|  |
| --- |
| **List the components of multivariate analysis.**   * **Trend**: * **Seasonality**: * **Cycle** * **Irregularity** |
| **Outline the different causal relationships between variables.** |
| State the unique aim for each of the following  analytical techniques  **Examine the unique aim for Principal Component Analysis (PCA).**  It is a popular technique for analyzing large datasets containing a high number of dimensions/features per observation, increasing the interpretability of data while preserving the maximum amount of information, and enabling the visualization of multidimensional data. |
| **What is the use of time series data analysis in organizations?**   * Time series analysis helps organizations understand the underlying causes of trends or systemic patterns over time. * Using data visualizations, business users can see seasonal trends and dig deeper into why these trends occur. * With modern analytics platforms, these visualizations can go far beyond line graphs. |
| **Define Time Series Analysis.**   * Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. * In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. |
| **What do you meant by trend in Time Series Analysis?**  **Trend**is a pattern in data that shows the movement of a series to relatively higher or lower values over a long period of time. In other words, a trend is observed when there is an increasing or decreasing slope in the time series. |
| **Name two methods of measuring trend.** |
| **List the uses of analysis of Time series.**   * Time series is used to predict future values based on previously observed values. * Time series analysis is used to identify the fluctuation in economics and business. * It helps in the evaluation of current achievements. * Time series is used in pattern recognition, signal processing, weather forecasting and earthquake prediction. |
| **What do you mean by moving averages?**  A moving average is a series of averages, calculated from historic data. Moving averages can be calculated for any number of time periods, for example a three-month moving average, a seven-day moving average, or a four-quarter moving average. The basic calculations are the same. |
| **Name the methods of measurement of trend in time series analysis.**  Following are the methods by which we can measure the trend.  (i) Freehand or Graphic Method.  (ii) Method of Semi-Averages.  (iii) Method of Moving Averages.  (iv) Method of Least Squares. |

**1.**

* + - 1. **Explain multivariate analysis. Illustrate the causal path models for three variables.**

Multivariate Analysis is defined as a process involving multiple dependent variables resulting in one outcome. This explains that the majority of the problems in the real world are Multivariate. For example, we cannot predict the weather of any year based on the season. There are multiple factors like pollution, humidity, precipitation, etc. Here, we will introduce you to multivariate analysis, its history, and its application in different fields.

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.

**Advantages and Disadvantages of Multivariate Analysis**

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.

Classification Chart of Multivariate Techniques

Selection of the appropriate multivariate technique depends upon-

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?

Multivariate analysis technique can be classified into two broad categories viz., This classification depends upon the question: are the involved variables dependent on each other or not?

If the answer is yes: We have Dependence methods.  
If the answer is no: We have Interdependence methods.

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.

The Objective of multivariate analysis

(1) Data reduction or structural simplification: This helps data to get simplified as possible without sacrificing valuable information. This will make interpretation easier.

(2) Sorting and grouping: When we have multiple variables, Groups of “similar” objects or variables are created, based upon measured characteristics.

(3) Investigation of dependence among variables: The nature of the relationships among variables is of interest. Are all the variables mutually independent or are one or more variables dependent on the others?

(4) Prediction Relationships between variables: must be determined for the purpose of predicting the values of one or more variables based on observations on the other variables.

(5) Hypothesis construction and testing. Specific statistical hypotheses, formulated in terms of the parameters of multivariate populations, are tested. This may be done to validate assumptions or to reinforce prior convictions.

1. **Summarize the levels of the techniques in multivariate analysis.**

**Pairwise plots**

Pairwise plots are a great way to look at multi-dimensional data, and at the same time maintain the simplicity of a two-dimensional plot. As shown in the figure below, it allows the analysts to view all combinations of the variables, each in a two-dimensional plot. In this way, they can visualize all the relations and interactions among the variables on one single screen.

