RURAL DATA DRIVEN INDIA
Data Analytics Business Analytics
Agro Analytics Rural Analytics

Rural Management

Business Analytics I

First Edition

Rural Management - Business Analytics I

सत्यमेव जयते

MoE | Government of India
Ministry of Education

First Edition: 2020

ISBN: 978-93-89431-24-7

Price: ₹ 750/-

Department of Higher Education

Ministry of Human Resource Development, Government of India

5-10-174, Shakkar Bhavan, Ground Floor, Fateh Maidan Road, Hyderabad - 500 004

Telangana State. Tel: 040-23422112, 23212120, Fax: 040-23212114

E-mail : editor@mgncre.org Website : www.mgncre.org

Published by: Mahatma Gandhi National Council of Rural Education (MGNCRE), Hyderabad

About the Book

Analytics has been defined as the extensive use of data, statistical and quantitative analysis, explanatory and predictive models and fact-based management to drive decisions and actions. Analytics is more than just analytical methodologies or techniques used in logical analysis. It is a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving. Analytics includes a range of activities, including business intelligence, which is comprised of standard, quantitative methods, including statistical analysis, forecasting/ extrapolation, predictive modeling (such as data mining), optimization and simulation.

Analytics refers to processes, technologies, and techniques that convert data into information and knowledge that controls business decisions. Business Analytics is a study of data through statistical and operations analysis, the formation of predictive models, applications used for optimizing techniques. Data Analytics is basically experimentation. Business Analytics is the application of data analytics for making decisions in the organization.

This text book aims at giving clear picture about the Analytics, Data Analytics and Business analytics. The course contains five units. In which three types of analysis, descriptive analysis, predictive analysis and prescriptive analysis has been covered. Chapter 1 covers basic concepts and role of analytics in Rural and Agriculture Development are covered. Chapter 2 explains about the Permutations and Combinations. Chapter 3 and 4 discusses how to visualize and summarise the data using line chart, pie chart, bar chart, area chart, scatter graph, Histogram, area graph and funnel chart. Chapter 5 discusses relationship between the variables and influence of one variable on other variables using Correlation and Regression Analysis and elaborates Qualitative and Judgmental time series forecasting models with stationary, linear, seasonal data. In this chapter problems are solved using Moving Average and Exponential Smoothing Methods.

All the cases discussed in the five chapters are based on the rural and agriculture data. This course examines how data analytics is used to transform rural and agriculture industries cases, using examples and case studies.

Finally, a recent McKinsey Report suggested that “the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions”.

I thank the contributors: Mr. Kaja Bantha Navas Raja Mohamed, Assistant Professor in the School of Mechanical Engineering at Sathyabama Institute of Science and Technology, Chennai and Ms. S. Sharmila Parveen, Assistant Professor in Department of Business Administration at Nazareth College of Arts and Science, Chennai to this book for their outstanding insights. Also, I would like to thank MGNCRE Team members for extending their extreme support in completing this text book.

Dr W G Prasanna Kumar
Chairman MGNCRE

About the Book

Analytics has been defined as the extensive use of data, statistical and quantitative analysis, explanatory and predictive models and fact-based management to drive decisions and actions. Analytics is more than just analytical methodologies or techniques used in logical analysis. It is a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving. Analytics includes a range of activities, including business intelligence, which is comprised of standard, quantitative methods, including statistical analysis, forecasting/ extrapolation, predictive modeling (such as data mining), optimization and simulation.

Analytics refers to processes, technologies, and techniques that convert data into information and knowledge that controls business decisions. Business Analytics is a study of data through statistical and operations analysis, the formation of predictive models, applications used for optimizing techniques. Data Analytics is basically experimentation. Business Analytics is the application of data analytics for making decisions in the organization.

This text book aims at giving clear picture about the Analytics, Data Analytics and Business analytics. The course contains five units. In which three types of analysis, descriptive analysis, predictive analysis and prescriptive analysis has been covered. Chapter 1 covers basic concepts and role of analytics in Rural and Agriculture Development are covered. Chapter 2 explains about the Permutations and Combinations. Chapter 3 and 4 discusses how to visualize and summarize the data using line chart, pie chart, bar chart, area chart, scatter graph, Histogram, area graph and funnel chart. Chapter 5 discusses relationship between the variables and influence of one variable on other variables using Correlation and Regression Analysis and elaborates Qualitative and Judgmental time series forecasting models with stationary, linear, seasonal data. In this chapter problems are solved using Moving Average and Exponential Smoothing Methods.

All the cases discussed in the five chapters are based on the rural and agriculture data. This course examines how data analytics is used to transform rural and agriculture industries cases, using examples and case studies.

Finally, a recent McKinsey Report suggested that “the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions”.

I thank the contributors: Mr. Kaja Bantha Navas Raja Mohamed, Assistant Professor in the School of Mechanical Engineering at Sathyabama Institute of Science and Technology, Chennai and Ms. S. Sharmila Parveen, Assistant Professor in Department of Business Administration at Nazareth College of Arts and Science, Chennai to this book for their outstanding insights. Also, I would like to thank MGNCRE Team members for extending their extreme support in completing this text book.

Dr W G Prasanna Kumar
Chairman MGNCRE

Contents

List of Tables

List of Figures

Chapter 1 Business Analytics - An Introduction

Introduction

A business analyst is someone who analyzes an organization or business domain (real or hypothetical) and documents its business, processes, or systems, assessing the business model or its integration with technology. However, organizational titles vary such as analyst, business analyst, business systems analyst or maybe systems analyst. Business analytics makes extensive use of analytical modeling and numerical analysis, including explanatory and predictive modeling and fact-based management to drive decision making. It is therefore closely related to management science. Analytics may be used as input for human decisions or may drive fully automated decisions.

Objectives

- To gain an understanding of how decision makers use business analytics to formulate and solve business problems and to support managerial decision making
- To understand statistics and its role in analytics in problem solving and decision making that is imperative for business organizations
- To become familiar with Arithmetic Progression and Geometric Progression

Structure

1.1 Data Analytics Vs Business Analytics

1.2 Applications of Analytics

1.3 Statistics in Business Management and Analytics

1.4 Arithmetic Progression

1.5 Geometric Progression

Data, they say, is the new oil. India could well be the New Arabia.

1.1 Data Analytics Vs Business Analytics

Analytics is the data driven decision making approach for a business problem. ***Analytics*** is a field which combines data, information technology, statistical analysis, quantitative methods and computer-based models into one. This all are combined to provide decision makers all the possible scenarios to make a well thought and researched decision. But Business Analytics has some specific focus. Analytics refers to processes, technologies, and techniques that convert data into information and knowledge that controls business decisions. ***Data Analytics*** is basically experimentation. ***Business Analytics*** is the application of data analytics for making decisions in the organization. Business analytics is a methodology or tool to make a sound commercial decision. Hence it impacts functioning of the whole organization/individual. Therefore, business analytics can help improve profitability of the business, increase market share and revenue and provide better return to a shareholder, managers and decision makers.

Definition of Business Analytics (BA)

Business analytics (BA) refers to the skills, technologies, practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning. Business analytics is used for prescriptive analysis, which is utilized to formulate optimization techniques for stronger business performance.
Business Analytics (BA) is the study of data through statistical and operations analysis, the formation of predictive models, application of optimization techniques, and the communication of these results to customers, business partners, and college executives.
Business Analytics (BA) is a combination of Data Analytics, Business Intelligence and Computer Programming. It is the science of analyzing data to find out patterns that will be helpful in developing strategies. Its usage can be found in almost every industry.

Analytics is the data driven decision making approach for a business problem. Analytics is a field which combines data, information technology, statistical analysis, quantitative methods and computer-based models into one. This all are combined to provide decision makers all the possible scenarios to make a well thought and researched decision. But Business Analytics has some specific focus. For example, business analytics in rural is used to determine pricing of various commodity in the nation based past and present set of information. For example, price data board is shown in table 1.

