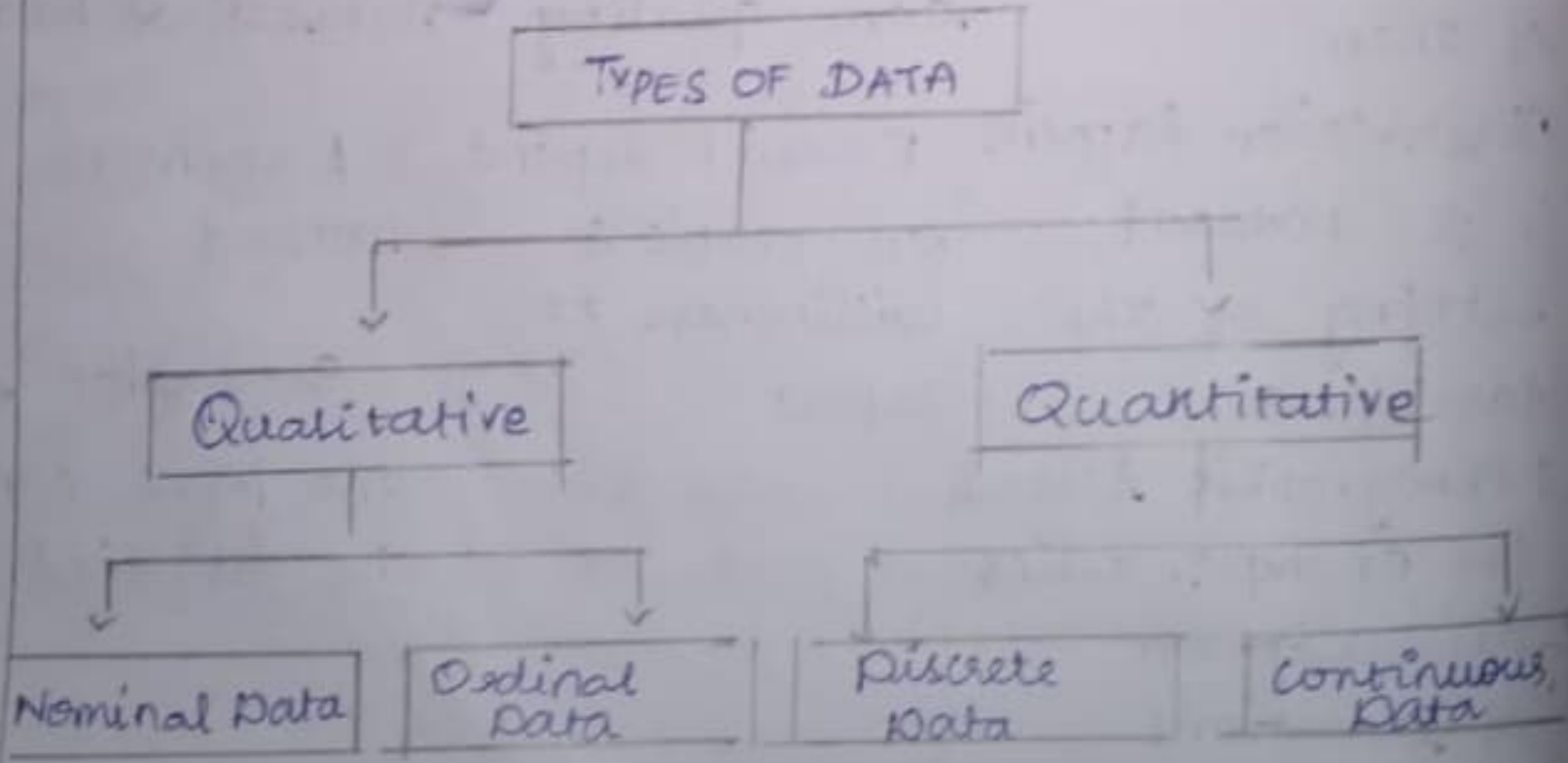
reports).

TYPES OF DATA



QUALITATIVE OR CATEGORICAL DATA.

Qualitative or categorical data that can't be measured or counted in the form of numbers. These types of data are sorted by category, not by number, That's why it is also known as categorical data. These data consist of audio, images, symbols or text, The gender of a person i.e male, female, or others, is qualitative data.

Qualitative data tells about the perception of people. This data helps market researchers understand the customers tastes and then design their ideas and strategies accordingly.

