

Data

B. Srirangacharyulu

BS

Data Analytics

1

1

Data

- Facts and figures collected, analysed, and summarised for presentation and interpretation
- Elements – are entities on which data are collected
- Variable (or attribute) – a characteristic of the element
- Levels of attribute (or observations) – enables to distinguish one element from the other

BS

Data Analytics

2

2

Data – scales of measurement

- Nominal scale
- Ordinal scale
- Interval scale
- Ratio scale

BS

Data Analytics

3

3

Data – scales of measurement

Nominal scale

- Labels/names – to identify an attribute of an element
- E.g. names, ID numbers, Pin codes, Telephone numbers
- Pin codes – not purely nominal because every digit has a meaning; still it is nominal because we can only distinguish
- One cannot get more information than to identify

BS

Data Analytics

4

4

Data – scales of measurement

Ordinal scale

- We get extra information here
- E.g. grades, ratings, ranks
- There is order – ascending or descending
- We cannot do any mathematical operations
- Though we know the order, we don't know the numerical distance that separates two successive entities

BS

Data Analytics

5

5

Data – scales of measurement

Interval scale

- Ordinal + interval is fixed (equidistant)
- Equal intervals and an arbitrary 'zero' (philosophically, 'zero' means nothing)
- Degrees centigrade:
 - $90^{\circ}C - 70^{\circ}C = 70^{\circ}C - 50^{\circ}C$
- $9 A.M. \neq 2 \times 4:30 A.M.$
- Shoe size '0' doesn't mean "No Shoe"

BS

Data Analytics

6

6

Data – scales of measurement

Ratio scale

- Interval + meaningful ratio
- Cost of pen 'A' is 100, and cost of pen 'B' is 50; a ratio of "2" implies cost of pen 'A' is twice that of pen 'B'
- E.g. height, weight, distance, length
- A '0' value exists, indicating that nothing exists

BS

Data Analytics

7

7

Scales of measurement

- Temperature in degrees Celsius
- Hours to complete a task
- Student ranks based on test scores
- Car colours
- Satisfaction level on a scale of 1 to 5
- Length of table
- T-shirt sizes
- 2000 years of elapsed time
- 2000 A.D.

BS

Data Analytics

8

8

Categorical and Quantitative Data

Categorical	Quantitative
<ul style="list-style-type: none">• Obtained by Nominal / Ordinal scales of measurement	<ul style="list-style-type: none">• Obtained by Interval / Ratio scales of measurement
<ul style="list-style-type: none">• Statistical analysis is limited	<ul style="list-style-type: none">• Lot of statistical analysis can be done
<ul style="list-style-type: none">• No arithmetic operations	<ul style="list-style-type: none">• Arithmetic operations can be done
<ul style="list-style-type: none">• Can count the no. of observations	<ul style="list-style-type: none">• Mean, variance

Data – other classifications

- Cross-Sectional and Time Series Data
- Continuous and Discrete