DATA PROCESSING AND ANALYSIS

- **EDITING**
- **CODING**
- **CLASSIFICATION**
- **TABULATION**

You have extracted or collected the required information……

- What to do with this information?
- How to find answers to your research questions?
- How to make sense of information collected?
- How to draw meaning from the data gathered?
- How to accept or reject the hypothesis?
- How to analyse the information to achieve the objectives of study?

To answer these questions you need to subject your data to number of procedures

DATA PROCESSING

Irrespective of method of data collection, the information collected is called ***raw data***, simply ***data.***

After collection, data has to be **processed** and **analysed** in accordance with the outline laid down for the purpose at the time of development of research plan.

Processing implies **editing, coding, classification and tabulation of collected data** so that they are amenable to analysis…summarizing the collected data and organizing these in a manner that they answer the research questions

DATA ANALYSIS

Analysis refers to the **computation of certain measures** along with searching for patterns of relationship that exist among data groups.

The relationships or differences supporting or conflicting with hypothesis are subjected to **statistical tests of significance** to determine "with what validity data indicates certain conclusions"?

EDITING…Cleaning the data

Sometimes even the best investigators can

- Forget to ask a question
- Forget to record a response
- Wrongly classify a response
- Write only half a response
- Write illegibly

Scrutinizing the collected research instruments (questionnaires/ schedules) to identify and *minimize errors, incompleteness, misclassification and inconsistencies* and to assure the data have been well arranged to facilitate coding and tabulation.

EDITING

The editing may be done at two stages

1. Field editing
2. Central editing

1. ***Field editing:*** review of the reporting forms by the investigator for completing if anything has been recorded in abbreviated/ illegible form at the time of recording the respondent's responses.
   - Should be done as soon as possible after the interview, preferably on the very day or on the next day.
   - ***Do not correct errors of omission***

2. ***Central editing:*** takes place when all forms and schedules have been completed and returned to the office.

   May be done by a single editor in a small study or by a team of editors in case of a large inquiry.

EDITING

Check if

- *The information has been recorded completely*
- *Internal consistency*
- *Entries have been made at right place*
- *Entries have been made in right unit (Eg. Recording duration in months when it is required in weeks)*
- *Responses in abbreviations*
- *Illegibility in writing*

EDITING

Procedures:

- **By inference -** in case of inappropriate or missing replies, the proper answer may be determined by reviewing the other information in the schedule (related questions).

  Be careful …new errors may be introduced into the data

- ***By recall -*** if the data is collected by means of interviews, it might be possible for the interviewer to recall respondent's answers …*be extremely careful*

- ***By going back to the respondent -*** If the questionnaires contains some identifying information, the respondent may be contacted through visit/phone to confirm or ascertain an answer….*expensive and time consuming*

EDITING

Two ways to proceed

1. Examine all the answers to **one question** or variable at a time
2. Examine all the responses given to all the questions by **one respondent** at a time

The second method helps to assess internal consistency.

The editor must **strike out the answer** if the same is inappropriate and he has no basis for determining the correct answer or response…enter 'no answer'

All **wrong replies**, which are quite obvious, **must be dropped** from the final results

EDITING

Rules for editors

1. They should be familiar with the instructions given to the interviewers, coders and editing instructions supplied to them
2. While crossing out some entry, just draw a single line on it so that it remains legible
3. If any entries are to made on the form, make them in a distinctive colour and that too in a standardized form
4. Initial all the answers that they change or supply
5. Editor's initials and the date of editing should be placed on each completed form or schedule

CODING

Process of assigning numerals or other symbols to answers so that responses may be put into a limited number of categories or classes

Think about the coding procedure before you start collecting data – precode where possible. If you do not precode..first step after collecting data is to CODE

Method of coding is determined by

1. The way variable has been measured (measurement scale) in your research instrument
2. Whether a question is open ended or closed
3. The way the findings for that variable are to be communicated to the readers – types of statistical procedures to be applied

CODING

Examine the responses to each question and place them in appropriate category. While developing category there are four important considerations.

1. The categories should be mutually exclusive (non-overlapping). It should not be possible to place a response in two categories.
2. The categories should be exhaustive, i.e. it should be possible to place almost every response in one of the categories
3. The use of 'other' category. A waste basket for those odd responses that cannot be put into any category.

   Use of 'other' category must be kept to an absolute minimum...should not include more than 5 percent of the total responses and should not contain more responses than any other category.

4. Categories should be uni-dimensional. Every class should be defined in terms of only one concept.

CODING

In quantitative research all responses can be classified into one of the following three categories..

1. Quantitative responses

   (measured on interval or ratio scale)
2. Categorical responses

   (which may be quantitative or qualitative)
3. Descriptive responses

   (invariably qualitative)

Assigning numerical codes to responses c/a Coding

Content Analysis

How we can analyse the information obtained in response to a question depends upon how a question was asked, and how a respondent answered it...i.e. the measurement scale.

CODING

Content Analysis

Identify main themes that emerge from descriptions given by respondents in answer to questions. Thereafter, three ways to deal with them:

1. **Examine verbatim responses and integrate them into the text of your report to support or contradict your statement.**
2. **Assign a code to each theme and count how frequently each has occurred**
3. **Combine both methods**

Choice is determined by the way your wish to communicate your findings to your readers

CODING

FOUR STEPS OF CODING

1. **DEVELOPING A CODE BOOK**
2. **PRE-TESTING THE CODE BOOK**
3. **CODING THE DATA**
4. **VERIFYING/CLEANING THE CODED DATA**

A code book is a document describing the **set of rules for assigning numerical values to answers obtained from respondents and computer file location of the data in a format that computers can use.**

Always make multiple copies- if you lose it ..will have to recode again

TWO formats for data entry – 'fixed' and 'free'

FIXED FORMAT

1. A piece of information obtained from an individual is entered in a specific column
2. Each column has a ‘column number’
3. ‘Col. No.’ in the code book refers to the column in which a specific type of information is to be entered in the code book
4. Information about individual is entered in ‘rows’

Table 15.1 An example of a code book

Col. no.	Q. no.	Variable name	Response pattern	Code
1	2	3	4	5
1–3	S. no.	ID	Actual serial number	Code actual
4	Record no.	RNO	First record	1
			Second record	2
			Third record	3
5	1(a)	Age	20–24	1
			25–29	2

Source: Ranjit Kumar, 2014. pp.301

questionnaire

1 Please indicate:

(a) Your current age in completed years: _____
(c) Your marital status: (Please tick)
Currently married _____
Living in a de facto relationship _____
Separated _____
Divorced _____
Never married _____

2 (b) If tertiary/university, please specify the level achieved and area of study. (Please specify all postgraduate qualifications.)