Table 1.1 Price Data Board for Commodity Price Date: 03.01.2020

State Name	District	Market Name	Group	Commodity
Variety	Grade	Min. Price (Rs/Quintal)	Max Price (Rs/Quintal)	Model Price (Rs/Quintal)

Source: https://agmarknet.gov.in/

Close to 97,000 positions related to analytics and data science are currently vacant due to a dearth of qualified talent, which is 45 percent over last year. The figure is poised to more than double to two lakhs by 2020. In India now also accounts for one in 10 advanced analytics job openings in the world. In India alone, demand for business analytics professionals grew by 400%; in contrast the supply grew by just 19%. Data Science The figure shows that in India is the place to be to participate in the booming industry of analytics and data science.

A study done by online ed-tech company Great Learning said here

Points to Remember

The Students should carry out market survey in your nearest market, observe the prices of the commodities particularly perishables (vegetables and fruits). From the ten days observation,

[1]. List out the prices for 5 vegetables for ten days

[2]. List out the prices for 5 fruits for ten days

[3]. Find out the minimum prices' vegetables

[4]. Find out the maximum prices' vegetables fruits

Case Study 1

Data Analytics (DA) to Business Analytics (BA)

Data Analytics is a field that uses technology, statistical techniques and big data to identify important business questions such as patterns and correlations. The implementation of data analytics in an organization may increase efficiency in gathering information and creating an actionable strategy for existing or new opportunities.

Business Analytics is focused on using the same big data tools as implemented with data analysis to determine business decisions and implement practical changes within an organization. Business analytics is implemented to identify weaknesses in existed procedures and to surface data that can be used to drive an organization forward in efficient and other measurements of growth. It is important to understand the similarities and differences between these fields when considering starting a career in either data analytics or a career in business analytics.

These fields will often share the same goal of increasing efficiency through data, but their differences are key. The skills, interests and background needed to be successful in these fields should be considered before you pursue one of these paths.

Required Skills for Business Analyst

Identify

Identifying the research questions within the business that can be (fully or partially) be solved using analytics. It should be noted that these questions might either be clearly stated by relevant stakeholders (eg. decision makers) or might require the analytics team to identify and clarify these research questions.

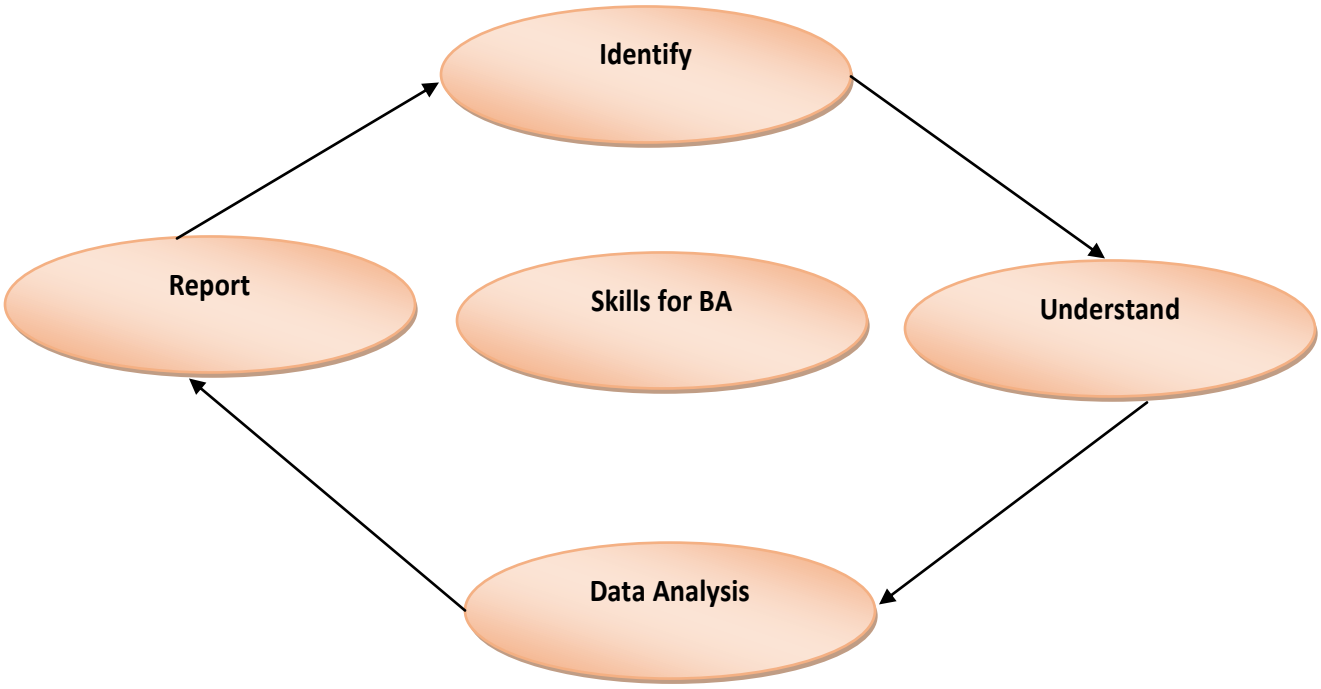


Fig. 1.1 Skills Required for Business Analytics

Understand

Understanding what data might be available to use within an Analytics Engine (an environment for analyzing these data). This includes both data collected internally within an organization, and data collected externally by other organizations.

Data Analytics

Data analysis (including both the initial analysis to obtain key metrics such as means and ranges of variables and levels of missing data, and a full analysis to understand the patterns and relationships between variables)

Report

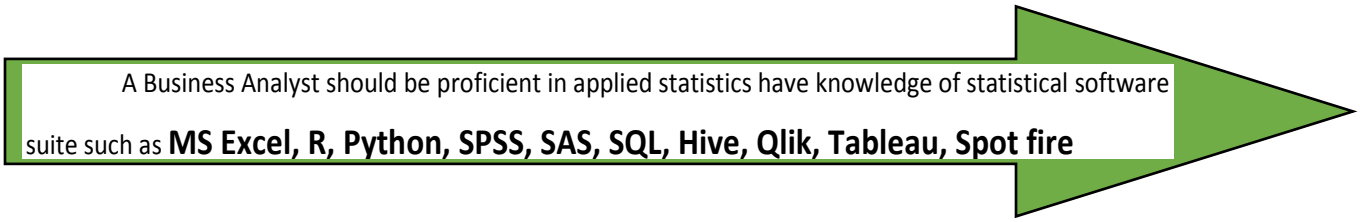
Reporting the results of analytics back to stakeholders (knowing that different stakeholders will be interested in different research questions, different levels of expertise in interpreting analytics results, different amounts of time available for reading through analytics results, potentially different levels of privacy and confidentiality and hence different types of results they have permission to view, and different needs in terms of timeliness – do they need to view results in real-time or would they prefer a monthly/annual report). Figure 1.1 shows skills required for business analytics.

The **Ministry of Electronics and Information Technology (MeitY)** has been trying to develop the **Common Service Centre (CSC)** scheme to help connect citizens in remote parts of the country. The government plans to use AI and data analytics to improve services across sectors such as finance, education, and healthcare, among others. CSCs are facilities created under the **Digital India programme**. CSCs function as access points for the delivery of government e-services. MeitY, through the **CSC SPV (special purpose vehicle) agency**, will collaborate with a private company to develop and deliver digital services to about 900 million citizens living in rural areas in the country.

Points to Remember

Software Skills

Software skills also required for business analytics. But it may vary depending on the organization's requirement. However, this is a basic knowledge framework required for making the cut.



