



# **SNS COLLEGE OF ENGINEERING**

Kurumbapalayam (Po), Coimbatore – 641 107

**An Autonomous Institution**

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade  
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

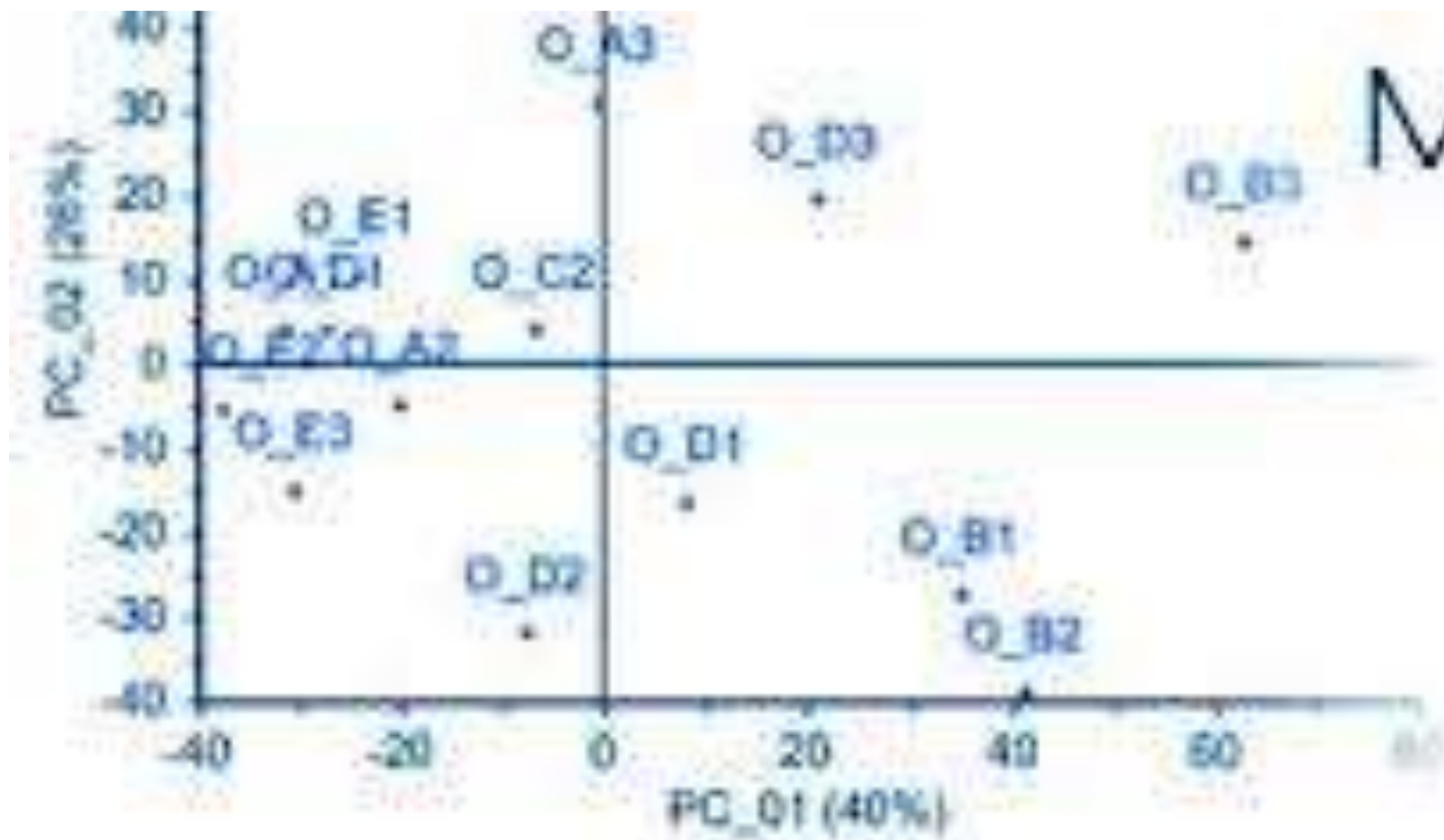
## **DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY**