Level of achievement	*Area of study: e.g. engineering, accounting*
Associate diploma	
Diploma	
Bachelor	
Graduate diploma	
Masters	
PhD	

11 What, in your opinion, are the main differences between your jobs prior to and after redeployment?

_____

_____

12 We would like to know your perception of the two jobs *before* and *after* redeployment with respect to the following aspects of your job. Please rate them on a five-point scale using the following guide: 5 = extremely satisfied, 4 = very satisfied, 3 = satisfied, 2 = dissatisfied, 1 = extremely dissatisfied

Before redeployment					*Areas*	*After redeployment*				
1	2	3	4	5		1	2	3	4	5
					Job status					
					Job satisfaction					
					Motivation to work					
					Interest in the job					
					Self-esteem					
					Professional competence					
					Peer interaction					
					Morale					
					Work environment					
					Social interaction					

Figure 15.2 Example of questions from a survey

Source: Ranjit Kumar, 2014. pp.299

FIXED FORMAT

Table 15.1 An example of a code book

Col. no.	Q. no.	Variable name	Response pattern	Code
1	2	3	4	5
1–3	S. no.	ID	Actual serial number	Code actual
4	Record no.	RNO	First record	1
			Second record	2
			Third record	3
5	1(a)	Age	20–24	1
			25–29	2

- ***Column 1*** **refers to columns in which a particular piece of information is to be entered.**
- ***Column 2*** **identifies the question number in the research instrument for which the information is being coded**
- ***Column 3*** **refers to name of the variable (should be unique & rules of respective softwares be kept in mind)**
- ***Column 4*** **lists possible responses to various questions**
- ***Column 5*** **lists the actual codes of the code book that you decide to assign to a response**

CONSIDERATIONS

Column 4 (possible responses to various questions)

- Developing a response pattern differs with type of questions (open ended or closed)
- If question is closed, response pattern is developed as part of instrument construction. All you need to do is assign a numerical value to each response category
- If the closed question includes 'other' category, to accommodate any response that you may not have listed

  (1) analyse the responses and

  (2) assign them to non-overlapping categories as we do for open ended questions and

  (3) add to the already developed categories assign each a numerical value

  (4) If it is multiple response question, number of columns assigned should be in accordance with number of responses to be asked

CONSIDERATIONS

Column 4 (Coding of open ended questions -OEQ) - CONTENT ANALYSIS

- Select a number of schedules/questionnaires randomly
- Select the first OEQ and write down responses on a sheet of paper
- If a person has given more than one response, write them separately on the same sheet
- Select another OEQ. Write down the responses in the same manner on a separate sheet
- Similarly for all OEQs…remember response to each question should be written on a separate sheet.
- Now select another questionnaire/schedule and add the responses given for the same question on the sheet for the question.
- Continue the process till you feel that responses are being repeated and that you are getting no new ones

CONSIDERATIONS

Column 4 (Coding of open ended questions -OEQ) - Content analysis

- Now examine the responses to each question to ascertain similarities and differences
- If two or more responses are similar in meaning, not necessarily language, combine them in one category
- Give a name to the category that is descriptive of the responses
- CODE THE CATEGORIES, NOT THE RESPONSES
- Write down the different responses under each category in the code book so that, while coding, you know the type of responses you have grouped under a category

REMEMBER THE RULES

1. **Exclusiveness**
2. **Exhaustiveness**
3. **< 5% in 'Other' category**

Important

- You need to keep the variation in respondents' answers
- On the other hand, you need to break them up into meaningful categories to identify the commonalities

CONSIDERATIONS

Column 5 (Assigning actual codes)

- You may assign any numerical value to any response as long as you don't repeat it within the same question.
- For responses as 'not applicable' or 'no response' select a number that can be used for these responses for all or most questions.
- For eg. If possible responses are limited to 2 or 3, then assign 8 or 9 to N.A. or N.R.

1 Please indicate:

(a) Your current age in completed years: _____
(c) Your marital status: (Please tick)
Currently married _____
Living in a de facto relationship _____
Separated _____
Divorced _____
Never married _____

2 (b) If tertiary/university, please specify the level achieved and area of study. (Please specify all postgraduate qualifications.)

Level of achievement	*Area of study: e.g. engineering, accounting*
Associate diploma	
Diploma	
Bachelor	
Graduate diploma	
Masters	
PhD	

11 What, in your opinion, are the main differences between your jobs prior to and after redeployment?

_____

_____

12 We would like to know your perception of the two jobs *before* and *after* redeployment with respect to the following aspects of your job. Please rate them on a five-point scale using the following guide: 5 = extremely satisfied, 4 = very satisfied, 3 = satisfied, 2 = dissatisfied, 1 = extremely dissatisfied

Before redeployment					*Areas*	*After redeployment*				
1	2	3	4	5		1	2	3	4	5
					Job status					
					Job satisfaction					
					Motivation to work					
					Interest in the job					
					Self-esteem					
					Professional competence					
					Peer interaction					
					Morale					
					Work environment					
					Social interaction					

Figure 15.2 Example of questions from a survey

Table 15.1 An example of a code book

Col. no.	Q. no.	Variable name	Response pattern	Code
1	2	3	4	5
1–3	S. no.	ID	Actual serial number	Code actual
4	Record no.	RNO	First record	1
			Second record	2
			Third record	3
5	1(a)	Age	20–24	1
			25–29	2
			30–34	3
			35–39	4
			40–44	5
			45–49	6
			No response	9
6	1(c)	MS	Currently married	1
			Living in a de facto relationship	2
			Separated	3
			Divorced	4
			Never married	5
			No response	9
	2(b)	TEDU	Assoc. Dip.	1
			Diploma	2
			Bachelors	3
			Grad. Dip.	4
			Masters	5
			PhD	6
			Not applicable	8
			No response	9

CODE BOOK

Open ended quantitative question

Question 1(a)

- Your current age in completed years …….

1	S. no.	ID	Actual serial number	Code actual
1-3	Record no.	RNO	First record	1
4			Second record	2
			Third record	3
5	1(a)	Age	20-24	1
			25-29	2
			30-34	3
			35-39	4
			40-44	5
			45-49	6
			No response	9

- Determine the range of responses
- Divide into required number of categories- decide in light of purpose of analysis, how you want to communicate findings of your study and whether the findings are going to be compared with some another study?
- Assign codes to categories. Assign a code to 'No response'

CODE BOOK

Closed categorical question

Question 1(c)

Your marital status : (Please tick)

Currently married ……..

