SNS COLLEGE OF TECHNOLOGY

Coimbatore-35.
An Autonomous Institution

COURSE NAME : DATA ANALYTICS

II YEAR/ IV SEMESTER

UNIT – II Getting Insights from Data

Topic: *Types of Scale*

Dr.K.Sangeetha
HoD
Department of Computer Science and Engineering

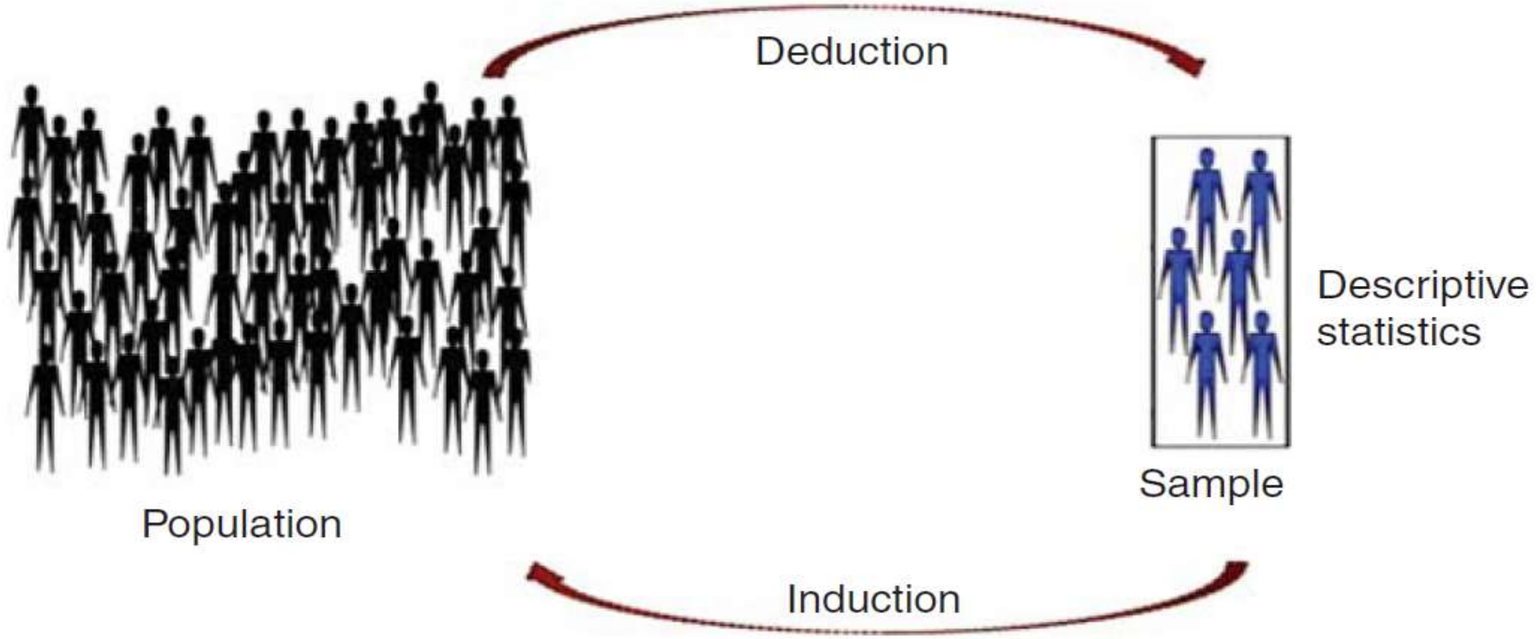
Brain Storming

1. What is Statistics ?

2. What is Descriptive Statistics ?

Descriptive Statistics ?

- **Descriptive statistics** is the branch of statistics that, sets out methods to describe data samples, through summarization and Visualization.



Descriptive Statistics (Cont..)

- Describing and visualizing data are usually categorized based on the **number of attributes** we are considering.
- Univariate analysis - analysis of single attributes
- Bivariate analysis - for pairs of attributes
- Multivariate analysis - for groups of more than two attributes

Descriptive Statistics (cont..)

Descriptive Statistics

- **Scale Types**
- Descriptive Univariate Analysis
  - Univariate Frequencies
  - Univariate Data Visualization
  - Univariate Statistics
  - Common Univariate Probability Distributions
- Descriptive Bivariate Analysis
  - Two Quantitative Attributes
  - Two Qualitative Attributes, at Least one of them Nominal
  - Two Ordinal Attributes

Descriptive Statistics (cont..)