**COURSE NAME :19CS407 DATA ANALYTICS WITH R**  
**II YEAR /IV SEMESTER**

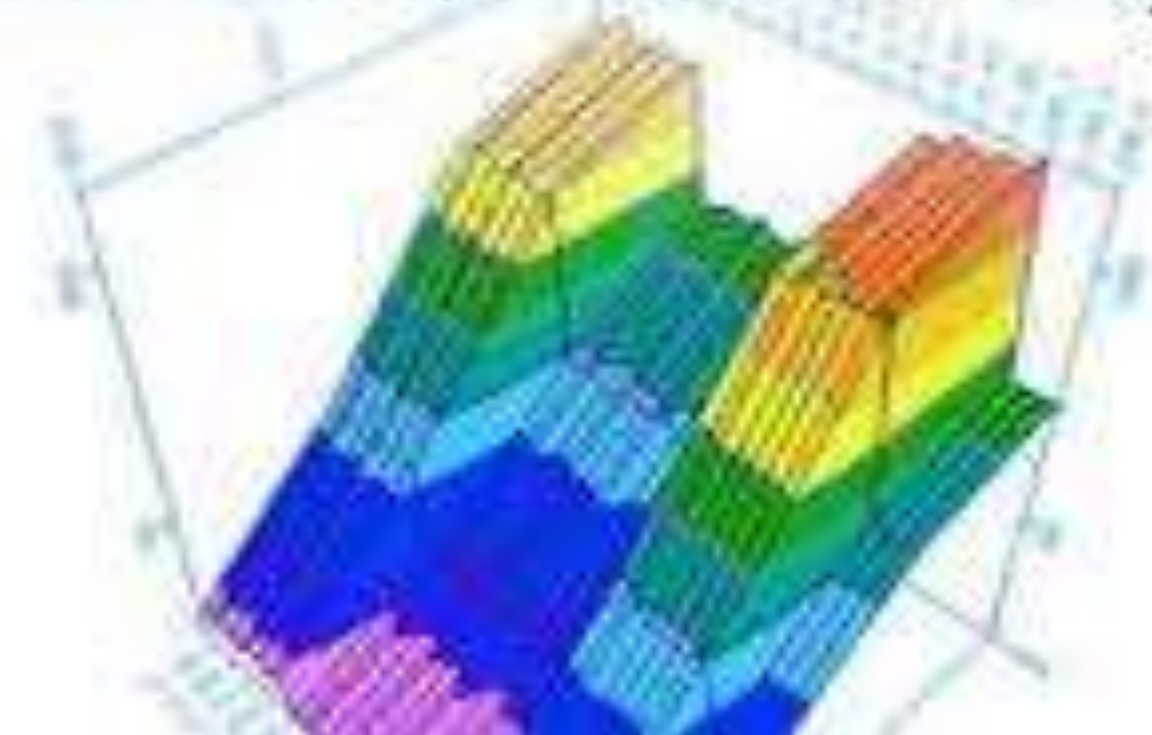
**Unit 2- GETTING INSIGHTS FROM DATA, DATA QUALITY  
AND PREPROCESSING**

**Topic : Multivariate Analysis - Multivariate Frequencies**





# Multivariate Data Analysis





# Descriptive Multi Variate Analysis



- ✓ Multivariate analysis (MVA) is a Statistical procedure for analysis of data involving more than one type of measurement or observation.
- ✓ It may also mean solving problems where more than one dependent variable is analyzed simultaneously with other variables



# Classification Chart of Multivariate Techniques

- ✓ a) Are the variables divided into independent and dependent classification?
- ✓ b) If Yes, how many variables are treated as dependents in a single analysis?
- ✓ c) How are the variables, both dependent and independent measured?
- ✓ **Dependence technique: Dependence Techniques are types of multivariate analysis techniques that are used when one or more of the variables can be identified as dependent variables and the remaining variables can be identified as independent.**



# Dependence Technique

Multiple Regression

Conjoint analysis

Multiple Discriminant Analysis

Linear Probability Models

Multivariate Analysis of Variance and Covariance

Canonical Correlation Analysis

Structural Equation Modeling



# Descriptive Multi Variate Analysis



- ✓ In real life, as we saw in our contacts data set, the number of attributes is usually more than two. It can be tens, hundreds or even more.
- ✓ Actually, in biology, for example, data sets with several hundreds, or even thousands, of attributes are very common
- ✓ When the analysis of a data set explores more than two attributes, it is termed “multivariate analysis”. As in univariate and bivariate analysis, frequency tables, statistical measures and plots can be used or adapted for multivariate analysis



# Descriptive Multi Variate Analysis



- ✓ Naturally, the larger the number of attributes, the more difficult the analysis becomes. It must be observed that all methods used for more than two attributes can also be used for two or one attributes
- ✓ In order to illustrate the methods described in this chapter for multivariate analysis, let us add a new attribute to the data set of the excerpt of our private list of contacts



# Descriptive Multi Variate Analysis



- ✓ seven columns, our multivariate analysis can use up to seven attributes.
- ✓ The columns (attributes) are the **name of the contact**, the **maximum temperature registered in the previous month** in their home town, their **weight, height**, how long we have known them (years), and their **gender**, finishing with our rating of how good their company is





# Descriptive Multi Variate Analysis



Table 3.1 Data set of our private list of contacts with weight and height.

