Statistics – Statistics is the science of collecting, organizing and analyzing data. Data is the piece of information that we can measure.

Statistics types –

1. Descriptive statistics – It describe, organize or summarize the data. Like Average, Sum etc.
2. Inference statistics – It is technique where we measure the data to form conclusions. Such as comparison etc. eg. compare average marks of this class room similar to maths class. This usually take sample against whole population to do comparison etc.

Population and Sample –

Population is denoted by (N) and sampling by (n)

Sampling Techniques –

1. Simple Random Sampling – Take random samples. Every member of population has an equal chance of being selected for sample(n).
2. Stratified Sampling- It is technique where the population split into non-overlapping groups. Here the population is divided into sub-categories(strat) then from each category the sampling will happen, may be with different probability sampling techniques. Eg. Gender, Race etc.
3. Systematic Sampling – The sampling based on predefined systematic technique. Eg. Every 10th sample should select.
4. Convenience/Volunteer Sampling – Only domain specific expert population will be selected. Eg. Survey for students. Then only students from population will be considered.

Variable Types –

1. Quantitative – These are numbers.
2. Discrete Variables- Whole numbers like total sales, total populate etc.
3. Continuous Variables – Continues numbers. Eg. River Length, Song Length, Blood Pressure.
4. Qualitative/Categorical Variables – These are categories or groups. Eg. Gender, Age Group, Marital Status, PIN Code.

Variable Measurement Scales- 4 types-

1. **Normal/Nominal** – Categorical data. This is part of Quantitative variables. These are just categories with no ordering, ranking or directions. Any computation such as Avg, Min, Max etc. will not work on this variable. Eg. Zip code, Race, Gender etc.
2. **Ordinal** – Order of data values matters but not the difference between values. Eg. Rank in class, Income level- Low income, high etc. Satisfaction – Low, high, extreme satisfied etc. This is part of Quantitative variables.
3. **Interval**- Order of data and difference between values both matters. This cannot have natural zero. Eg. Temperature, credit scores (300-850). This is part of Categorical variables.
4. **Ratio** – A variable has Interval properties that value and order both matters, but it has definition for 0 as well. This is part of Categorical/ Qualitative variables. Here the ration of two values has meaningful interpretation. For Eg. Weight of two persons can says ratio of weight.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Computation** | **Nominal** | **Ordinal** | **Interval** | **Ratio** |
| Frequency Distribution | Yes | Yes | Yes | Yes |
| Median & percentiles | No | Yes | Yes | Yes |
| Add/Subtract | No | No | Yes | Yes |
| Mean/Std | No | No | Yes | Yes |
| Ratio/Co-efficient of variation | No | No | No | Yes |

Computation Types-

Frequency Distribution – Occurrence of value. Eg. Bar Chart of Flowers.

Cumulative Distribution Frequency (CDF) – The frequency of previous category will get added to next category. Chart of Percentile or Populate over age where as age increases the person belongs to next age category as well.