**Spider Plots**

While there are various ways of visualizing multi-dimensional data, spider plots are one of the easiest ways to decipher the meaning of data. From the figure below, we can see how easily we can compare three mobile phones based on attributes such as their speed, screen, camera, memory and apps.

**Correlation Analysis**

Often, data sets contain variables that are either related to each other or derived from each other. It is important to understand these relations that exist in the data. In statistical terms, correlation can be defined as the degree to which a pair of variables are linearly related. In some cases, it is easy for the analyst to understand that the variables are related, but in most cases, it isn’t. Thus, performing a correlation analysis is very critical while examining any data. Furthermore, feeding data which has variables correlated to one another is not a good statistical practice, since we are providing multiple weightage to the same type of data. To prevent such issues, correlation analysis is a must.

**Cluster Analysis**

In many business scenarios, the data belongs to different types of entities; and fitting all of them into a single model might not be the best thing to do. For example, in a bank dataset, the customers might belong to multiple income groups which leads to different spending behaviors. If we use the data having all these customers into a single model, we would be comparing apples to oranges. In that regard, clustering provides analysts a good way to segment their data and therefore avoid this problem. Clustering also allows us to visually understand and therefore compare the different attributes of the segments formed.

K-means clustering is a well-renowned approach used by a lot of data analysts and scientists. This separates the data points into clusters such that the inter-cluster distances are maximized. What this means is that each point in a particular cluster is similar to every other point in that cluster; and, points in a particular cluster are very different from every point in any other cluster. Other popular approaches for clustering include the hierarchical clustering algorithm, the DBSCAN algorithm, Partitioning Around Medoids (PAM) algorithm, etc.

1. **Calculate 3 years moving averages for the given data.**

[**https://www.youtube.com/watch?v=oDn\_j40inMA**](https://www.youtube.com/watch?v=oDn_j40inMA)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 3- Yearly moving Average in Time Series Analysis | | | | |
|  |  |  |  |  |  |
|  | **Years** | **values** | **3-yearly moving total** | **3-yearly moving Average** |  |
|  | 2011 | 3 |  |  |  |
|  | 2012 | 4 | 15 | 5 |  |
|  | 2013 | 8 | 18 | 6 |  |
|  | 2014 | 6 | 21 | 7 |  |
|  | 2015 | 7 | 24 | 8 |  |
|  | 2016 | 11 | 27 | 9 |  |
|  | 2017 | 9 | 30 | 10 |  |
|  | 2018 | 10 | 33 | 11 |  |
|  | 2019 | 14 | 36 | 12 |  |
|  | 2020 | 12 |  |  |  |
|  |  |  |  |  |  |

1. hort notes on
2. **Secular Trend or Simple trend or Long term movement:**Secular trend refers to the general tendency of data to increase or decrease or stagnate over a long period of time. Time series relating to Economic, Business, and Commerce may show an upward or increasing tendency. Whereas, the time series relating to death rates, birth rates, share prices, etc. may show a downward or decreasing tendency.
3. **Seasonal variations:**Seasonal variations refer to the changes that take place due to the rhythmic forces which operate in a regular and periodic manner. These forces usually have the same or most similar pattern year after year. When we record data weekly, monthly or quarterly, we can see and calculate seasonal variations. Thus, when a time series consists of data only based on annual figures, there will be seen no seasonal variations. These variations may be due to seasons, weather conditions, habits, customs or traditions. For example, in summers the sale of ice-cream increases and at the time of Diwali the sale of diyas, crackers, etc. go up.
4. **Cyclical variations:**Cyclical variations are due to the ups and downs recurring after a period from time to time. These are due to the business cycle and every organization has to phase all the four phases of a business cycle some time or the other. Prosperity or boom, recession, depression, and recovery are the four phases of a business cycle.
5. **Random or irregular variations:**Random variations are fluctuations which are a result of unforeseen and unpredictable forces. These forces operate in an absolutely random or erratic manner and do not have any definite pattern. Thus, these variations may be due to floods, famines, earthquakes, strikes, etc.