Living in a de facto relationship ……

Separated…….

Divorced……..

Never married……

Just assign a numerical value to each category

Add 'no response' as another category and assign a code to it

In which column of the code sheet the response to this question will be entered??

			No response	
6	1(c)	MS	Currently married	1
			Living in a de facto relationship	2
			Separated	3
			Divorced	4
			Never married	5
			No response	9

CODE BOOK

Multiple response question

Question 2(b)

If tertiary/university, please specify the level achieved and the area of study (Please specify all post graduate qualifications).

Level of achievement	*Area of study: e.g. engineering, accounting*
Associate diploma	
Diploma	
Bachelor	
Graduate diploma	
Masters	
PhD	

- *The question asks for two aspects*
- *(1) level of achievement which is categorical*
- *(2) Area of study which is open ended*

A person may have more than one qualification - thus it is a multiple response question

CODE BOOK

- *Scrutinize the responses*
- *Determine the maximum number of levels of achievement identified by the study population, say 3*
- *First code the levels of achievement , and then the areas of study*

	2(b)	TEDU	Assoc. Dip. 1
			Diploma 2
			Bachelors 3
			Grad. Dip. 4
			Masters 5
			PhD 6
			Not applicable 8
			No response 9
7		TEDU1	Same as in TEDU Code
8		TEDU2	Same as in TEDU as in
9		TEDU3	Same as in TEDU TEDU
		STUDY	Behavioural Sciences 1
			Business 2
			Economics/Commerce 3
			Communication 4
			Engineering 5
			Geography 6
			History 7
			Graphics 8
			Librarianship 9
			Nursing 10
			11

(1) level of achievement

(2) Area of study

CODE BOOK

- *Codes 01 to 22 assigned to different qualifications (STUDY), 88 to 'Not Applicable' and 99 for 'No Response'*
- *If a respondent has only one qualification, assign 88 to STUDY 2 and STUDY 3*

			Psychology	14
			Education	15
			Chartered Acct.	16
			Zoology	17
			Anthropology	18
			Social Sciences	19
			Public/Health Admin.	20
			English	21
			Audio-visual Aids Education	22
			Not applicable	88
			No response	99
10-11	2(b)	STUDY 1	Same as in STUDY	Code as in STUDY
12-13		STUDY 2	Same as in STUDY	
14-15		STUDY3	Same as in STUDY	

CODE BOOK

Open ended question

Question 11

What in your opinion are the main differences between your jobs prior to and after redeployment?

.....

.....

.....

.....

.....

.....

Common themes identified

Do not worry about large no. of categories in beginning.. they can be combined later on

11	DIFWK		– current job	01
			– previous job	02
			Less responsibility	
			– current job	03
			– previous job	04
			Low morale	
			– current job	05
			– previous job	06
			Lack of recognition of hard work	
			– current job	07
			– previous job	08
			Skills irrelevant	
			– current job	09
			– previous job	10
			Repetitious nature	
			– current job	11
			– previous job	12
			Job more segmented	
			– current job	13
			No change	78
			Irrelevant response	79
			Other	80
			Not applicable	88
			No response	99
16-17	11	DIFWK1	Same as in DIFWK	Code as
18-19		DIFWK2	Same as in DIFWK	in
20-21		DIFWK3	Same as in DIFWK	DIFWK
22-23		DIFWK4		
24-25		DIFWK5		

- **Content analysis**

Example of responses to open end question - Content analysis.pdf

CODE BOOK

Two way categorical likert scale

Question 12

12 We would like to know your perception of the two jobs *before* and *after* redeployment with respect to the following aspects of your job. Please rate them on a five-point scale using the following guide: 5 = extremely satisfied, 4 = very satisfied, 3 = satisfied, 2 = dissatisfied, 1 = extremely dissatisfied

Before redeployment					***Areas***	***After redeployment***				
1	2	3	4	5		1	2	3	4	5
					Job status					
					Job satisfaction					
					Motivation to work					
					Interest in the job					
					Self-esteem					
					Professional competence					
					Peer interaction					
					Mor					
					Wor					
					Soc					

	12	BEAFTER	Extremely dissatisfied	1
			Dissatisfied	2
			Satisfied	3
			Very satisfied	4
			Extremely satisfied	5
			Same as in BEAFTER	
26		JOBSTA	Same as in BEAFTER	Code as
27		JOBSTB	Same as in BEAFTER	in BEAFTER
28		JOBSTA	Same as in BEAFTER	Code as in
29		JOBSTA	Same as in BEAFTER	BEAFTER
			and so on	

- We assign one abbreviation to one aspect, e.g. job status and differentiate between ‘before’ and ‘after’ by adding ‘B’ and ‘A’ in tail of the abbreviation
- Common coding is assigned to Likert Scale on both sides

CODING

PRE-TESTING THE CODE BOOK

1. It is important to pre-test the code book for any problems before you code your data
2. Pre-test involves selecting a few questionnaires/schedules and actually coding the responses to ascertain problems in coding
3. May be you have not provided for some responses and therefore are unable to code them
4. Change your code book , if needed

C O D I N G

CODING THE DATA

Once code book is finalized , code the raw data

Three ways of doing :

1. **Coding on the questionnaire/ schedule itself (provide space for coding at the time of designing the research instrument.)**
2. **Coding on separate code sheets that are available for purchase**
3. **Code directly into computer using s/w …e.g. SPSS**
4. **If during coding you find some responses that do not fit your pre-determined categories, assign them a code and add these to your code book**