Scale Types :

- Two large scale types: **qualitative** and **quantitative.**

- Qualitative scales categorize data in a **nominal or ordinal** way.

Friend	Max temp (°C)	Weight (kg)	Height (cm)	Gender	Company
Andrew	25	77	175	M	Good
Bernhard	31	110	195	M	Good
Carolina	15	70	172	F	Bad
Dennis	20	85	180	M	Good
Eve	10	65	168	F	Bad

The simplest measurement scale we can use to label variables is a **nominal scale**.

Nominal scale: A scale used to label variables that have no quantitative values.

Some examples of variables that can be measured on a nominal scale include:

Gender: Male, female

Eye color: Blue, green, brown

Hair color: Blonde, black, brown, grey, other

Blood type: O-, O+, A-, A+, B-, B+, AB-, AB+

Political Preference: Republican, Democrat, Independent

Place you live: City, suburbs, rural

Features:

•**They have no natural order.** For example, we can't arrange eye colors in order of worst to best or lowest to highest.

•**Categories are mutually exclusive.** For example, an individual can't have *both* blue and brown eyes. Similarly, an individual can't live *both* in the city and in a rural area.

•**The only number we can calculate for these variables are *counts*.** For example, we can count how many individuals have blonde hair, how many have black hair, how many have brown hair, etc.

•**The only measure of central tendency we can calculate for these variables is *the mode*.** The mode tells us which category had the most counts. For example, we could find which eye color occurred most frequently.

Ordinal

Ordinal scale: A scale used to label variables that have a **natural order**, but no **quantifiable difference** between values.

Satisfaction: Very unsatisfied, unsatisfied, neutral, satisfied, very satisfied
Socioeconomic status: Low income, medium income, high income
Workplace status: Entry Analyst, Analyst I, Analyst II, Lead Analyst
Degree of pain: Small amount of pain, medium amount of pain, high amount of pain

Ordinal” indicates “order”. It can be named, grouped and also ranked.

“How satisfied are you with our products?” •1- Totally Satisfied •2- Satisfied •3- Neutral •4- Dissatisfied •5- Totally Dissatisfied	“How happy are you with the customer service?” •1- Very Unhappy •2- Unhappy •3- Neutral •4- Unhappy •5- Very Unhappy

Features:

- **They have a natural order**. For example, “very satisfied” is better than “satisfied,” which is better than “neutral,” etc.
- **The difference between values can’t be evaluated**. For example, we can’t exactly say that the difference between “**very satisfied and “satisfied**” is the same as the difference between “satisfied” and “neutral.”

- The two measures of central tendency we can calculate for these variables are the mode and the median.

- The mode tells us which category had the most counts and the median tells us the “middle” value.

Collected by companies through surveys who are looking for feedback about their product or service. For example, a grocery store might survey 100 recent customers and ask them about their overall experience.

Question: How satisfied were you with your most recent visit to our store?
Possible Answers: Very unsatisfied, unsatisfied, neutral, satisfied, very satisfied.
Using this data, the grocery store can analyze the total number of responses for each category, identify which response was most common, and identify the median response.

Descriptive Statistics (cont..)

Scale Types :

- There are two types of scale for quantitative data:
  absolute (ratios) – absolute Zero
  relative (intervals) no absolute zero

Relative or Interval Scale

Interval scale: A scale used to label variables that have a natural order and a **quantifiable difference between values**, *but no "**true zero**" value*.

Temperature: Measured in Fahrenheit or Celsius

Credit Scores: Measured from 300 to 850

- **These variables have a natural order.**
- **We can measure the mean, median, mode, and standard deviation of these variables.**
- **These variables have an exact difference between values.** Recall that ordinal variables have no exact difference between variables – we don't know if the difference between "very satisfied" and "satisfied" is the same as the difference between "satisfied" and "neutral." For variables on an interval scale, though, we know that the difference between a credit score of 850 and 800 is the exact same as the difference between 800 and 750.
- **These variables have no "true zero" value.**
- For example, it's impossible to have a credit score of zero. And for temperatures, it's possible to have negative values (e.g. -10° F) which means there isn't a true zero value that values can't go below.

Ratio or Absolute

Ratio scale: A scale used to label variables that **have a natural order**, a quantifiable difference between values, and a “true zero” value.

Height: Can be measured in centimeters, inches, feet, etc. and cannot have a value below zero.
Weight: Can be measured in kilograms, pounds, etc. and cannot have a value below zero.
Length: Can be measured in centimeters, inches, feet, etc. and cannot have a value below zero.

