consistent with other facts gathered uniformly entered, as completed as possible and have been well arranged to facilitate coding and tabulation.

- **EDITING** → **FIELD EDITING**
  → **CENTRAL EDITING**

FIELD EDITING

- Field editing consists in the review of the reporting forms by the investigator for completing (translating or rewriting) what the latter has written in abbreviated and/or in illegible form at the time of recording the respondents' responses.

- This type of editing is necessary in view of the fact that individual writing styles often can be difficult for others to decipher.

CENTRAL EDITING

- Central editing should take place when all forms or schedules have been completed and returned to the office.

- This type of editing implies that all forms should get a thorough editing by a single editor in a small study and by a team of editors in case of a large inquiry.

3. Coding

Coding refers to the process of translating answers into a form of numbers

Coding may be numeric , alphab coding

Classification

- Data having common character placed in one class. In this way is classified into different group
- Different types of classification:
- 1. according to attributes. i.e. a education etc
- 2. According to class intervals

Sample record: Excel sheet for two-wheeler owners

Unit Column 1	occupation Column 2	Vehicle Column 3	Km/day Column 4	Marital status Column 5	Family size Column 6
1	4	1	20	1	3
2	3	2	25	2	1
3	5	1	25	1	4
4	2	1	15	2	2
5	4	2	20	2	4
6	5	2	35	2	6
7	1	1	40	1	3
8	5	2	20	2	4

2) CODING

- Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes.
- Coding is necessary for efficient analysis and through it the several replies may be reduced to a small number of classes which contain the critical information required for analysis.
- Coding decisions should usually be taken at the designing stage of the questionnaire.
- It makes it possible to precode the quesistionnaire choices and which in turn is helpful for computer tabulation as one can straight forward key punch from the original questionnaires.

Family	Time	Mode	Activity	Satisfaction	Playground
1	30	1	1, 2, 3	0	N
2	30	3	4,6	1	Y
3	60	2	1, 2	2	Y
4	45	1	5	-1	N
5	30	1	6	1	N
6	60	2	2	2	Y
7	30	3	4	1	N
8	45	2	3, 4	-1	N
9	15	1	6	1	Y
10	60	2	2	2	Y
11	180	4	1, 2, 3, 4	2	Y
12	120	2	1,2,4	2	Y

3) CLASSIFICATION

❑Classification of data which happens to be the process of arranging data in groups or classes on the basis of common characteristics.

❑ Data having a common characteristic are placed in one class and in this way the entire data get divided into a number of groups or classes.

❑ **TYPES OF CLASSIFICATION**
- ACC. TO ATTRIBUTES
- ACC. TO CLASS INTERVAL

ACC. TO ATTRIBUTES

data are classified on the basis of common characteristics which can either be descriptive (such as literacy, sex, honesty, etc.) or numerical (such as weight, height, income, etc.)

ACC. TO CLASS INTERVAL
- EXCLUSIVE TYPE
- INCLUSIVE TYPE

EXCLUSIVE TYPE CLASS INTERVALS: They are usually stated as follows:

10–20
20–30
30–40
40–50

The above intervals should be read as under:

10 and under 20
20 and under 30
30 and under 40
40 and under 50

An item whose value is exactly 30 would be put in 30–40 class interval and not in 20–30 class interval.

INCLUSIVE TYPE CLASS INTERVALS: They are usually stated as follows:

11–20
21–30
31–40
41–50

Thus, an item whose value is 20 will be put in 11–20 class interval.

4. Data transcription

- The collected data are easily transferred into data sheets
- Two types of transcriptions used
- 1. Manual transcription
- 2. Computerized transcription

5. Tabulation

- Tabulation implies tabular representation of data
- Parts of table:
- Table number
- Subheads
- Caption and stubs
- Body of table
- Foot notes

6. Summarization

- Manual data sheets
- Computerized data sheets
- Compilation sheets
- Matrices
- Figures
- Tables
- Graphs

Types of table

1. Frequency distribution table: these table present the frequency and percentage distribution of the information collected
2. Contingency table: this shows the distribution of two nominal variables simultaneously and their totals
3. Multiple response tables: when classification of cases to be done in categories ,then this table is used
4. Miscellaneous tables: when the presentation of data cannot be classified under any of the tables….it is represented in this table.

