**Descriptive Statistics**

1. Different Types of Data
2. Qualitative Data – Attributes which describes the object under consideration using finite set of values. E.g. colour of shirt, pattern of shirt, size of shirt.

I ) Nominal data – No natural ordering in qualitative Data. (colour, pattern)

II ) Ordinal data – Natural ordering in qualitative Data. (size, rating)

A screenshot of a computer

Description automatically generated with medium confidence

1. Quantitative Data – Attributes which is used to count or measure using numbers and infinite set of values. e.g. price of shirt, discount on price.

I ) Discrete data – Only Integers

II ) Continuous data – All rational Number

A person in a white shirt

Description automatically generated with medium confidence

*“****The Type of Statistical Analysis depends on type of variable/data”.***

Exercise

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Domain | Ordinal | Nominal | Discrete | Continuous |
| Banking | Employee Review | Gender | No. of employees | Amount of Money |
| Insurance | Employee Review | Gender | No. of Insurance sold | Amount of Money |
| Education | Rank of Toppers | Subjects | No. of courses taught | Fees increment % |
| Healthcare | Rank | Race | No. of patients | Medicine’s weight |
| Sports | Rank of winners | Nationality | No. of medals | Size of field |

1. How to describe Qualitative Data
2. Using Frequency – No. of times an Attribute appears in a sample. **Absolute Frequency** **Plots** are used to represent frequency (Sorted Plot is better for analysis).

A screenshot of a computer

Description automatically generated with medium confidence

1. Using Relative Frequency – Easier to interpret (e.g. percentages)
2. Grouped Frequency Bar Charts – To compare Different **Sets** of data.

A screenshot of a computer

Description automatically generated with medium confidence

1. Grouped Relative Frequency Bar Charts – Relative way of interpreting.

Use Case in ML: - To analyse Errors ML algorithm, designing features for ML system (e.g., review classification for some product)

Exercise

|  |  |  |
| --- | --- | --- |
| Domain | Long-Tailed Distribution | Uniform Distribution |
| Banking |  |  |
| Insurance |  |  |
| Healthcare |  |  |
| Education |  |  |
| Sports |  |  |
| Business |  |  |

1. How to Describe Quantitative Data
2. Histograms – Ordered X and Y axis
3. Histograms with bins (Ideal for **Continuous Data)**– bin size depends on range of data. Ideal Bin size should neither hide nor reveal too many details.
4. Class-Boundary – Left-end inclusion convention

Graphical user interface, text, application, chat or text message

Description automatically generated