- SQL various databases
- Excel Spreadsheets
- Tableau Software: Simple drag and drop tools for visualizing data from spreadsheets and other databases.
- IBM Cognos Express: An integrated business intelligence and planning solution designed to meet the needs of midsize companies, provides reporting, analysis, dashboard, scorecard, planning, budgeting and forecasting capabilities
- SAS / SPSS / Rapid Miner: Predictive modeling and data mining, visualization, forecasting, optimization and model management, statistical analysis, text analytics, and more using visual workflows
- R / Python: Advanced programming-based data preparation, analytics and visualization

1.2 Application of Business Analytics

Business analytics uses data construction of the business model. It uses business data such as annual reports, financial ratios, marketing research, etc. It uses the database which contains various computer files and information coming from data analysis. Business analytics has a wide range of application from customer relationship management, financial management, and marketing, supply-chain management, human-resource management, pricing and even in sports through team game strategies. Some of the applications are listed as below:

- Financial services analytics
- Fraud analytics
- Health care analytics
- Marketing analytics
- Pricing analytics
- Retail sales analytics
- Behavioral analytics
- Cyber analytics
- Supply chain analytics
- Talent analytics
- Transportation analytics

Customer Journey Analytics

For example, your ***OLA app***, might frequently issue discount coupons tailored for each customer based on his past traveling patterns. This practice encourages the customer to consider travelling the discounted but favorite place repeatedly, while building customer loyalty. This practice is possible, since a smart use of business analytics allows the ***OLA app*** to figure out which items are likely to be purchased by which customer in his next travelling trip. Likewise, application potentials of business analytics are enormous given the abundant data available from the digital and mobile data sources.

Application of Business Analytics in Rural Development

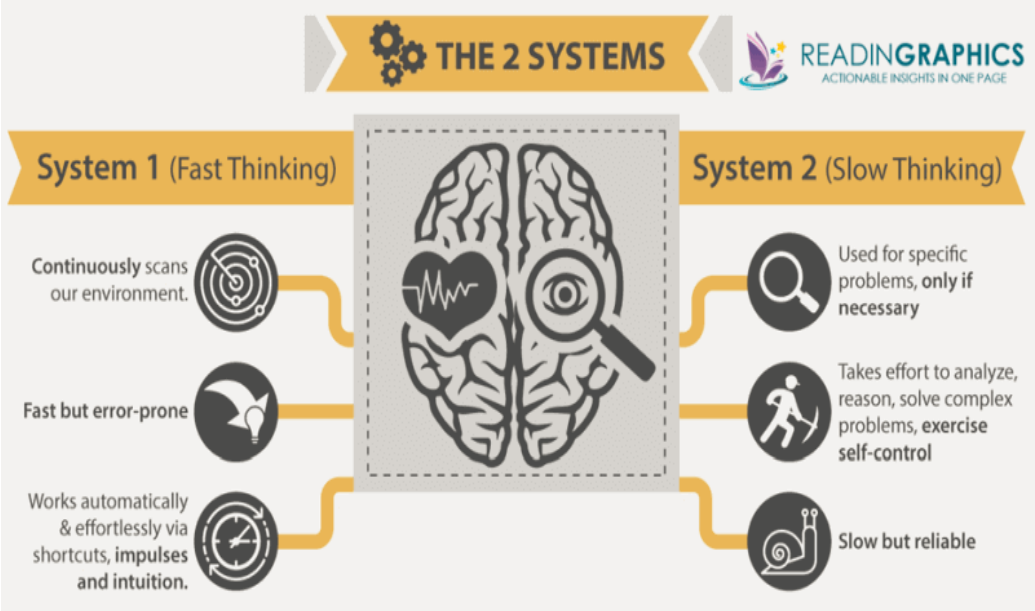
The following apps are making agriculture easy based on the business analytics

IFFCO किसान	**IFFCO Kisan Agriculture:** This app was launched in 2015 and is managed by IFFCO Kisan, a subsidiary of Indian Farmers' Fertilizer Cooperative Ltd. Its aim is to help Indian farmers make informed decisions through customized information related to their needs. The user can access a variety of informative modules including agricultural advisory, weather, market prices, agriculture information library in the form of text, imagery, audio and videos in the selected language at profiling stage. The app also offers helpline numbers to get in touch with Kisan Call Centre Services.
	RML Farmer – KrishiMitr : RML Farmer is a one of its kind agricultural app where farmers can keep up with the latest commodity and mandi prices, precise usage of pesticides and fertilizers, farm and farmer related news, weather forecast and advisory. Users can choose from over 450 crop varieties, 1300 mandis, and 3500 weather locations across 50,000 villages and 17 states of India. It works with the help of specific tools designed to analyze or provide information on different aspects of farming habits. Eg. CropDoc helps the farmers in identifying problems that affects their crops at the right time and suggests corrective actions; Farm Nutri provides general and personalized nutrient recommendations, which are presented in the form of a schedule of fertilizer dosage.

Points to Remember

The Two Types of Decision-Making in Our Brain

Chess prodigy Joshua Waitzkin writes about this distinction in The Art of Learning: A Journey in the Pursuit of Excellence. In his book, Waitzkin states that the best chess players are those that can take in the most information in a short span of time. They do so using the "fast thinking" system of their brain.Let us explain. As laid out by the Nobel winning psychologist Daniel Kahneman in his book Thinking, Fast and Slow, your brain has two different learning and decision making systems:



Source: Chess prodigy Joshua Waitzkin, The Art of Learning: A Journey in the Pursuit of Excellence

1.3 Statistics in Business Management and Analytics

Statistics is an important field of study because of its application in almost all walks of life. It is an important branch of mathematics. It is the *analysis, interpretation and presentation of data*. The role of statistics in business management and analytics are pivotal with the proper skill to implement statistical methods, the managers can increase the production capacity of any plant or find out the optimum production capacity, efficient management of work and employee performance, limit the wastage of resource, etc.

Applications of Statistics for Managers

A ***Marketing manager*** needs to gather and analyse a large amount of data pertaining to market dynamics and target customers. Ideally, marketing strategy depends up on the outcomes of a Market research, which involves statistical methods for collecting and analysing data, application of sampling techniques and evaluating the effect of various marketing strategies.

A ***Production manager*** would ideally use Statistical Process Control techniques to improve productivity and quality. Knowledge and application of Control Charts, Sampling techniques and Probability Distributions ensures better processes and products. This also leads to the reduction in production cost and higher profits. A ***HR manager*** would be interested in identifying the best approach to train employees and evaluate the impact of training. There is a need to measure attrition and understand the underlying factors.

For Finance ***manager***, crunching financial data and using financial techniques is an integral part of day-to-day job. Knowledge of Statistics enhances competency and proficiency of a manager as a researcher and therefore provides an edge.

To Do Activity

Applications of Statistics in Agriculture

Statistics offers information to answer some basic questions in agriculture –

What type of rice variety to produce?

How to produce the rice variety?

For whom to produce the rice variety?

This statistical information helps to understand the farmers' problems and formulation of agriculture policies. We will be able to see the measure of various farmers' income components and their compilations. It collects information on income, investment, saving, expenditure, etc and establishes the relationship between them.

1.4 Arithmetic Progression

A progression is a special type of sequence for which it is possible to obtain a formula for the n^{th} term. The Arithmetic Progression is the most used sequence in mathematics with easy to understand formulas. Let us see its three different types of definition.

It is a mathematical sequence in which the difference between two consecutive terms is always a constant

An arithmetic sequence or progression is defined as a sequence of numbers in which for every pair of consecutive terms, the second number is obtained by adding a fixed number to the first one

The fixed number that must be added to any term of an AP to get the next term is known as the common difference of the AP.