Contact	Maxtemp	Weight	Height	Years	Gender	Company
Andrew	25	77	175	10	M	Good
Bernhard	31	110	195	12	M	Good
Carolina	15	70	172	2	F	Bad
Dennis	20	85	180	16	M	Good
Eve	10	65	168	0	F	Bad
Fred	12	75	173	6	M	Good
Gwyneth	16	75	180	3	F	Bad
Hayden	26	63	165	2	F	Bad
Irene	15	55	158	5	F	Bad
James	21	66	163	14	M	Good
Kevin	30	95	190	1	M	Bad
Lea	13	72	172	11	F	Good
Marcus	8	83	185	3	F	Bad
Nigel	12	115	192	15	M	Good



# 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 for the attribute “height”.



# Advantages



- ✓ The main advantage of multivariate analysis is that since it considers more than one factor of independent variables that influence the variability of dependent variables, the conclusion drawn is more accurate.
- ✓ The conclusions are more realistic and nearer to the real-life situation



# Disadvantages



- ✓ The main disadvantage of MVA includes that it requires rather complex computations to arrive at a satisfactory conclusion.
- ✓ Many observations for a large number of variables need to be collected and tabulated; it is a rather time-consuming process.



# Multivariate Frequencies

- ✓ 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.



# Multivariate Frequencies



- ✓ absolute frequency
- ✓ relative frequency
- ✓ absolute cumulative frequency
- ✓ relative cumulative frequency.



# absolute frequency

- ✓ Absolute frequency is a statistical term describing the number of times a particular piece of data or a particular value appears during a trial or set of trials. Essentially, absolute frequency is a simple count of the number of times a value is observed.
- ✓ if you ask 10 friends if blue is their favorite color and three say yes and seven say no, you have enough information to determine absolute frequency: the absolute frequency of "yes" is equal to three and that of "no" is equal to seven. The number of values tracked often increases with sample size or trial scope.
- ✓ if you ask 100 people if their favorite color is blue, the absolute frequency will likely increase. However, there is no additional complexity in the tracking of how many times a given value occurs.



# Relative frequency



This is a frequency table to see how many students have got marks between given intervals in Maths.

Marks	Frequency	Relative Frequency
45 – 50	3	$3 / 40 \times 100 = 0.075$
50 – 55	1	$1 / 40 \times 100 = 0.025$
55 – 60	1	$1 / 40 \times 100 = 0.075$
60 -65	6	$6 / 40 \times 100 = 0.15$
65 – 70	8	$8 / 40 \times 100 = 0.2$
70 – 80	3	$3 / 40 \times 100 = 0.275$
80 -90	11	$11 / 40 \times 100 = 0.075$
90 – 100	7	$1 / 40 \times 100 = 0.025$





## absolute cumulative frequency

- ✓ An absolute cumulative frequency is the number of scores which fall at or below a given score value. It is computed by adding up the number of scores which are equal to or less than a given score value.

Absolute Cumulative Frequency of Shoe Size

Shoe Size	Absolute Frequency	Absolute Cumulative Freq
6.0	1	1
6.5	2	3
7.0	3	6
7.5	2	8
8.0	0	8
8.5	2	10
9.0	1	11
9.5	2	13
10.0	0	13
10.5	2	15
	15	



## relative cumulative frequency

- ✓ A cumulative relative frequency distribution is a tabular summary of a set of data showing the relative frequency of items less than or equal to the upper class limit of each class. Relative frequency is the fraction or proportion of the total number of items.

Count	Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	5	5	$\frac{5}{25} = 0.2$ 20%
3	5	10	$\frac{10}{25} = 0.4$ 40%
5	6	16	$\frac{16}{25} = 0.64$ 64%
7	1	17	$\frac{17}{25} = 0.68$ 68%
11	8	25	$\frac{25}{25} = 1.0$ 100%



# Assessment 1





# References



1. João Moreira, Andre Carvalho, Tomás Horvath – “A General Introduction to Data Analytics” – Wiley -2018
2. <https://www.mygreatlearning.com/blog/introduction-to-multivariate-analysis/>

**Thank You**