- These variables have a **natural order**.
- We can calculate the mean, median, mode, standard deviation, and a variety of other descriptive statistics for these variables.
- These variables have an **exact difference** between values.
- These variables have a "**true zero**" value.

- For example, length, weight, and height all have a minimum value (zero) that can't be exceeded. It's not possible for ratio variables to take on negative values.

- For this reason, the *ratio* between values can be calculated. For example, someone who weighs 200 lbs. can be said to weigh *two times* as much as someone who weights 100 lbs.

- Likewise someone who is 6 feet tall is 1.5 times taller than someone who is 4 feet tall.

The following table provides a summary of the variables in each measurement scale:

Property	Nominal	Ordinal	Interval	Ratio
Has a natural “order”	YES	YES	YES	YES
Mode can be calculated	YES	YES	YES	YES
Median can be calculated		YES	YES	YES
Mean can be calculated			YES	YES
Exact difference between values			YES	YES
Has a “true zero” value				YES

Data set of our private list of contacts with weight and height

Friend	Maxtemp (∘C)	Weight(kg)	Height(cm)	Gender	Company
Andrew	25	77	175	M	Good
Bernhard	31	110	195	M	Good
Carolina	15	70	172	F	Bad
Dennis	20	85	180	M	Good
Eve	10	65	168	F	Bad
Fred	12	75	173	M	Good
Gwyneth	16	75	180	F	Bad
Hayden	26	63	165	F	Bad
Irene	15	55	158	F	Bad
James	21	66	163	M	Good
Kevin	30	95	190	M	Bad
Lea	13	72	172	F	Good
Marcus	8	83	185	F	Bad
Nigel	12	115	192	M	Good

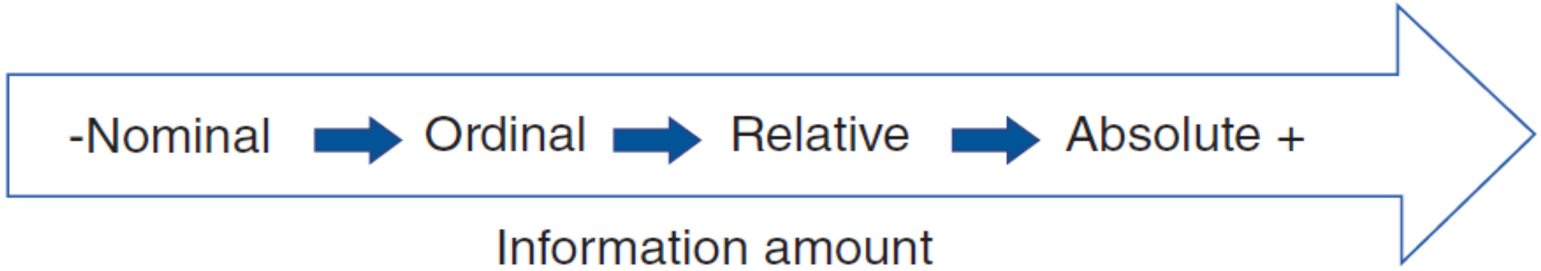
Example 1 The name of the contact is expressed on a nominal scale, while the information on how **good their company** is can be expressed on an ordinal scale because we can define an order of magnitude, ranging from good to bad. Good expresses a higher level of fellowship than bad. This notion of magnitude does not exist in the names.

Example 2.When the attribute “height” is **zero** it means there is no height. This is also true for the weight. But for the temperature, when we have 0∘C it does not mean there is no temperature. When we talk about weight, we can say that Bernhard weighs twice as much as Irene, but we cannot say that the maximum temperature last week in Dennis’ home town was twice that in Eve’s. This is why we usually use a change in temperature to characterize how the temperature varied in a given day instead of a ratio.