Common Difference and First Term

In this progression, for a given series, the terms used are the first term, common difference between the two terms and n^{th} term. Suppose, a_1, a_2, a_3... a_n is an AP, the common difference "d" can be obtained as;

$$d = a_2 - a_1 = a_3 - a_2 = a_4 - a_3 = a_5 - a_4 = \ldots\ldots\ldots\ldots = a_n - a_{n-1}$$

Where "d" is a common difference. It can be positive or negative. The AP can also be written in terms of common difference, as follows

$$a, a + d, a + 2d, a + 3d, a + 4d, \ldots\ldots\ldots\ldots\ldots a + (n-1)\, d$$

Where "a" is the first term of the progression

n^{th} Term of an AP

The formula for finding the n^{th} term of an AP is

$$a_n = a + (n-1)\, d$$

Now, let us consider the sequence, 5, 10, 15, 20, 25, 30… is considered as an arithmetic sequence with common difference 5.

$$a_1 = 5;\ a_2 = 10;\ a_3 = 15;\ a_4 = 20;\ a_5 = 25;$$

$$d = 10 - 5 = 15 - 10 = 20 - 15 = 25 - 20 = 5$$

Consider an AP to be: $a_1, a_2, a_3 \ldots a_n$. The general form of an A. P is shown in table 1.2

Table 1.2 General Terms in AP

Position of Terms	Representation of Terms	Values of Term	AP	AP
1	a_1	A	5	5
2	a_2	a + d	5 + 5	10
3	a_3	a + 2d	5 + (5x2)	15
4	a_4	a + 3d	5 + (5x3)	20
5	a_5	a + 4d	5 + (5x4)	25
.				
N	a_n	a + (n-1) d	..	..

In this table, a = First term; d = Common difference; n = number of terms; a_n = n^{th} term

Problem 1.1

Find out the Common difference of the AP the following series

The sequence is 2, 6, 10, 14…

Given Data:

$a_1 = 2;\ a_2 = 6$

Solution:

$d = a_2 - a_1$

$= 6 - 2$

$= 4$

Ans: d = 4

Problem 1.2

Find out the Common difference of Sequence is 3, 6, 9, 12….

Given Data:

$a_1 = 2$

$a_2 = 6$

Solution:

$d = a_2 – a_1$

$= 6 – 3$

$= 3$

Ans: d = 3

Problem 1.3

What is the common difference of the arithmetic progression 10, 5, 0, -5?

Given Data:

$a_1 = 10$

$a_2=5$

Solution:

$d = a_2 – a_1$

$= 5 – 10$

$= – 5$

Ans: d = – 5

Problem 1.4

Let an be an arithmetic progression, for $a_2=5$ and $a_1= –11$. Find the difference of the progression.

Given Data:

$a_1 = –11$

$a_2 = 5$

Solution:

$d = a_2 – a_1$

$= 5 – (– 11)$

$= 16$

Ans: d = 16

Problem 1.5

Do the numbers 2, 6, 10, 12, 16… Form an arithmetic progression?

Given Data:

$a_1 = 2$

$a_2 = 6$

$a_3 = 10$

$a_4 = 12$

$a_5 = 16$

Solution:

$d = a_2 - a_1 = a_3 - a_2 = a_4 - a_3 = a_5 - a_4$

$d = (6 - 4) = (10 - 6) = (12 - 10) = (16 - 12)$

$= 2 \neq 4 \neq 2 \neq 4$

Ans: If they formed, they would be 2, 6, 10, 14, and 18

Problem 1.6

Find the 20^{th} term for the given AP: 3, 5, 7, 9 …

Given Data:

$a = 3, d = 5 - 3 = 2, n = 20$

Solution:

$a_n = a + (n - 1) \times d$

$a_n = 3 + (20 - 1) \times 2$

$= 3 + 38$

$= 41$

Ans: $a_n = 41$

Problem 1.7

What is the seventh term of the arithmetic progression 2, 7, 12, 17…?

Given Data:

Given $a = 2, d = 7 - 2 = 5, n = 7$

Solution:

$a_n = a + (n - 1) d$

$a_n = 2 + (6) 5$

$a_n = 2 + 30$

$= 32$

Ans: $a_n = 32$

Problem 1.8

Find the 10^{th} term of the arithmetic progression 1, 3.5, 6, 8.5...

Given Data:

$n = 10$

$d = 2.5$ = Common Difference

$a = 1$ = First Term

Solution:

$a_n = a + (n - 1) d$

$= 1 + (9 * 2.5)$

$= 23.5$

Problem 1.9

Let an be an arithmetic progression, for which the first term a=1 and common difference d=1. Find a_{1083}

Solution:

$a = 1$

$d = 1$

$a_n = a + (n-1). d$

$a_{1083} = 1 + (1082)1$

$= 1083$

Ans: $a_{1083} = 1083$

Problem 1.10

If $a_1 = 4$; $a_2 = 7$; then determine a_{11}^{th} term

Given Data:

$a_1 = 4$

$a_2 = 7$

$d = 3$

$n = 11$

Solution:

$a_n = a + (n - 1) \times d$

$a_{11} = 4 + (10) * 3$

$= 34$

Ans: $a_{11} = 34$

Problem 1.11

Let an be an arithmetic progression, for which d=12 and a_3=43. Find initial term

Given Data:

$d = 12$

a_3=43

n=3

Solution:

$a_n = a+(n-1).d$

$a_3 = a + (3-1) d = 43$

$a + (2)12 = 43$

a= 43 – 24

a = 19

Ans: a = 19

Sum of First n Terms of an AP (We don't know the Last term in this case)

For any progression, the sum of n terms can be easily calculated. For an AP, the sum of the first n terms can be calculated if the first term and the total terms are known. The formula for the arithmetic progression sum is explained below:

$$S = n/2[2a + (n - 1) \times d]$$

Sum of AP when Last Term is given

Formula to find sum of AP when first and last terms are given as follows:

$$S = n/2 \text{ (first term + last term)}$$

Problem 1.12

What is the sum of the first 50 odd positive integers?

Given Data:

The Sequence is 1, 3, 5, 7……. (We don't know the Last term in this case)

n = 50

d = 2 = Common Difference

a = 1 = First Term

Solution:

$S = n/2[2a + (n - 1) \times d]$

$S = 50/2 [2 (1) + (49) * 2]$

$= 2500$

Ans: S = 2500

Problem 1.13

What is the sum of the first 100 positive integers?

Given Data:

The Sequence: 1, 2, 3, 4……100

First term = 1

Last term = 100

Solution:

S = n/2 (first term + last term)

S = 100/2 (1 + 100)

= 50 (101)

= 5050

Ans: S = 5050

Problem 1.14

Find the sum of the first 10 natural numbers.

Given Data:

As we know, the natural numbers form an arithmetic progression with The Sequence is 1, 2, 3….10

First Term = a = 1

Last Term = l = 10

Solution:

S = n/2 (a + l)

S = 10/2 (1+10)

=5 (11)

Ans: S = 55

Problem 1.15

The sum of five consecutive numbers is 100. Find the first number.

Given Data:

Consecutive numbers form an arithmetic progression with difference d = 1

S = 100

n = 5

Solution:

S = n/2[2a + (n − 1) × d]

S = 5/2 [2(a) + (4)(1)] = 100

2.5[2a + 4] = 100

2a + 4 = 100/2.5

2a + 4 = 40

a = (40 – 4)/2

Ans: a = 18

First Term = 18

The first number is 18, and the other numbers are 18, 19, 20, 21, and 22.

1.5 Geometric Progression

The geometric sequence is sometimes called the geometric progression or GP, for short. A geometric sequence is a sequence such that any element after the first is obtained by multiplying the preceding element by a constant called the common ratio which is denoted by r. The common ratio (r) is obtained by dividing any term by the preceding term, i.e.,

$$r = a_2/a_1 = a_3/a_2 = a_4/a_3 = \ldots\ldots\ldots\ldots\ldots a_n/a_{n-1}$$

r	common ratio
a_1	first term
a_2	second term
a_{n-1}	the term before the $n\text{-}1^{th}$ term
a_n	the n^{th} term

The sequence 1, 3, 9, 27, 81 is a geometric sequence. Note that after the first term, the next term is obtained by multiplying the preceding element by 3.