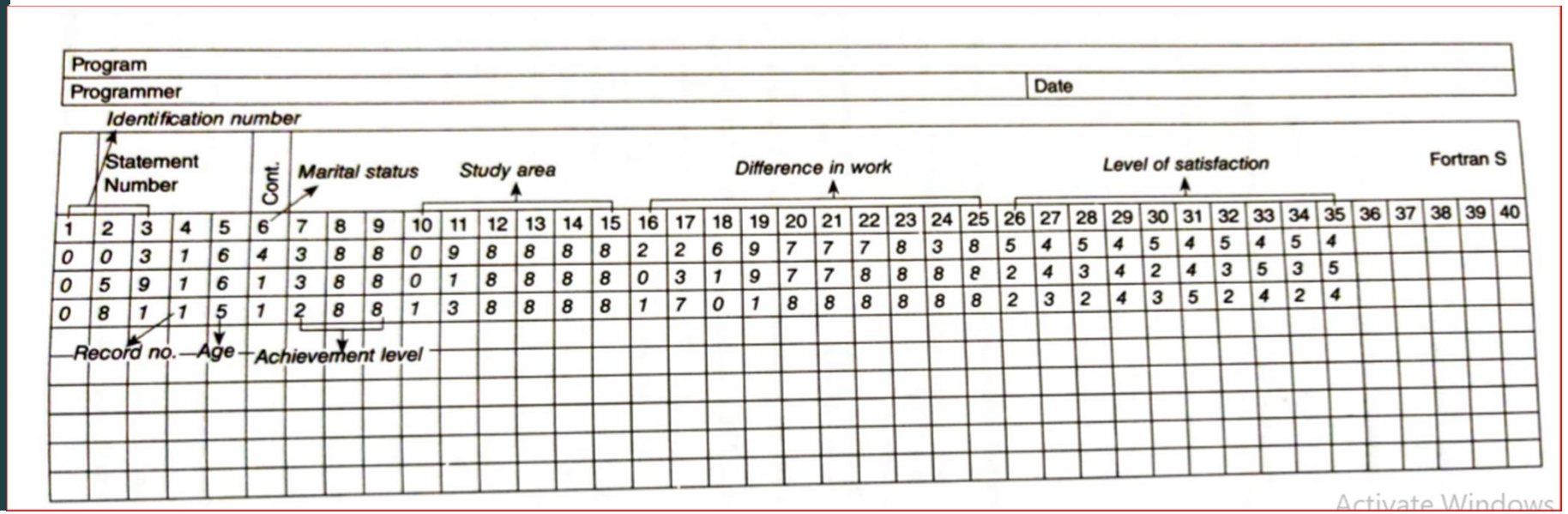
Following is the sample of a coding sheet wherein responses of three respondents (nos. 3,59 and 81) have been coded as per the scheme

CODING

CODING THE DATA

Most computer programs require data to be in grid format

- Each row represents a respondent or case – c/a **data records**
- A column or set of columns represent specific variables – c/a **data fields**
- Each cell, say row 3 column 6, identifies the original source of data – the respondent, questionnaire item and his response to the said question.



CODING

CLEANING THE CODED DATA

You may have selected a perfect sample, perfect measures, no errors in data collection but errors in coding may ruin the entire research project

1. Select few research instruments at random (10 to 15 percent) and code the responses again to identify any discrepancies in coding
2. If you discover no coding errors in recoded sample, proceed
3. If you find errors, recheck all the coding
4. Continue to verify till you are assured of no discrepancies in coding

CODING

VERIFYING CODING

1. **Wild code checking :** Checking all variables for impossible codes

   (If gender has been coded in two categories - 1 & 2, a 4 found in the field for gender variable indicates coding error

2. **Contingency cleaning :** Cross classify variables and look for logically impossible combinations

   For eg. a respondent may have passed only eighth grade and has been recorded as a medical doctor in occupation field

CODING

FOUR METHODS OF DATA ENTRY

1. CODE SHEET - Manual
2. DIRECT ENTRY - pre-programmed entry (For e.g. Google forms)
3. OPTICAL SCAN
4. BAR CODE

DATA CLASSIFICATION

CLASSIFICATION

- Large volume of raw data cannot be easily understood and is not fit for further analysis and interpretation
- Must be reduced into homogenous groups if we were to get meaningful relationships and analyse it.
- Classification is the process of arranging data in homogeneous groups or classes on the basis of common characteristics.
- Separating them into different related parts
- Data having common characteristics are placed in one class and in this way the entire data get divided into a number of groups or classes

OBJECTIVES

Primary objectives of data classification are

1. To consolidate volume of data in such as way that similarities and differences can be quickly understood. Figures can consequently be ordered in a few sections holding common traits.
2. To aid comparison.
3. To point out important characteristics of data at a flash
4. To give importance to prominent data while separating the optional elements
5. To make data amenable to statistical analysis.

CLASSIFICATION

TYPES - FOUR

- Geographical classification
- Chronological classification
- Qualitative classification
- Quantitative qualification

CLASSIFICATION

GEOGRAPHICAL CLASSIFICATION

- Also c/a Spatial classification
- Data are classified with reference to geographical locations or areas such as countries, states, cities….

Eg. Production of food grains in different states of India (in ‘000 tonnes)

CLASSIFICATION

CHRONOLOGICAL CLASSIFICATION

- Also c/a Temporal classification
- Data are grouped according to time, like months, years…
- Classified either in ascending or descending order

Eg. Total rainfall over last 10 years (in cms)

2011	x
2012	y
2013	Z

…..

CLASSIFICATION

QUALITATIVE CLASSIFICATION

- **Classification according to attributes**: Data are classified on the basis of descriptive characteristics (attribute or quality) such as literacy, honesty, gender etc.

  Attribute cannot be measured. Only presence or absence of the qualitative attribute is considered.

  Such classification can be simple or manifold

  Simple : Only one attribute is considered, and divide the universe into two classes-one class consisting of items possessing given attribute and the other class which does not possess the given attribute.

  Manifold : two or more attributes are considered simultaneously. Number of classes 2 raised to power n where n is number of attributes considered.

CLASSIFICATION

QUANTITATIVE CLASSIFICATION

- Data are classified on the basis of some numerical characteristics measuring quantitative phenomenon such as (eg. Income, age, weight…)

Classification according to class interval

Two types of quantitative classification

1. Discrete Frequency distribution
2. Continuous frequency distribution

Two elements

1. Variable
2. Frequency

CLASSIFICATION

VARIABLE

A characteristic (quantity or attribute) which varies or differs/ changes in magnitude or quality from case to case

- Variables which are capable of taking only exact value and not any fraction are c/a Discrete Variables
- Those which can take all possible values (integer as well as fraction) are c/a Continuous Variables.

FREQUENCY

Number of times each variable gets repeated

CLASSIFICATION

STATISTICAL SERIES

Systematic arrangement of statistical data in some logical order

General Characteristics

On what basis data have been arranged?