Eg of Qualitative data are

- ☆ What language do you speak.
- ☆ Favorite holiday destination
- ☆ Opinion on something (agree, disagree, or neutral)
- ☆ Colors.

QUALITATIVE - NOMINAL DATA.

Nominal Data is used to label variables without any order or quantative value. The color of hair can be considered nominal data, as one color can't be compared with another color.

The name "nominal" comes from the Latin name "nomen", which means "name". With the help of nominal data, we can't do any numerical tasks or can't give any order to sort the data. These data don't

have any meaningful order; their values are distributed into distinct categories.

Eqs.

* Colour of hair (Blonde, red, brown, black etc)
* Martial status (Single, Married, Widowed).
* Nationality (Indian, German, American)
* Gender (Male, Female, Others)
* Eye color (Black, Brown, etc).

ORDINAL DATA

Ordinal data have ~~neutral~~ natural ordering where a number is present. In some kind of order by their position on the scale. These data are used for observation like customer satisfaction, happiness etc. but we can't do any arithmetical tasks on them.

Ordinal data is qualitative data for which their values have same kind of relative position. These kinds of data can be considered "in-between" qualitative and quantitative data. The ordinal data only shows the sequences and cannot use for statistical analysis. Compared to nominal data, ordinal data have same kind of order that is not present in nominal data.

Eg:s:

* When companies ask for feedback, experience or statisfaction on a scale of 1 to 10
* Letters grades in the exam (A, B, C, D, etc)
* Ranking of people in a competition (First, second, third, etc).

* Economic Status (High, Medium, and Low)
* Education level (Higher, secondary, primary)

Diff between Nominal & Ordinal Data.

NOMINAL DATA	ORDINAL DATA
* Nominal Data can't be quantified, neither they have any intrinsic ordering.	* Ordinal data gives some kind of sequential order by their position on the scale.
* Nominal data is qualitative data or categorial data.	* Ordinal data is said to be in-between "qualitative and quantative data.
* They don't provide any quantitative value, neither can we perform any arithmetical operation.	* They provide sequence and can assign numbers to ordinal data but cannot perform the arithmetical operation.
* Nominal data cannot be used to compare with one another.	* Ordinal data can help to compare one item with another by ranking or ordering.
Egs: Eye color, housing style, gender, hair color, religion, marital status, ethnicity etc.	Egs: Economic status, customer satisfaction, education level, letter, grades etc.

QUANTITATIVE OR NUMERICAL DATA

Quantitative data is also known as numerical data which represents the numerical value (i.e, how much, how often, how many). Numerical data gives information about the quantities of a specific things. Some eg. of numerical data are height, length, size, weight, and so on. The quantitative can be classified

into two different types based on the data sets. The two different classifications of numerical data are discrete data and continuous data

DISCRETE DATA

Discrete data can take only discrete values. Discrete information contains only a finite number of possible values. Those values cannot be subdivided meaningfully. Here, things can be counted in whole numbers.

Eg: No. of students in the class

CONTINUOUS DATA.

Continuous data in data that can be calculated. It has an infinite number of probable values that can be selected within a given specific range. Continuous data is information that occurs in a continuous series. Continuous data includes any values that falls inside a range.

Eg: The weight of the truck, the height of children, the speed of the cars.

- Gender (women, men)
- Hair color (blonde, brown)
- Ethinicity (Hispanic, Asian)

Nominal Data

- First, second, and third
- Letter grades: A, B, C
- Economic Status: low, medium

Ordinal data

QUANTITATIVE DATA

TYPES OF DATA

QUANTITATIVE DATA

Discrete Data

- The no. of students in a class
- The no. of workers in a company
- The no. of home runs in a basketball game

Continuous Data

- The height of children
- The square footage of a two-bedroom house
- The speed of cars.

DATA TABLE: DEFINITION

If the no. of observations is huge, then organising data in ascending or descending or serial order is a tedious job. The directive doesn't tell us much except maybe the minimum(s)

and maximum(s) of data. So the data make it easy to understand and clear, we can tabulate the data in the form of a table. this table is called a data table.

DATA TABLES REPRESENTATION.

After collecting data, the investigator must find ways to condense them in tabular form to study their salient features

* Serial order of alphabetic order
* Ascending order
* Descending order

The raw data, when put in ascending or descending order of magnitude, it is called an array. The raw data is called an array or array data when placed in ascending or descending order of magnitude. The collected data is arranged in rows and columns in a table called a data table.