The GP can also be written in terms of common ratio, as follows:

$$a_1, a_1 * r, a_1 * r^2, a_1 * r^3, a_1 * r^4, \ldots\ldots\ldots\ldots\ldots\ldots a_1 * r^{n-1}, a_1 * r^n$$

To find the nth term of a geometric sequence we use the formula:

$$a_n = a_1 * r^{n-1}$$

Sum of Terms in a Geometric Progression

Finding the sum of terms in a geometric progression is easily obtained by applying the formulas: n^{th} partial sum of a geometric sequence

$$S_n = (a_1 * (1-r^n)) \div (1-r) \qquad r \neq 1$$

sum to infinity

$$S_\alpha = \sum(a_1 * r^{n-1}) = a_1 \div (1-r) \qquad -1 < r < 1$$

Limit of n = 1 to n = α

S_n sum of GP with n terms

S_∞ sum of GP with infinitely many terms

Problem 1.16

Find the number of terms in the geometric progression 6, 12, 24, ..., 1536

Given Data:

$a_1 = 6$

$a_2 = 12$

$a_3 = 24$

$a_n = 1536$

Solution:

$r = a_2/a_1$

$= 12/6$

$r = 2$

$a_n = a_1 * r^{n-1}$

$1536 = 6 * 2^{n-1}$

$1536/6 = 2^{n-1}$

Changes 256 to its exponential form

$256 = 2^{n-1}$

$2^8 = 2^{n-1}$

Equate the indices since they both have the same base

$8 = n - 1$

$n = 8 + 1$

Hence the ninth term is 1536

Ans: n = 9

Problem 1.17

Find the Common ratio of the geometric series: – 2, 1/2, – 1/8, ……– 1/37268

Given Data:

$a_1 = -2$

$a_2 = ½$

Solution:

$r = a_2/a_1$

$= (1/2)/(-2)$

Ans: r = - 1/4

Problem 1.18

Ram gives his son Rs. 100 on one day, Rs. 50 on the second day, Rs. 25 on third day and so on. What will be total amount given by Ram to his son starting from the first day, if he lives forever?

Given Data:

In this question, every day he is giving half the amount, he has given the previous day and he must pay forever. This makes it an infinite GP series.

Geometric series: 100, 50, 25 ……. α

Solution:

$a_1 = 100$

$a_2 = 50$

$r = a_2/a_1$

$= 50/100$

r = 1/2

$S_\alpha = a_1 \div (1-r)$

$= 100 \div (1-(1/2))$

$= 100 \div (1/2)$

Ans: S_α = 200

Problem 1.19

Find the sixth term in the geometric progression 6, 12, 24, 48 ……

Given Data:

$a_1 = 6$

$a_2 = 12$

Solution:

$r = a_2/a_1$

$= 12/6$

$r = 2$

$n = 6$

$a_n = a_1 * r^{n-1}$

$= 6 * 2^5$

Ans: $a_n = 192$

Problem 1.20

1^{st} term is 64 and the 5^{th} term is 4. If the sum of all terms is 128, what is the common ratio?

Given Data:

$a_1 = 64$; $S_5 = 128$

Series - sum to infinity

Solution:

$S_\alpha = a_1 \div (1-r)$

$128 = 64 \div (1-r)$

$1/2 = (1-r)$

$r = 1 - (1/2)$

Ans: r = 1/2

Problem 1.21

Find the sum of the first 5 terms of the following series. Given that the series is finite: 3, 6, 12 …

Given Data:

$a_1 = 3$; $n = 5$; $a_2 = 6$

Solution:

$r = a_2/a_1$

$= 6/3$

$r = 2$

$S_n = (a_1*(1-r^n)) \div (1-r)$

$S_5 = (3 * (1 - 2^5)) \div (1-2)$

Ans: $S_5 = 93$

To check the answer:

$a_1 = 3$; r = 2; geometric series: 3, 6, 12, 24, 48; $S_5 = 3 + 6 + 12 + 24 + 48$

Ans: $S_5 = 93$

Additional Problems

Problem 1.22

Find the 15^{th} term of an arithmetic progression whose first term is 2 and the common difference is 3.

Given Data:

Given a = 2, d = 3

Solution:

$a_n = a + (n - 1) d$

$a_n = 2 + (15-1) 3$

$a_n = 2 + (14) 3$

$a_n = 2 + 42$

$= 44$

Ans: $a_n = 44$

Problem 1.23

What is the sum of the first 15 terms of an A.P whose 11^{th} and 7^{th} terms are 5.25 and 3.25 respectively?

Given Data:

$a_{11} = 5.25$; $a_7 = 3.25$; (We don't know the Last term in this case)

Solution:

$a_{11} = 5.25$

$a_n = a + (n - 1) d$

$a + 10 d = 5.25$ .....(1)

$a_7 = 3.25$

$a_n = a + (n - 1) d$

$a + 6 d = 3.25$ .....(2)

Solve the equation 1 and 2, we can get

$a = 1/4$ and $d = 1/2$

$S = n/2[2a + (n - 1) \times d]$

$S = 15/2 [2 (1/4) + (14) * (1/2)]$

$S = 7.5 [(1/2) + 7)]$

$S = 7.5 [7.5]$

= 56.25

Ans: S = 56.25

Problem 1.24

Find the value of n. If a = 10, d = 5, a_n = 95.

Given Data:

Given a = 10, d = 5, a_n = 95

Solution:

$a_n = a + (n - 1) d$

$95 = 10 + (n-1) 5$

$95 = 10 + 5n – 5$

$95 – 10 + 5 = 5n$

$90 = 5n$

Ans: n = 18

Problem 1.25

The first term of an Arithmetic Progression is 15 and the last term is 85. If the sum of all terms is 750, what is the 6^{th} term?

Given Data:

First term = 15

Last term = 85

Solution:

Step 1:

S = n/2 (first term + last term)

750 = n /2 * (15 + 85)

750 = n /2 * (100)

1500 = n * (100)

n = 15

Step 2:

$a_n = a + (n - 1) d$

$85 = 15 + (15-1) d$

$85 = 15 + (14) d$

$85 = 15 + 14 d$

$85 - 15 = 14 d$

$70 = 14 d$

d = 5

Step 3:

6^{th} term of this Arithmetic Progression a = 15, d = 5

$a_n = a + (n - 1) d$

$a_n = 15 + (6-1) 5$

$a_n = 15 + (5) 5$

$a_n = 15 + 25$

$= 40$

Ans: $a_n = 40$

Problem 1.26

The first term of an AP is 10 and the last term is 28. If the sum of all terms is 190, what is the common difference?

Given Data:

First term = 10

Last term = 28

Solution:

Step 1:

190 = n/2 (first term + last term)

190 = n /2 * (10 + 28)

190 = n /2 * (38)

380 = n * (38)

n = 10

Step 2:

$a_n = a + (n - 1) d$

28 = 10 + (10-1) d

28 = 10 + (9) d

28 - 10 = 9 d

18 = 9 d

d = 2

Ans: d = 2

Problem 1.27

The sum of three numbers in an Arithmetic Progression is 45 and their product is 3000. What are the three numbers?

Given Data:

Assuming that the numbers

a - d, a , a + d

Solution:

Sum of the three number

(a – d) + a + (a + d)= 45

3a = 45

a = 15

Product of the three numbers

(a – d) * a * (a + d) = 3000

(a – d) * (a + d) = 3000/a

(a – d) * (a + d) = 200

(15-d)*(15+d) = 200

(15-d)*(15+d) = 200

$(a+b) * (a-b) = a^2 - b^2$

$(15^2) - (d^2) = 200$

$(225) - (d^2) = 200$

$(d^2) = 25$

d = 5

Ans: The Numbers are 15,20,25

Problem 1.28

The 5^{th} term of an AP is 17/6 and the 9^{th} term is 25/6. What is the 12^{th} term?