- Time Series (Chronological)
- Spatial Series (Geographical)

  location considerations
- Condition Series

  quality or quantity

Construction

- Individual Series
- Discrete Series
- Continuous Series

CLASSIFICATION

On basis of construction

- **Individual Series** – Items are listed individually
- **Discrete Series** – Frequency by individual values
- **Continuous Series** – Represents continuous variables. Shows range of values of different items of the series

SERIES	Marks of 10 students out of 10	
Individual	4, 7, 6, 7, 6, 9, 4, 7, 9, 4	
Discrete	Marks	No. of students
	4	3
	6	2 ……and so on
Continuous	Marks	No. of students
	0 - 5	3
	6 - 10	7

CLASSIFICATION

Quantitative classification

- **Frequency distribution**: Data are grouped on the basis of come measurable property in classes

  Each group of class intervals has an '**upper limit**' and '**lower limit**' known as '**class limits**'.

  Difference between the class limits is c/a **class magnitude/ class interval**

  There may be classes with **equal** or **unequal** class magnitude

  Number of items which fall in a particular class is known as **frequency** of that class.

  All classes or groups with their respective frequencies taken together and put in form of a table are described as **frequency distribution**

CLASSIFICATION

Three main questions

- QUESTION 1: How many classes should be there? What should be their magnitudes?
  - No specific rule
  - Objective should be to display data in such a way to make it meaningful for the analysis
  - Typically 5 to 15 classes
  - To the extent possible, equal magnitude should be preferred
  - In some cases unequal magnitude may result in better classification
  - Researcher's objective judgement, skill and experience plays major role
  - Multiples of 2, 5 and 10 are generally preferred while determining class magnitude

Sturges formula for determining class interval

$$i = R/K$$

$$K = (1+3.3 \log N)$$

i = size of class interval

K = No. of Classes

R = Range of data

N = Number of items to be grouped

If one or two items have very high or low values, open ended intervals may be used

Example

For Example: If the total number of observations is 50, the number of classes would be

$$K = 1 + 3.322 \log N$$
$$K = 1 + 3.322 \log 50$$
$$K = 1 + 3.322(1.69897)$$
$$K = 1 + 5.644$$
$$K = 6.644$$

7 Classes approximately.

QUESTION 2 : How to choose class limits? Open or close?

- **Thumb rule :** Mid point of a class interval and the actual average of items of that class interval should remain as close to each other as possible
- Class limits should be located at multiples of 2,5, 10, 20, 100..

Exclusive or **inclusive???**

Exclusive	Inclusive
10-20 20-30 Read as 10 and under 20 20 and under 30 Items whose values are equal to the upper limit of a class are grouped into next higher class Should be preferred for continuous data	11-20 21 - 30 Upper limit of a class interval is also included in the concerned class If stated upper limit is, say, 20 - the real limit is 20.9999999 11-20 means 11 and under 21 Should be preferred for discrete data

Types of continuous series classification

EXCLUSIVE

Age (in Years)	No. of Students
0 – 10	3
10 – 20	5
20 – 30	12
30 – 40	6
40 – 50	4

Upper limit of the class not considered for grouping

INCLUSIVE

Age (in Years)	No. of Students
0 – 9	3
10 – 19	5
20 – 29	12
30 – 39	6
40 – 49	4

Both upper limit and lower limit of the class are considered for grouping

Types of continuous series classification

OPEN ENDED

Age (in Years)	No. of Students
BELOW 10	3
10 – 20	5
20 – 30	12
30 – 40	6
40 & ABOVE	4

Lower limit of first class and upper limit of last class is not given

MID VALUE

Mid Values	(f)
5	3
15	5
25	12
35	6
45	4

Mid values are central value of class interval

Types of continuous series classification

UNEQUAL INTERVAL

(X)	(f)
0 – 10	3
10 – 15	5
15 – 30	12
30 – 40	6
40 – 45	4

Can be converted into equal by splitting or merging classes

CUMULATIVE FD

Less than

Age (in Years)	No. of Students
Less Than 10	3
Less Than 20	8
Less Than 30	20
Less Than 40	26
Less Than 50	30

More than

Age (in Years)	No. of Students
More Than 10	30
More Than 20	27
More Than 30	22
More Than 40	10
More Than 50	4

Obtained by successively adding the frequencies of the classes

- **QUESTION 3 : How to determine class frequencies?**
  - Use tally sheets or mechanical aids

Favorite Pets		
Pet	Tally Marks	Number
	𝍸 𝍸	10
	𝍸	4
	𝍸 I	6

HR Complaints Tally Sheet

Objective To record the frequency/nature of user complaints
Filled in By: Stephanie Famuyide
Location: HR Department
Period 1-5 April 2012

TYPE OF COMPLAINT	FREQUENCY	SUBTOTAL
Personal details incorrect in database	||||	4
Error in salary calculation	||||	4
Leave history not updated	𝍸 𝍸 𝍸	**15**
Delayed training arrangements	|	1
Pension not remitted	|||	3
Incorrect tax calculation	|	1
Working conditions not favourable	|||	3
Other		0

TABULATION

TABULATION

- **Summarizing raw data and arranging it in a concise, logical and compact form – orderly arrangement of data in columns and rows.**
- **Useful means of presenting large amounts of detailed information in a small space**

May be Simple or Complex

WHY?

Conserves space

Reduces descriptive statement to minimum in the text

Serves as a quick reference for the reader

- Facilitates

  Clarification of text

  Understanding of data

  Summation of items

  Comparison

  Detection of errors and omissions

  Statistical computation

STRUCTURE OF A TABLE

Title

Table 1. Attitude towards problem X by age

Column Heading

Attitude towards X Problem	Age of Respondent					Total
	<25	25-34	35-44	45-54	55+	
Strongly in favour	CELL					
In favor						
Uncertain						
Against						
Strongly against						
Total						

Sub-category

Source:.....Hypothetical data

Body

Stub

Notes

TITLE

- Indicates table number and describes the type of data the table contains
- Always placed above the table
- Give a unique number to each table sequentially

  Common method is to identify a table by chapter number followed by number of table within the chapter
- Caption should clearly specify the contents of the table
  - Clearly identify the information contained in the table
  - In case of two variables, identify the dependent variable first in the title
  - Should answer three questions – what, where and when
  - Be short and concise

Eg. “Attitude towards mining by gender, Udaipur (2019) ”

STUB

- First column on the left
- Lists the items about which information is provided in the horizontal rows to the right - *McGraw Hill Style Manual*
- A vertical listing of the categories or individuals about which information is given in the columns of the table – *The Chicago Manual of Style*
- Should be clear and brief

COLUMN HEADING

In bivariate or multivariate tables, subcategories of second variable displayed in the column heading

Should be concise and clear

BODY

- The cells housing the analysed data

SUUPLEMENTARY NOTES OR FOOTNOTES

FOUR TYPES

- Source notes
- Other general notes
- Notes on specific parts of the table
- Notes on level of probability/ statistical significance

Placed directly beneath the table, along with reference symbols used in the table.