For eg:

Let the marks obtained by \(30\) students of class in a class test, out of marks, according to their roll numbers, be.

\(39, 25, 5, 33, 19, 21, 12, 41, 12, 21, 19, 1, 10, 8, 12, 17, 19, 17, 17, 41, 40, 12, 41, 33\) \(19, 21, 33, 5, 1, 21\)

In the very first column of the table, we write all marks from lowest to highest. We now look at the first value on the given raw data and put a vertical line in the second column parallel to it. We see the second value in the given raw data and

put a bar opposite it in the second column. This process is repeated till all observations in the given raw data are exhausted.

The bars drawn in the second column are known as tally marks, and to facilitate, we record tally marks in bunches of five, the fifth tally mark is drawn diagonally across the first four. We eventually count the number of tally marks corresponding to each observation and write in the third column.

MARKS	TALLY MARKS	NO. OF STUDENTS.
1	||	2
5	||	2
8	|	1
10	|	1

This way of presentation of data is known as frequency distribution. Marks are called variates, and the no. of students who have secured a particular number of marks is called the frequency of the variate. The number of times an observation occurs in the given data is called the frequency of the observation.

What is frequency distribution?

The frequency of a value is the number of times it occurs in a dataset. A frequency distribution is the pattern of frequencies of a variable. It's the no. of times each possible value of a variable occurs in a dataset.

TYPES OF FREQUENCY DISTRIBUTIONS.

There are four types of frequency distributions

i) UNGROUPED FREQUENCY DISTRIBUTIONS:
The no. of observations of each value of a variable.

You can use this type of frequency distribution for categorical variables.

ii) GROUPED FREQUENCY DISTRIBUTIONS
The no. of observations of each class interval of a variable. Class intervals are ordered groupings of a variable's values.

You can use this type of frequency distribution for quantitative variables.

iii) RELATIVE FREQUENCY DISTRIBUTIONS:
The proportion of observations of each value or class interval of a variable.

You can use this type of frequency distribution for any type of variable when you're more interested in comparing frequencies than the actual no. of observations.

iv) COMULATIVE FREQUENCY DISTRIBUTIONS:
The sum of the frequencies less than or equal to each value or class interval of a variable.

You can use this type of frequency distribution for ordinal or quantitative variables when you want to understand how often observations fall below certain values.

How to make a grouped frequency table.

1. Divide the variable into class intervals.

Below is one method to divide a variable into class intervals. Different methods will give different answers, but there's no agreement on the best method to calculate class intervals.

Calculate the range:

Subtract the lowest value in the dataset from the highest.

Decide the class interval with:

There are no firm rules on how to choose the width, but the following formula is a rule of thumb:

$$\text{width} = \frac{\text{range}}{\sqrt{\text{sample size}}}$$

You can round this value to a whole number or a number that's convenient to add (such as a multiple of 100).

Calculate the class intervals:

Each interval is defined by a lower limit and upper limit. Observations in a class interval are greater than or equal to the lower limit and less than the upper limit.

$$\text{lower limit} \leq x < \text{upper limit}$$

The lower limit of the first interval is the lowest value in the dataset. Add the class interval width to find the upper limit of the first interval and the lower limit of

the second variable. Keep adding the interval width to calculate more class intervals until you exceed the highest value.

2. Create a table:

Create a table with two columns and as many rows as there are class intervals. Label the first column using the variable name and the label the second column "Frequency". Enter the class intervals in the first column.

3. Create the frequencies.

The frequencies are the number of observations in each class interval. You can count by tallying if you find it helpful. Enter the frequencies in the second column.

Example:.

Grouped Frequency Distribution.

A sociologist conducted a survey of 20 adults. She wants to report the frequency distribution of the ages of the survey respondents. The respondents were the following ages in years:.

52, 34, 32, 29, 63, 40, 46, 54, 36, 36, 24, 19, 45, 20, 28, 29, 38, 33, 49, 37.

range = highest - lowest

range = 63 - 19

range = 44

$$\text{width} = \frac{\text{range}}{\sqrt{\text{sample size}}}$$

$$\text{width} = \frac{44}{\sqrt{20}} \Rightarrow 9.84$$

Round the class interval width to 10
The class interwals are

$19 \leq a < 29$, $29 \leq a < 39$, $39 \leq a < 49$, $49 \leq a < 59$ and $59 \leq a$ 69.