Given Data:

$a_5 = 17/6$; $a_9 = 25/6$

Solution:

$a_5 = 17/6$

$a_n = a + (n - 1) d$

a + 4 d = 17/6 .....(1)

$a_9 = 25/6$

$a_n = a + (n - 1) d$

a + 8 d = 25/6 .....(2)

Solve the equation 1 and 2, we can get

a = 3/2 and d = 1/3

$a_n = 3/2 + (12 - 1) (1/3)$

$a_n = 3/2 + (11) (1/3)$

$a_n = 3/2 + (11/3)$

= 31/6

Ans: a_n = 31/6

Problem 1.29

The sum of the first six terms of an AP is 48 and the common difference is 2. What is the 4^{th} term?

Given Data:

n = 6

d = 2

s = 48

Solution:

S = n/2[2a + (n − 1) × d]

48 = 6/2 [2 a + (5) * 2)]

48 = 3 [2 a + 10]

48 = 3 [2 a + 10]

16 = 2 a + 10

2 a = 16 - 10

2 a = 6

= 3

$a_n = a + (n - 1)\ d$

$a_n = 3 + (4-1)\ 2$

$a_n = 3 + (3)\ 2$

$a_n = 3 + 6$

= 9

Ans: $a_n = 9$

Problem 1.30

Find the fifth term in the geometric progression 1/7, 1/14, 1/28…….

Given Data:

$a_1 = 1/7$

$a_2 = 1/14$

Solution:

$r = a_2/a_1$

= (1/14)/(1/7)

r = 1/2

n = 5

$a_n = a_1 * r^{n-1}$

$= 1/7 * (1/2)^4$

= 1/7 * (1/16)

= 1/116

Ans: $a_n = 1/116$

Problem 1.31

Find 9th term of the following series: 5, 10, 20, 40 ……

Given Data:

$a_1 = 5$

$a_2 = 10$

Solution:

$r = a_2/a_1$

= (10)/(5)

r = 2

n = 9

$a_n = a_1 * r^{n-1}$

$= 5 * (2)^8$

= 5 * (256)

= 1280

Ans: $a_n = 1280$

Problem 1.32

The 3rd and the 8th term of a G. P. are 4 and 128 respectively. Find the G. P.

Given Data:

$a_3 = 4$; $a_8 = 128$

Solution:

n = 3

$a_n = a_1 * r^{n-1}$

$a_3 = a_1 * r^2$

$a_1 * r^2 = 4$ …… (1)

n = 8

$a_8 = a_1 * r^{n-1}$

$a_8 = a_1 * r^7$

$a_1 * r^7 = 128$ …… (2)

Equation (2) divided by equation (1)

$128/4 = (a_1 * r^7)/(a_1 * r^2)$

$32 = r^5$

$r = 2$

Sub suite equation 1

$a_1 * 4 = 4; a_1 = 1$

Ans: G.P= 1, 2, 4, 8, 16,.....

Problem 1.33

Which term of the G. P.: 6, –12, 24, – 48 ... is 384

Given Data:

$a_1 = 6$

$a_2 = -12$

Solution:

$r = a_2/a_1$

$= -12/6$

$r = -2$

$a_n = a_1 * r^{n-1}$

$384 = 6 * (-2)^{n-1}$

$64 = (-2)^{n-1}$

$(2)^6 = (-2)^{n-1}$

$6 = n - 1$

$n = 6 + 1$

Ans: n = 7

Problem 1.34

Find the sum of GP: 1, 2, 4, 8 ... up to the 10^{th} term.

Given Data:

$a_1 = 1; r = 2$

Solution:

$n = 10$

$a_n = a_1 * r^{n-1}$

$= 1 * 2^9$

Ans: 2^9

Problem 1.35

Find the sum of an infinite GP 3, 1, 1/3…?

Solution:

$a_1 = 3$

$a_2 = 1$

$r = a_2/a_1$

$= 1/3$

r = 1/3

The sum of an infinite GP

$S_\alpha = a_1 \div (1-r)$

$= 3 \div (1-(1/3))$

$= 3 \div (2/3)$

Ans: $S_\alpha = 9/2$

To Do Activity

Take any one commodity and do the following activities

Pricing setting prices for the commodity	**Customer segmentation** identifying and targeting key customer groups	**Merchandising** determining brands to buy, quantities, and allocations	**Location** finding the best location for where to buy

Business Analytics can be used to implement personal causes as well. Like maybe it can be used to calculate or analyze the household applications or energy consumptions till a certain level. Well, this can be done as one can take care of the amount of consumption going on and based on that a lot of prediction or further calculation can be done as well. For example, if it is somehow possible to check and calculate the regular consumption of electricity it will soon be possible to trace out a monthly figure for the same. This prediction can help find out the amount of consumption that might be a common figure for the nearby households as well. Similarly, the same logic or procedure can be applied to trace the rate of consumption in the entire local area and then maybe finally it will be possible to make some further estimation for the entire year. Apart from that, it can help one decide how much energy is being conserved and how it can be used for better energy efficiency or for better utilization of energy resources. In fact, this can help improve the utilization of natural and other resources as well. ***Analysis in your home electricity bill with respect to electricity consumption and cost***

Case Study 2: Can Business Analytics be used for Personal Purpose?

Further can Learn Business Analytics Courses from
Coursera, Greatlearning, MIT, NPTEL

1. https://www.coursera.org/specializations/business-analytics
2. https://www.greatlearning.in/business-analytics-courses-chennai
3. https://executive-ed.mit.edu/applied-business analytics/index.php?utm_source=Google&utm_medium=CPC&utm_campaign=B-9629_IN_GG_SE_ABA_MAY_20_Generic&utm_content=c&utm_term=%2Bbusiness%20%2Banalytics%20%2Bcourse
4. https://nptel.ac.in/courses/110105089/

Points to Remember

Summary of the Chapter

1. **Business Analytics (BA) is the study of data through statistical and operations analysis, the formation of predictive models, application of optimization techniques, and the communication of these results to customers, business partners and college executives.**
2. **The AP can be written in terms of common difference, as follows**
3. **a, a + d, a + 2d, a + 3d, a + 4d, ……a + (n-1) d**
4. **The formula for finding the nth term of an AP is an = a + (n-1) d**
5. **Sum of AP when Last Term is given S = n/2 (first term + last term)**
6. **Sum of First n Terms of an AP (We don't know the Last term in this case)**
7. **S = n/2 [2a + (n – 1) × d]**
8. **Geometric Progression: a_1, a_1* r, a_1* r^2, a_1* r^3, a_1* r^4, ……a_1* r^{n-1}, a_1* r^n**
9. **nth term of a geometric sequence we use the formula: an = a1* rn-1**
10. **Sum of Terms in a Geometric Progression $S_n = (a_1*(1-r^n)) \div (1-r)$ $r \neq 1$**
11. **Sum to infinity $S_\alpha = \sum(a_1*r^{n-1}) = a_1 \div (1-r)$ $-1 < r < 1$**

Model Questions

Part A

[1]. Define: Business Analytics
[2]. List out the difference between Business Analytics and Data Analytics
[3]. Remember: Skills required for Business Analytics
[4]. Mention any four applications for Business Analytics
[5]. List out the applications of statistics in agriculture
[6]. Write down the formula for sum of terms in Geometric Progression
[7]. Write down the formula for n^{th} term of an AP
[8]. List out the difference between Arithmetic Progression and Geometric Progression

Part B

[1]. What is the common difference of the arithmetic progression 5, 10, 15, 20?

[2]. Find the 20^{th} term for the given AP: 6, 12, 18, 24 …

[3]. Find the 10^{th} term of the arithmetic progression 4, 8, 12, 16, 20 ...

[4]. If $a_1 = 4$; $a_2 = 7$; then determine a_{11}^{th} term

[5]. 1^{st} term is 64 and the 5^{th} term is 4. If the sum of all terms is 128, what is the common ratio?

[6]. Find the sixth term in the geometric progression 2, 4, 8, 16. ……

[7]. Find the number of terms in the geometric progression 6, 12, 24, ...60

Chapter 2 Permutations and Combinations

Introduction

Life is full of uncertainties. We don't know the outcomes of a situation until it happens.