SIMPLE Vs COMPLEX

- **Simple tabulation** gives information about one or more groups of independent questions - One way tables - which deal with one characteristic of data at a time

- **Complex tabulation** gives information concerning one or more sets of inter-related questions - **Bivariate** or **Multivariate -** give information about several inter-related characteristics of data. **C/A CROSS TABULATION**

UNIVARIATE (FREQUENCY TABLES)

Table 16.1 Respondents by age (frequency table for one population – hypothetical data)

Age	No. of respondents
<20 years	2 (2.0)
20–24	12 (12.0)
25–29	22 (22.0)
30–34	14 (14.0)
35–39	17 (17.0)
40–44	10 (10.0)
45–49	11 (11.0)
50–54	
55+	
Total	

Note: Figures in parentheses are percentages.

- **Contain information about one variable**

Table 16.2 Respondents by age (frequency table comparing two populations – hypothetical data)

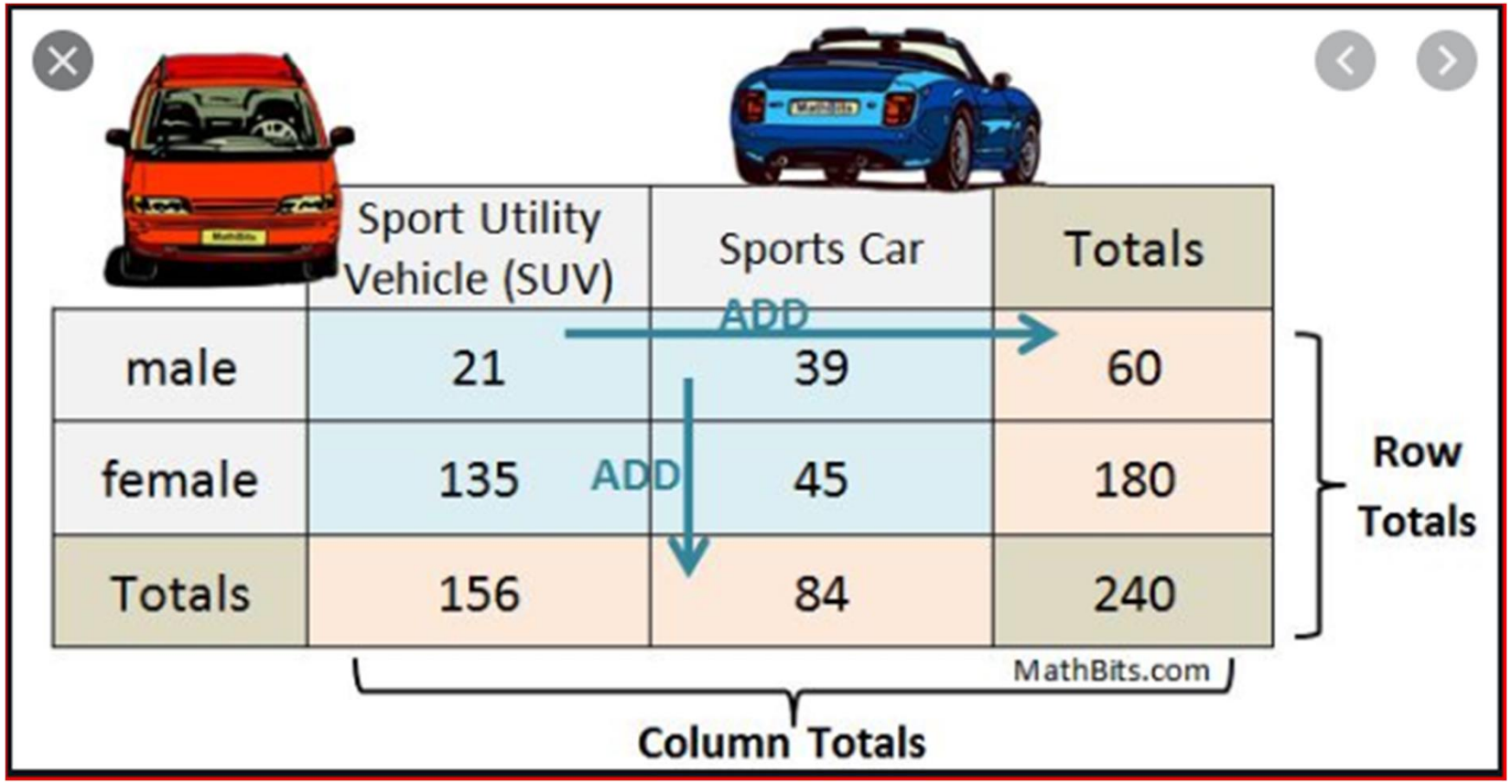
Age	Population A	Population B
<20	2 (2.0)	1 (0.6)
20–24	12 (12.0)	17 (10.9)
25–29	22 (22.0)	23 (14.7)
30–34	14 (14.0)	18 (11.5)
35–39	17 (17.0)	26 (16.7)
40–44	10 (10.0)	16 (10.3)
45–49	11 (11.0)	18 (11.5)
50–54	9 (9.0)	27 (17.3)
55+	3 (3.0)	10 (6.4)
No response	0 (0.0)	0 (0.0)
Total	100 (100.0)	156 (100.0*)

Note: Figures in parentheses are percentages (*these sum to 99.9 due to rounding).