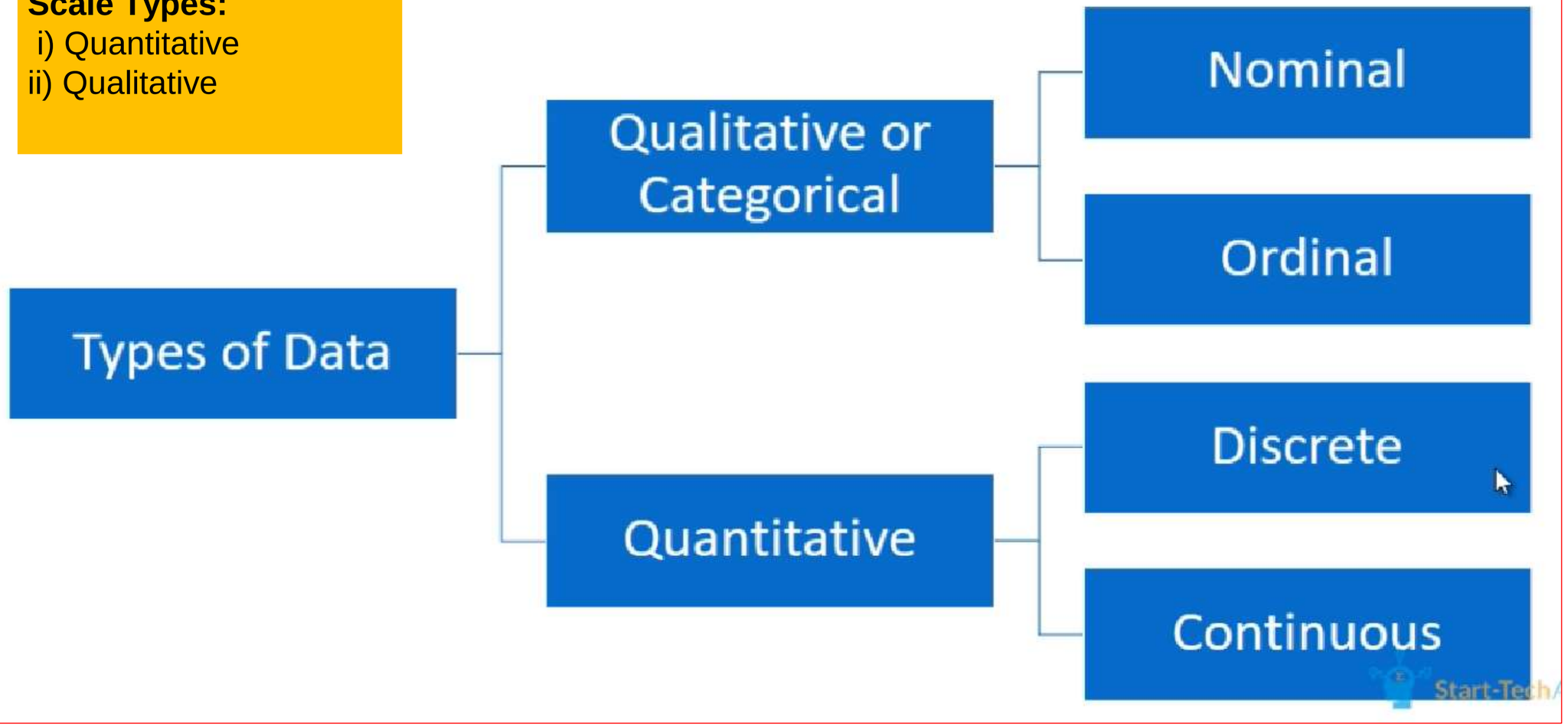
Descriptive Statistics (cont..)

Scale Types :

❖ Express the data according to the operations we can perform on their values equal (=) or different (≠) , larger than (>), larger than or equal to (≥), smaller than (<) or smaller than or equal to (≤), add to (+) or subtract from (−),



Scale Types:
i) Quantitative
ii) Qualitative



Descriptive Statistics (cont..)

Scale Types :

- **Example:** consider the attribute “weight” expressed in an absolute scale in kilograms. We can convert it to any other scale:
- ***Relative****:* weight converted to a relative scale by, for instance , subtracting a value of 10. The old zero becomes -10 and the new zero is the old 10.

- ***Ordinal****:* We can define, for instance, levels of fatness: “fat” (larger than 80kg, “normal” when the weight is larger than 65kg but less than or equal to 80kg) and “thin” (less than or equal to 65kg)

- ***Nominal****:* fat, normal and thin – into B, A and C, respectively.

References

TEXT BOOKS

1. Joao Moreira, Andre Carvalho, Tomás Horvath – “A General Introduction to Data Analytics” – Wiley -2018

REFERENCES

1 Dean J, —Big Data, Data Mining and Machine learning, Wiley publications, 2014.

2 Provost F and Fawcett T, —Data Science for Business, O‘Reilly Media Inc, 2013.

3 Janert PK, —Data Analysis with Open Source Tools, O‘Reilly Media Inc, 2011. .

4 Weiss SM, Indurkhya N and Zhang T, —Fundamentals of Predictive Text Mining, Springer-Verlag London Limited, 2010.

5. Runkler T A, - Data Analytics: Models and Algorithms for Intelligent data analysis,Springer, 2012