[1]. Will it rain today?
[2]. Will I pass the next business analytics test?
[3]. Will my favorite Indian cricket team win the toss?
[4]. Will I get a promotion in next 6 months?

All these questions are examples of uncertain situations we live in.

Future events are far from certain in the business world. This is especially true for smaller businesses/agriculture, which tend to have more volatility than larger organizations, or newer businesses without a proven track record of sales and costs. For this reason, probability can be a great tool for estimating future returns and profitability. Concepts of probability theory are the backbone of many important concepts in business analytics like inferential statistics to Neural Networks.

Objectives

- To understand Probability and its role in business analytics in problem solving and decision making that is imperative for business organizations
- To become familiar with Permutations, Combinations and Binomial Theorem

Structure

2.1 Probability in Business Analytics

2.2 Permutations

2.3 Combinations

2.4 Relationship between Permutations and Combinations

2.5 Binomial Theorem

It would not be wrong to say that the journey of mastering Business Analytics begins with probability.

2.1 Probability in Business Analytics

Probability is a numerical description of how likely an event is to occur or how likely it is that a proposition is true. Probability is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility and 1 indicates certainty.

Probability Terminologies

Few of the common terminologies which is playing major role in probability.

- Experiment – are the uncertain situations, which could have multiple outcomes. ***Whether it rains on daily basis is an experiment***.
- Outcome is the result of a single trial. So, if it rains today, the outcome of today's trial from the experiment is "***It rained***"
- Event is one or more outcome from an experiment. "***It rained***" is one of the possible event for this experiment. "***Not rained***" is another possible event for this experiment.
- Probability is a measure of how likely an event is. So, if it is 60% ***chance*** that it will rain tomorrow, the probability of Outcome "it rained" for tomorrow is 0.6

Application of Probability in Business Analytics

Case 1:

Scenario Analysis

Probability distributions can be used to create scenario analyses. A scenario analysis uses probability distributions to create several, theoretically distinct possibilities for the outcome of a course of action or future event. For example, a business might create three scenarios: worst-case, likely and best-case. The worst-case scenario would contain some value from the lower end of the probability distribution; the likely scenario would contain a value towards the middle of the distribution; and the best-case scenario would contain a value in the upper end of the scenario. For example, ***day to day stock market position***

Case 2:

Sales Forecasting

One practical use for probability distributions and scenario analysis in business is to predict future levels of sales. It is essentially impossible to predict the precise value of a future sales level; however, businesses still need to be able to plan for future events. Using a scenario analysis based on a probability distribution can help a company frame its possible future values in terms of a likely sales level and a worst-case and best-case scenario. By doing so, the company can base its business plans on the likely scenario but still be aware of the alternative possibilities.

To Do Activity

Take any location or state in India

[1]. Collect rain fall data for last 25 years

[2]. Link rainfall and agriculture productivity

[3]. Do risk evaluation for agriculture productivity

[4]. Do rainfall forecasting for the coming years

2.2 Permutations

Permutation is a term used to understand the concept of the arrangement of things. The word arrangement is referred if the order of things is considered. In mathematics, permutation relates to the act of arranging all the members of a set into some sequence or order, or if the set is already ordered, rearranging its elements, a process called permuting. Permutations occur, in prominent ways, in almost every area of mathematics. They often arise when different orderings on certain finite sets are considered.

A permutation is defined as an arrangement in a definite order of a number of objects taken some or all at a time. Counting permutations are merely counting the number of ways in which some or all objects at a time are rearranged.

The different arrangements of a given number of things by taking some or all at a time, are called permutations.

A permutation is an ordering, or arrangement, of the elements in a finite set.

All permutations (or arrangements) made with the letters a, b, c by taking all at a time are (abc, acb, bac, bca, cab, cba)

All permutations (or arrangements) made with the letters a, b, c by taking two at a time are (ab, ba, ac, ca, bc, cb).

Factorial

Where the symbol "!" denotes the factorial, which means that the product of all the integer less than or equal to n but it should be greater than or equal to 1.

For example,

$1! = 1$

$2! = 1 \times 2 = 2$

$3! = 1 \times 2 \times 3 = 6$

$4! = 1 \times 2 \times 3 \times 4 = 24$, which are the factors of the given number.

Points to Remember

Number of Permutations

A permutation is the choice of r things from a set of n things without replacement and where the order matters.

The convenient expression to denote permutation is defined as ${}_nP_r$.

$$d = nPr = (n!) / (n-r)! \qquad 0 \le r \le n$$

Number of all permutations of n things, taken r at a time, is given by:

$$nP_r = n * (n - 1) * (n - 2) \dots * (n - r + 1) = \frac{n!}{(n - r)!}$$

Cor. number of all permutations of n things, taken all at a time = n!

Examples:

$6P_2 = (6 \times 5) = 30.$

$7P_3 = (7 \times 6 \times 5) = 210.$

Problem 2.1

Evaluate: 10! / 6!

Solution:

10! = 10 x 9 x 8 x 7 x 6 x 5 x 4 x 3 x 2 x1

6! = 6 X 5 x 4 x 3 x 2 x 1

10! / 6! = 10 x 9 x 8 x 7 = 5040

Ans: 5040

Problem 2.2

Find the number of permutations if n = 12 and r = 2?

Given Data:

n = 12

r = 2

Solution:

Permutation:

$nP_r = (n!) / (n-r)!$

$12P_2 = (12!) / (12-2)!$

= 12! / 10!

= (12 x 11 x 10!)/ 10!

= 132

Ans: 132

Problem 2.3

From a team of 6 students, in how many ways can we choose a captain and vice-captain assuming one person can not hold more than one position?

Given Data:

From a team of 6 students, two students are to be chosen in such a way that one student will hold only one position. Here, the no. of ways of choosing a captain and vice-captain is the permutation of 6 different things taken 2 at a time.

n = 6

r = 2

Solution:

Permutation:

$nP_r = (n!) / (n-r)!$

$6P_2 = (6!) / (6-2)!$

$= 6! / 4!$

$= (6 \times 5 \times 4!) / 4!$

$= 30$

Ans: 30

Problem 2.4

How many words can be formed using all the letters of the word Dinomite, using each letter exactly one time?

Given Data:

From a team of 6 students, two students are to be chosen in such a way that one student will hold only one position. Here, the no. of ways of choosing a captain and vice-captain is the permutation of 6 different things taken 2 at a time.

n = 6

r = 2

Solution:

Permutation:

$nP_r = (n!) / (n-r)!$

$6P_2 = (6!) / (6-2)!$

$= 6! / 4!$

$= (6 \times 5 \times 4!) / 4!$

$= 30$

Ans: 30

Problem 2.5 How many words can be formed using all the letters of the word Dinomite, using each letter exactly one time?

Given Data:

As there are 7 different letters in the word – Dinomite. So the number of different words formed using these 7 letters

n = 8

r = 7

Solution:

Permutation:

$nP_r = (n!) / (n-r)!$

$8P_7 = (8!) / (8-7)!$

$= 8! / 1!$ $(1! = 1)$

$= 8!$

Ans: 8 !

Problem 2.6 how many ways can the letters of the word 'LEADING' be arranged in such a way that the vowels always come together?

Given Data:

The word '**LEADING**' has **7** different letters.

When the vowels **EAI** are always together, they can be supposed to form one letter.

Then, we have to arrange the letters **LNDG (EAI)**

Solution:

Now, 5 (4 + 1 = 5) letters can be arranged in 5! = 120 ways.

The vowels (EAI) can be arranged among themselves in 3! = 6 ways.

$\therefore$ Required number of ways = (120 x 6) = 720.

Ans: 720

Problem 2.7 In how many ways can the letters of the word 'CORPORATION' be arranged so that the vowels always come together?

Given Data:

In the word 'CORPORATION', we treat the vowels OOAIO as one letter.

Thus, we have CRPRTN (OOAIO).

Solution:

This has 7 (6 + 1) letters of which R occurs 2 times and the rest are different.