3) TABULATION

❖Tabulation is the process of summarizing raw data and displaying the same in compact form (i.e., In the form of statistical tables) for further analysis.

❖ In A broader sense, tabulation is an orderly arrangement of data in columns and rows.

Tabulation is essential because of the following reasons

❖It conserves space and reduces explanatory and descriptive statement to a minimum.

❖It facilitates the process of comparison.

❖It facilitates the summation of items and the detection of errors and omissions.

❖It provides a basis for various statistical computations.

8) Those columns whose data are to be compared should be kept side by side. Similarly, percentages and/or averages must also be kept close to the data.

9) It is generally considered better to **approximate figures** before tabulation as the same would reduce unnecessary details in the table itself.

10) It is important that all **column figures** be properly aligned. Decimal points and (+) or (–) signs should be in perfect alignment.

11) **Abbreviations** should be avoided to the extent possible and ditto marks should not be used in the table.

12) **Miscellaneous and exceptional items**, if any, should be usually placed in the last row of the table.

13) **The arrangement of the categories** in a table may be chronological, geographical, alphabetical or according to magnitude to facilitate comparison.

Class (Rs.)	Tally Marks	Frequency Students
20 - 30	𝍸	5
30 - 40	𝍸 III	8
40 - 50	𝍸 IIII	9
50 - 60	𝍸 𝍸	10
60 - 70	𝍸 I	6
70 - 80	II	2
Total		**40**

	Pizza Rolls	Chips and Dip	Cookies	**Totals**
Poker	10	3	12	**25**
Trivial Pursuit	8	14	7	**29**
Monopoly	14	17	7	**38**
Wii Bowling	12	7	4	**23**
Totals	**44**	**41**	**30**	**115**

Activities	Male		Female		Total	
	No.	%	No.	%	No.	%
At home						
TV	6	3.6	2	1.1	8	2.2
Radio	3	1.8	1	0.5	4	1.1
Cassette player	2	1.2	0	0	2	0.6
Lying idle	49	29.3	162	86.6	211	59.6
Others	22	13.2	13	6.9	35	9.9
Outside home	93	55.7	15	8.0	108	30.5
Base	167		187		354	

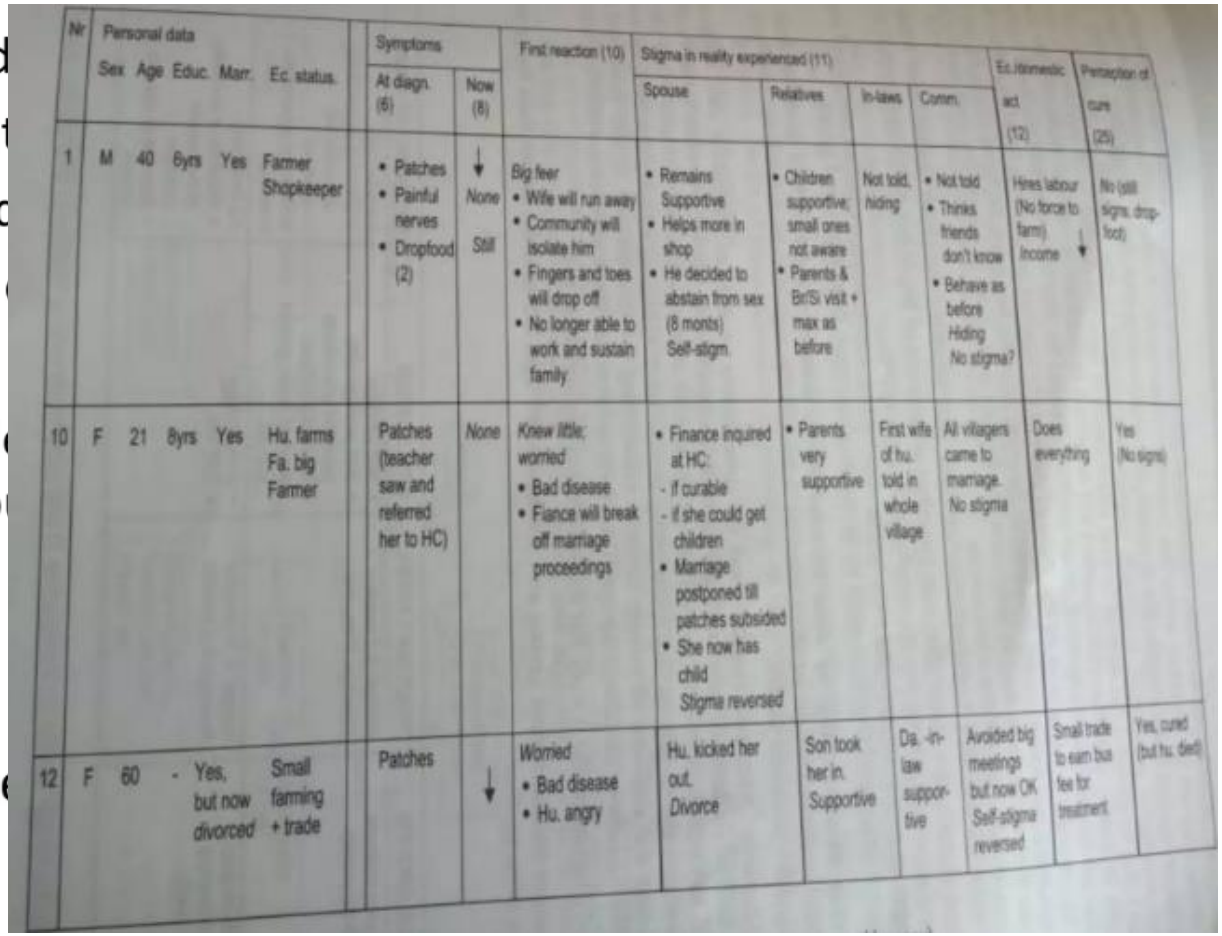
Sex $\chi^2_{(1)} = 96.124$, $P < 0.001$