Grouped frequency table of the ages of survey participants.

Age, a(years)	Frequency
$19 \leq a < 29$	4
$29 \leq a < 39$	9
$39 \leq a < 49$	3
$49 \leq a < 59$	3
$59 \leq a < 69$	1

What is Tabular Presentation of Data?

It is a table that helps to represent even a large amount of data in an engaging, easy to read, and coordinated manner. The data is arranged in rows and columns. This is one of the most popularly used forms of presentation of data as data tables are simple to prepare and read.

The most significant benefit of tabulation is that it coordinates data for additional statistical treatment and decision making. The analysis used in tabulation is of four types. They are

* Qualitative
* Quantitative
* Temporal
* Spatial.

1. Qualitative classification:
When the classification is done according to traits such as physical status, nationality, social status etc, it is known as qualitative classification

2. Quantitative classification:
In this, the data is classified on the basis of features that are quantitative in nature. In other words, these features can be estimated quantitatively.

3. Temporal classification:
In this classification, time becomes the categorising variable and data are classified according to time. Time, maybe in years, months, weeks, days, hours, etc.

4. Spatial classification
When the categorisation is done on the basis of location, it is known as spatial classification. The place may be a country, state, district, block, village/town etc.

Basics of Tabular Presentation.

Concept of Tabulation:
* Tabulation i.e., tabular presentation of data is a method of presentation of da
* It is a systematic and logical arrangement of data in the form of rows and columns with respect to the characteristics of data.
* It is an orderly arrangement which is compact and self-explanatory

* Its objective is to present the data in a simple form, economies (save) space, facilitate comparison, facilitate statistical analysis, reduce the chances of errors.

Objectives of Tabulation

* To simplify the complex data.
* To bring out essential features of the data.
* To facilitate comparison.
* To facilitate statistical analysis.
* Saving of Space.

Limitations of a Table.

i) Lacks description.
   * The table represents only figures and not attributes.
   * It ignores the qualitative aspects of the facts.

ii) Incapable of presenting individual items.
   * It doesn't present individual items.
   * It presents aggregate data.

iii) Needs special knowledge.
   * The understanding of the table require special knowledge.
   * It cannot be easily used by layman.

Eg:
Compute the arithmetic mean of the first 6 odd natural numbers

Solution:
The first 6 odd natural numbers : 1, 3, 5, 7, 9, 11

$\bar{x} = (1+3+5+7+9+11)/6 = 36/6 = 6$

Thus, the arithmetic mean is 6.

Eg: Discrete grouped data

Find the arithmetic mean of the following distribution.

x	10	30	50	70	89
f	7	8	10	15	10

x_i	f_i	$x_i f_i$
10	7	70
30	8	240
50	10	500
70	15	1050
89	10	890

Total $\Sigma f_i = 50$ $\Sigma x_i f_i = 2750$

The Arithmetic mean formula

$\bar{x} = \Sigma x_i f_i / \Sigma f_i = 2750/50 = 55$

Arithmetic mean = 55

Calculate the A.M using the Shortcut method.

Wages (x)	100	110	120	140	130
No. of workers (f)	10	20	30	50	40

Here the value of A is 30.

Wages (d)	No. of workers (f)	d = f - A	fxd
100	10	-20	-200
110	20	-10	-200
120	30	0	0
130	40	10	400
140	50	20	1000
	$\Sigma f = 150$		$\Sigma f.d = 1000$

Shortcut Arithmetic method.

$$\bar{x} = A + \frac{\Sigma f.d}{\Sigma f}$$

$$\bar{x} = 30 + \frac{1000}{50}$$

$$\bar{x} = 36.667$$

STEP DEVIATION METHOD FORMULA.

Estimated or Direct mean $= \Sigma x_i f_i / \Sigma f_i$,

where f_i is the frequency and x_i is the midpoint of the class interval.

Assumed Mean $= A + \Sigma d_i f_i / \Sigma f_i$, where A is the assumed mean, f_i is the frequency and deviation $d_i = x_i - A$

Since this method is the extension of the assumed mean method, the formula is.

Step Deviation of Mean $= A + h[\Sigma u_i f_i / \Sigma f_i]$,

where,

A is the assumed mean

h is the class size