



SNS COLLEGE OF TECHNOLOGY

Coimbatore-35.



An Autonomous Institution

COURSE NAME : 19CST203 - DATA ANALYTICS

II YEAR /IV SEMESTER

1.MULTIVARIATE FREQUENCIES

The multivariate frequency values can be computed independently for each attribute. We can represent the frequency values for each attribute by a matrix, in which the number of rows is the number of values assumed by the attribute and the columns are frequency values, just as in Table 2.3 for the attribute “height”.

TEMPERATURE(IN CELSIUS)	ICE CREAM SALES
20	2000
25	2500
35	5000
43	7800

depending on the attribute values being discrete or continuous, the attribute values are defined by, respectively, a probability mass function or a probability density function. Thus, different procedures are used for qualitative and quantitative value scales. Nevertheless, for each attribute, the following frequency measures can be taken:

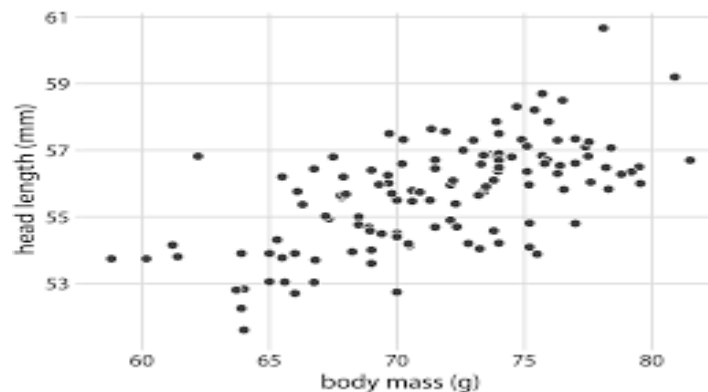
- absolute frequency
- relative frequency
- absolute cumulative frequency
- relative cumulative frequency

2. MULTIVARIATE DATA VISUALIZATION

We have already seen, for univariate and bivariate analysis, it is easier to understand data and experimental results when they are illustrated using visualization techniques.

The good news is that some of the previous plots can be extended to represent a small number of extra attributes. Additionally, new visualization approaches and techniques are continuously being created to deal with new types of data, new approaches to results interpretation and new data analysis tasks. Depending on the number of attributes, and the need to represent spatial and/or temporal aspects of the data, different plots can be used. This section explores how multivariate data can be visually represented in different ways and the main benefits of each of these alternatives.

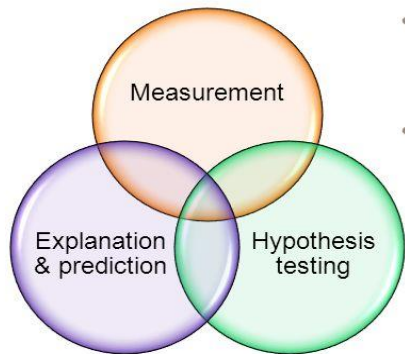
When the multivariate data has three attributes, or one can only analyze three attributes from a multivariate data set, the data can still be visualized in a bivariate plot, associating the values of the third attribute with how each data object is represented in the plot. If the third attribute is quantitative, the value can be represented by the size of the object representation in the plot.



3. MULTIVARIATE STATISTICS

At first sight, the extraction of statistical measures from more than two attributes can seem complicated. However multivariate statistics are just a simple extension of the univariate statistics seen in the previous chapter. As we will see in the next sections, some of the statistical measures previously described for univariate and bivariate analysis, such as the mean and standard deviation, can easily be extended to multivariate analysis.

Why Multivariate Statistics?



- Difficulty of addressing complicated research questions with univariate tools
- Several drivers for increasing popularity, e.g.
 - Availability of nicely packaged software
 - Greater complexity of contemporary research
 - Large amounts of data
 - Emergence of data mining perspective (finding unforeseen patterns and associations)