Basic Statistics

Dr. Sudipta Das

Department of Computer Science,
Ramakrishna Mission Vivekananda Educational & Research Institute

Outline I

Chapter 2: Different Data Types in Statistics

Data Types: Qualitative/Quantitative I

Table: Information on 15 new-born babies, collected from a hospital

Sl No	Sex	Eye Color	Mother's Education	Weight (in k.g.)	Height (in c.m.)	No. of Siblings
1	female	blue	UG	3.85	52.51	1
2	female	black	HS	3.01	51.76	0
3	male	blue	UG	2.54	49.19	1
4	female	grey	PG	2.68	49.98	0
5	male	blue	HS	3.81	55.97	2
6	male	grey	HS	3.35	54.63	1
7	female	blue	UG	2.80	46.43	1
8	female	grey	PG	2.75	50.68	0
9	male	grey	UG	3.86	54.06	0
10	male	black	UG	3.76	54.04	1
11	female	blue	PG	2.59	47.01	2
12	male	black	HS	2.68	52.14	1
13	male	grey	PG	3.62	51.77	0
14	female	black	UG	3.12	50.23	1
15	male	black	PG	2.88	47.91	1

Data Types: Qualitative/Quantitative II

- Attribute/Categorical (Non-numerical to start with)
    - Nominal:-
        - Name or label without any order
        - Sex, Religion
        - Matching
        - *character* in R
    - Ordinal:-
        - Name or label with an order
        - Level of satisfaction, Degree of pain
        - Matching and sorting
        - *factor* in R

- Variable (Numerical)
    - Discrete:-
        - Take distinct and isolated values
        - No. of children, No. of car
        - All arithmetic operations
        - *integer* in R
    - Continuous:-
        - Take any value in a range
        - Height or weight of a person
        - All arithmetic operations
        - *numeric* in R

Data Types: Originality Aspects

- Primary Data
    - Investigators/Enumerators go directly to the field of inquiry and through observation, interview or direct measurement collect the relevant data.
    - First-hand data
- Secondary Data
    - Data collected from secondary source (research organization, journal etc.), which collected data for their own purpose and later stored the data for future use
    - Second-hand data

Data Types: Identity Aspects

- Non-frequency Data
  - Identity of the individual is important and is kept in mind throughout the study
  - Example: Annual production of steel in India; only the production values without reference to the years have little significance
- Frequency Data
  - Identity of the individual is not important, and a particular value is important
  - Example: a particular individual height figure is important without reference to the individual concerned