Source: Ranjit Kumar pp 335

BI-VARIATE (CROSS TABULATION)

- **Contain information about two related variables**



BIVARIATE (CROSS TABULATIONS)

- Contain information about two variables

- Understand the difference between Row, Column and Total percentages

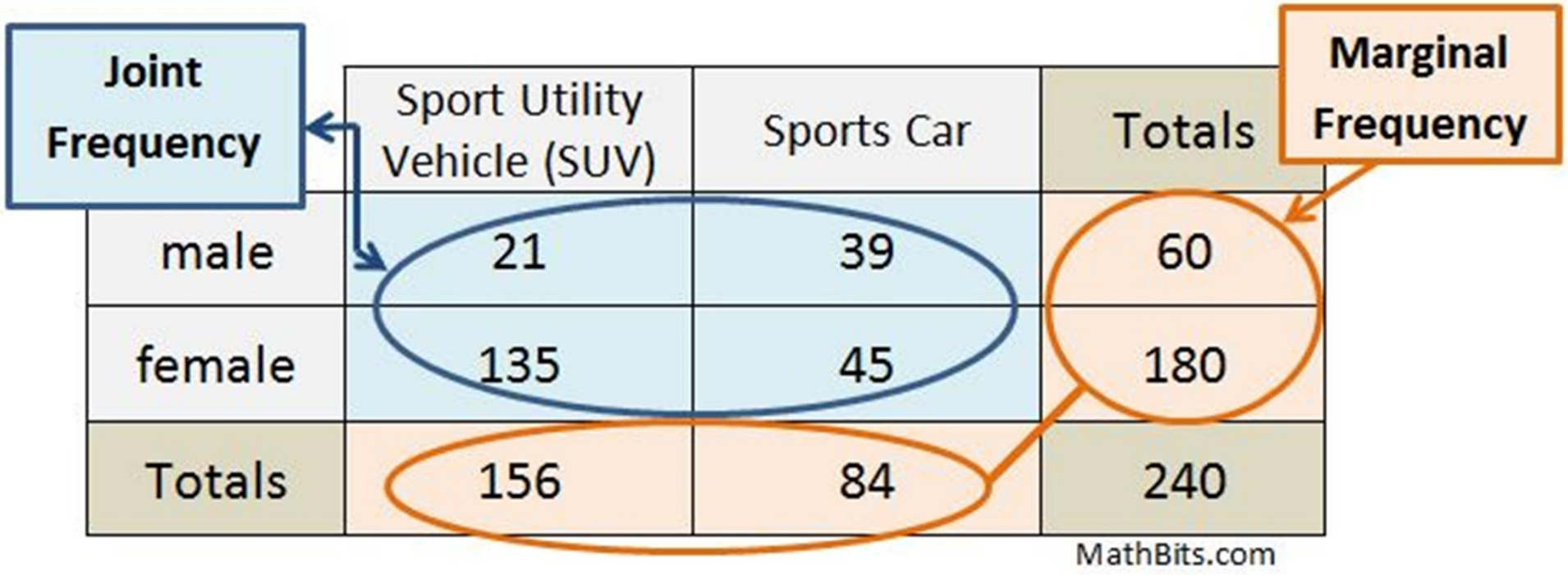
Source: Ranjit Kumar pp 336

Table 16.3 Respondents by attitude towards uranium mining and age (cross-tabulation – hypothetical data)

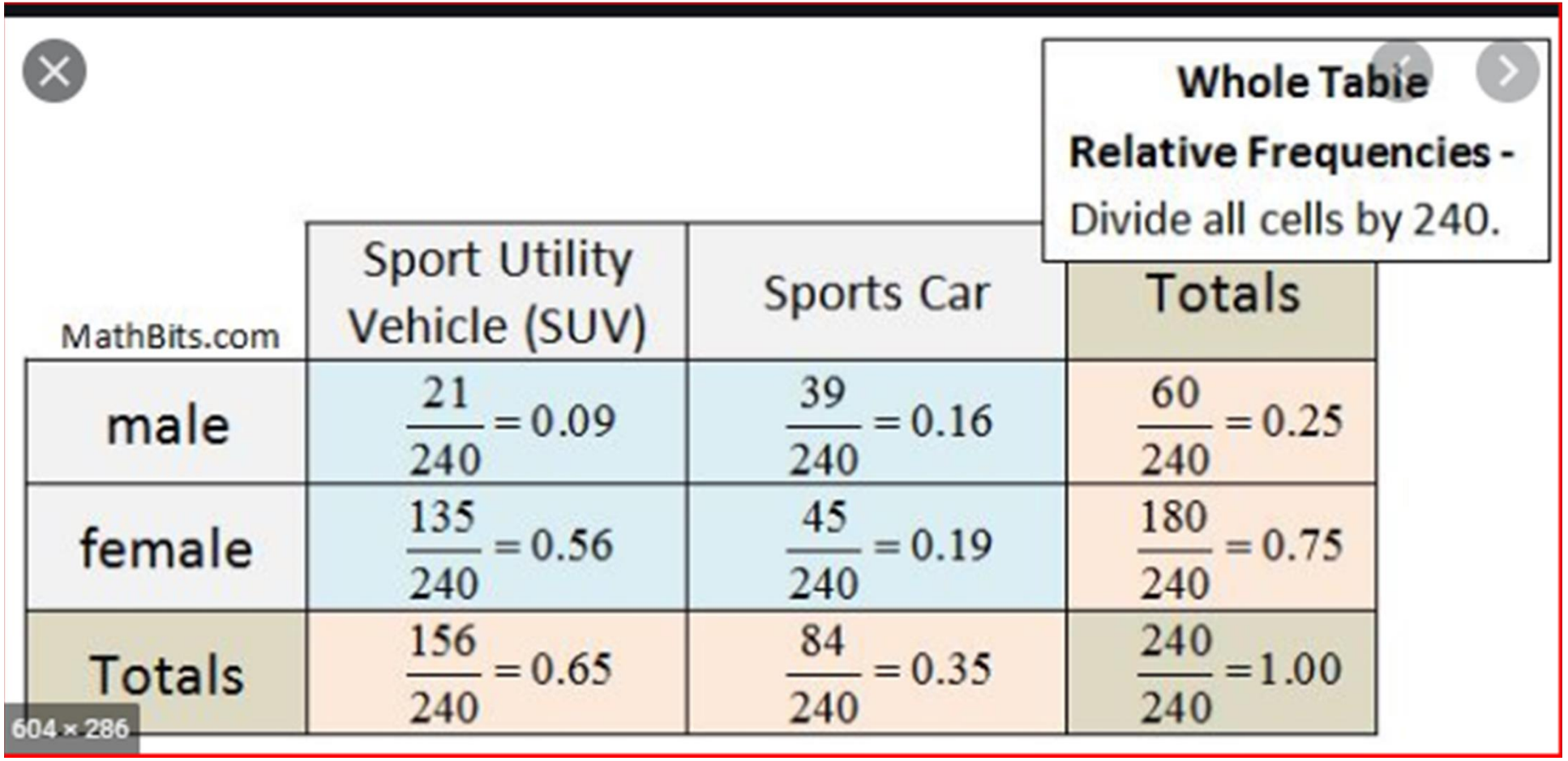
Attitude towards uranium mining	Age					
	<25	25–34	35–44	45–54	55+	Total
Strongly in favour	(0.0)*	(5.5)	(14.8)	(35.0)	(100.0)	
	0	2	4	7	3	16
	(0.0)†	(12.5)†	(25.0)†	(43.6)†	(18.6)†	(100.0)†
	(0.0)*	(8.3)	(18.5)	(20.0)	(0.0)	
In favour	0	3	5	4	0	12
	(0.0)	(25.0)	(41.7)	(33.3)	(0.0)	(100.0)
	(0.0)*	(0.0)	(7.4)	(20.0)	(0.0)	
Uncertain	0	0	2	4	0	6
	(0.0)	(0.0)	(33.3)	(66.7)	(66.7)	(100.0)
	(14.3)*	(19.4)	(3.7)	(0.0)	(0.0)	
Against	2	7	1	0	0	10
	(20.0)	(70.0)	(10.0)	(0.0)	(0.0)	(100.0)
	(85.7)*	(66.7)	(55.6)	(25.0)	(0.0)	
Strongly against	12	24	15	5	0	56
	(21.4)	(42.9)	(26.8)	(8.9)	(0.0)	(100.0)
Total	(100.0)*	(100.0)	(100.0)	(100.0)	(100.0)	(100.0)
	14	36	27	20	3	100