Summarizing data

- Manual analysis of data is d
  - Master sheet for quantit
  - Compilation sheets for c
- Master sheets are made in
- Simple checklist

All answers must be code of the sheet and answers shou appropriate columns

- Compilation sheets
- In this the topics are cate different categories and
- Other



Nr	Personal data: Sex	Age	Educ.	Marr.	Ec. status	Symptoms: At diagn. (6)	Now (8)	First reaction (10)	Stigma in reality experienced (11): Spouse	Relatives	In-laws	Comm.	Ec./domestic act. (12)	Perception of cure (25)
1	M	40	6yrs	Yes	Farmer Shopkeeper	• Patches • Painful nerves • Dropfood (2)	↓ None Still	Big fear • Wife will run away • Community will isolate him • Fingers and toes will drop off • No longer able to work and sustain family	• Remains Supportive • Helps more in shop • He decided to abstain from sex (8 months) Self-stigm	• Children supportive; small ones not aware • Parents & Br/Si visit + max as before	Not told, hiding	• Not told • Thinks friends don't know • Behave as before Hiding No stigma?	Hires labour (No force to farm) ↓ Income	No (still signs, drop-foot)
10	F	21	8yrs	Yes	Hu. farms Fa. big Farmer	Patches (teacher saw and referred her to HC)	None	Knew little; worried • Bad disease • Fiance will break off marriage proceedings	• Finance inquired at HC: - if curable - if she could get children • Marriage postponed till patches subsided • She now has child Stigma reversed	• Parents very supportive	First wife of hu. told in whole village	All villagers came to marriage. No stigma	Does everything	Yes (No signs)
12	F	60	-	Yes, but now divorced	Small farming + trade	Patches	↓	Worried • Bad disease • Hu. angry	Hu. kicked her out. Divorce	Son took her in. Supportive	Da. -in-law suppor-tive	Avoided big meetings but now OK Self-stigma reversed	Small trade to earn bus fee for treatment	Yes, cured (but hu. died)

GENERALLY ACCEPTED PRINCIPLES OF TABULATION:

1) Every table should have a clear, concise and **adequate title and** this title should always be placed just above the body of the table.

2) Every table should be given a **distinct number** to facilitate easy reference.

3) The column **headings** (captions) and the row headings (stubs) of the table should be clear and brief.

4) The **units of measurement** under each heading or sub-heading must always be indicated.

5) **Explanatory footnotes**, if any, concerning the table should be placed directly beneath the table, along with the reference symbols used in the table.

6) **Source or sources** from where the data in the table have been obtained must be indicated just below the table.

7) Usually the columns are separated from one another by lines which make the table more readable and attractive.