* =column percentage; † = row percentage.

HOW TO READ ??


Joint Frequency
Marginal Frequency
Sport Utility Vehicle (SUV)
Sports Car
Totals
male
21
39
60
female
135
45
180
Totals
156
84
240
MathBits.com

HOW TO READ ??



Whole Table

Relative Frequencies -

Divide all cells by 240.

MathBits.com	Sport Utility Vehicle (SUV)	Sports Car	Totals
male	$\frac{21}{240} = 0.09$	$\frac{39}{240} = 0.16$	$\frac{60}{240} = 0.25$
female	$\frac{135}{240} = 0.56$	$\frac{45}{240} = 0.19$	$\frac{180}{240} = 0.75$
Totals	$\frac{156}{240} = 0.65$	$\frac{84}{240} = 0.35$	$\frac{240}{240} = 1.00$

604 × 286

HOW TO READ ??

Column Relative Frequencies -
Divide SUV column by156.
Divide Sports Car column by 84.
Divide Totals column by 240.

Conditional relative frequencies

	Sport Utility Vehicle (SUV)	Sports Car	Totals
male	$\frac{21}{156} = 0.13$	$\frac{39}{84} = 0.46$	$\frac{60}{240} = 0.25$
female	$\frac{135}{156} = 0.87$	$\frac{45}{84} = 0.54$	$\frac{180}{240} = 0.75$
Totals	$\frac{156}{156} = 1.00$	$\frac{84}{84} = 1.00$	$\frac{240}{240} = 1.00$

MathBits.com

HOW TO READ ??

Row Relative Frequencies -
Divide male row by 60.
Divide female row by 180.
Divide Totals row by 240.

Conditional relative frequencies

	Sport Utility Vehicle (SUV)	Sports Car	Totals
male	$\frac{21}{60} = 0.35$	$\frac{39}{60} = 0.65$	$\frac{60}{60} = 1.00$
female	$\frac{135}{180} = 0.75$	$\frac{45}{180} = 0.25$	$\frac{180}{180} = 1.00$
Totals	$\frac{156}{240} = 0.65$	$\frac{84}{240} = 0.35$	$\frac{240}{240} = 1.00$

MathBits.com

POLY VARIATE (MULTI-VARIATE TABLES)

- **Contain information about more than two variables**

Table 16.4 Attitude towards uranium mining by age and gender (hypothetical data)

Attitude towards uranium mining	Number of respondents												
	<25		25–34		35–44		45–54		55+		Total		
	F	M	F	M	F	M	F	M	F	M	F	M	T
Strongly in favour	0	0	1	1	3	1	5	2	3	0	12	4	16
In favour	0	0	1	2	3	2	3	1	0	0	7	5	12
Uncertain	0	0	0	0	1	1	2	2	0	0	3	3	6
Against	1	1	4	3	1	0	0	0	0	0	6	4	10
Strongly against	4	8	17	7	8	7	2	3	0	0	31	25	56
Total	5	9	23	13	16	11	12	8	3	0	59	41	100

Source: Ranjit Kumar pp 336

PRINCIPLES OF TABULATION

- Every table should be referred in the text
- Units of measurement must be indicated under each heading/ sub-heading
- Columns may be separated by lines which make the table more readable and attractive. Lines should always be drawn at top and bottom of the table and below the headings.
- Thick lines may be used to separate the data under one class from the data under another class. Thin lines may be used to separate sub-divisions of individual class

USE OF BORDERS

Number of Employees in an organization According to age Group

Age (in years)	No. of Employees
Below 25	50
25-35	67
35-45	43
45-55	15
55 and above	5
	Total 180

		Gender	
		Male	Female
		Count	Count
Age category	Less than 25	108	134
	25 to 34	276	351
	35 to 44	309	370
	45 to 54	221	260
	55 to 64	136	184
	65 or older	178	301

PRINCIPLES OF TABULATION

- Columns may be numbered to facilitate reference in text
- Those columns whose data are to be compared should be kept side by side. Percentages should be kept close to the data.
- Better approximate figures before tabulation.
- Relative significance of different categories may be emphasised through variation in indentation, spacing, font etc.
- All column figures/ cell values, + or - signs should be identically aligned

State	Total Population
Uttar Pradesh	199,581,477
Maharashtra	112,372,972
Bihar	103,804,637
West Bengal	91,347,736
Andhra Pradesh	84,665,533
Madhya Pradesh	72,597,565

Type	Freq	cumulative frequency
Up to 1000	22	22
1001-2000	45	67
2001-3000	57	124
3001-4000	97	221
4001-5000	152	373
5001-6000	241	614
6001-7000	153	767

X

OTHER PRINCIPLES OF TABULATION

- Avoid abbreviations in column/row headings
- Place miscellaneous and exceptional items in the last row of the table
- Table should be made as logical, clear, accurate and simple as possible
- DO NOT crowd too large data in a single table
- Total of rows must be placed in extreme right column and total of column at the bottom.
- Arrangement of categories may be chronological, geographical, alphabetical or according to magnitude to facilitate comparison.