Number of ways arranging these letters = $\frac{7!}{2!}$ = 2520.

Now, 5 vowels in which O occurs 3 times and the rest are different, can be arranged in

Number of ways arranging these letters = $\frac{5!}{3!}$ = 20.

Required number of ways = (2520 x 20) = 50400.

Ans: 50400

Permutation when all the Objects are Distinct

There are some theorems involved in finding the permutations when all the objects are distinct. They are:

Theorem 1:

If the number of permutations of n different objects taken r at a time, it will satisfy the condition $0 < r \leq n$ and the objects which do not repeat is $n (n – 1) (n – 2)……(n – r + 1)$, then the notation to denote the permutation is given by " $\mathbf{nP_r}$"

Theorem 2:

The number of permutations of different objects "n" taken "r" at a time, where repetition is allowed and is given by $\mathbf{n^r}$

Permutation When all the Objects are not Distinct Objects

Theorem 3:

To find the number of permutations of the object's 'n', and 'p' are of the objects of the same kind and rest is all different is given as **n! / p!**

Theorem 4: The number of permutations of n objects, where p_1 are the objects of one kind, p_2 are of the second kind, …,p_k is of the k^{th} kind and the rest, if any, are of a different kind then the permutation is given by

$n! / (p_1! * p_2! * … * P_k!)$

Problem 2.8 It is needed to seat 5 boys and 4 girls in a row so that the girl get the even places. How many such arrangements are possible?

Solution:

5 boys and 4 girls are to be seated in a row so that the girl gets the even places.

The 5 boys can be seated in 5! ways.

For each of the arrangement, the 4 girls can be seated only at the places which are cross marked to make girls occupy the even places).

B x B x B x B x B

So, the girls can be seated in 4! ways.

Hence, the possible number of arrangements = 4! × 5! = 24 × 120 = 2880

Ans: 2880

Problem 2.9 In how many ways can the letters of the word 'LEADER' be arranged?

Given Data:

The word 'LEADER' contains 6 letters, namely 1L, 2E, 1A, 1D and 1R.

Solution:

The number of permutations of n objects, where p_1 are the objects of one kind, p_2 are of the second kind, …, p_k is of the k^{th} kind and the rest, if any, are of a different kind then the permutation is given by

$n! / (p_1! * p_2! * \ldots * P_k!)$

$\therefore$ Required number of ways = $\frac{6!}{(1!)(2!)(1!)(1!)(1!)}$

Ans: 360

2.3 Combination

The combination is a way of selecting items from a collection, such that (unlike permutations) the order of selection does not matter. In smaller cases, it is possible to count the number of combinations. Combination refers to the combination of n things taken k at a time without repetition. To refer to combinations in which repetition is allowed, the terms k-selection or k-combination with repetition are often used.

The combination is a selection of a part of a set of objects or selection of all objects when the order doesn't matter

Each of the different groups or selections which can be formed by taking some or all a number of objects is called a combination.

The term combination means selection of things; it is used when the order of things is not important.

Combinations

Suppose we want to **select two out of three boys -A, B, C**. Then, possible selections are **AB, BC and CA**.

Note: AB and BA represent the same selection.

All the combinations formed by a, b, c taking **ab, bc, ca**.

The only combination that can be formed of three letters a, b, c taken all at a time is **abc**.

Various groups of **2 out of four persons A, B, C, D** are:

AB, AC, AD, BC, BD, CD.

Note that ab, ba are two different permutations but they represent the same combination.

Points to Remember

The convenient expression to denote combinations is defined as nC_r. The number of combinations of n objects taken r at a time and the combination formula is given by.

$$nC_r = \frac{n!}{r\,!\,(n-r)!}$$

Problem 2.10 Evaluate $10C_3$

Given Data:

n = 10

r = 3

Solution:

combinations:

$nC_r = (n!) / r\,!\,(n-r)!$

=10! / (10-3)! 3!

=10! / 7! 3!

=120

Ans: 120

Problem 2.11 Teacher asks a student to choose 6 items from the table. If the table has 20 items to choose, how many ways could the students choose the things?

Given Data:

Here, student has to choose 6 items from 20 items. Here, r= 6 and n= 20

Solution:

Combination,

$nC_r = (n!) / r\,!\,(n-r)!$

$20C_6$ = 20! / (20-6)! 6!

Ans: 38760

Problem 2.12 In how many ways a committee consisting of 5 men and 3 women, can be chosen from 9 men and 12 women.

Solution:

Choose 5 men out of 9 men = $9C_5$ways = 126 ways

Choose 3 women out of 12 women = $12C_3$ ways = 220 ways

The committee can be chosen in 126 x 220 = 27720 ways

Ans: 27720

Problem 2.13 How many words can be formed each of 2 vowels and 3 consonants from the letters of the given word – DAUGHTER ?

Given Data:

No. of Vowels in the word – DAUGHTER are 3.

No. of Consonants in the word Daughter are 5.

Solution:

No of ways to select a vowel = $3C_2 = 3!/ 2! (3 – 2)! = 3$

No. of ways to select a consonant = $5C_3 = 5!/ 3! (5 – 3)! = 10$

Now you know that the number of combinations of 3 consonants and 2 vowels = 10 x 3 = 30

And these can be arranged in 30 x 5! = 3600 ways

Ans: 3600

Problem 2.14 Find the number of 5-card combinations out of a deck of 52 cards if each selection of 5 cards has exactly one king.

Given Data:

Take a deck of 52 cards,

To get exactly one king, 5-card combinations have to be made. It should be made in such a way that in each selection of 5 cards, or in a deck of 52 cards, there will be 4 kings.

Solution:

To select 1 king out of 4 kings = $4C_1$

To select 4 cards out of remaining 48 cards = $48C_4$

To get the needed number of 5 card combination = $4C_1 \times 48C_4$

Ans: $4C_1 \times 48C_4$

Problem 2.15 In a class, there are 27 boys and 14 girls. The teacher wants to select 1 boy and 1 girl to represent the class for a function. In how many ways can the teacher make this selection?

Solution:

Here the teacher is to perform two operations:

(i) Selecting a boy from among the 27 boys $27C_1$

(ii) Selecting a girl from among 14 girls $14C_1$

Therefore, the total required ways is = $27C_1 * 14C_1$

Ans: $27C_1 * 14C_1$

Problem 2.16 In a group of 6 boys and 4 girls, four children are to be selected such that at least one boy should be there. In how many ways can they be selected such that at least one boy should be there?

Solution:

Option 1

Hence, we have 4 choices as given below; we can select 4 boys Number of ways to this = $6C_4$

Option 2

We can select 3 boys and 1 girl

Number of ways to this = $6C_3 \times 4C_1$

Option 3

We can select 2 boys and 2 girls

Number of ways to this = $6C_2 \times 4C_2$

Option 4

We can select 1 boy and 3 girls

Number of ways to this = $6C_1 \times 4C_3$

Total number of ways

$= (6C_4) + (6C_3 \times 4C_1) + (6C_2 \times 4C_2) + (6C_1 \times 4C_3)$

$= (6C_2) + (6C_3 \times 4C_1) + (6C_2 \times 4C_2) + (6C_1 \times 4C_1)$

$$nC_r = nC_{(n-r)}$$

= 15 + (20 x 4) + (15x6) + (6x4)

= 15 + 80 + 90 + 24

= 209

Ans: 209

Problem 2.17A group of 12 women and 5 men are used to pick a committee of 6 people. What is the possible outcomes if

a) 5 women and 1 man is selected b) any mixture of women and men

Solution:

a) From the FCP we know that two decisions will be made, choosing 5 women out of 12 and choosing 1 man out of 5. Since order does not matter and there is no replacement, we use combinations.

$= 12C_5 * 5C_1$

= 12! / 5! (12 – 5)! x 5

= 792 x 5

= 3960 combinations

b) Any combination of men and women means only one choice or category is made, people.

$17C_6$

= 17! /6! (17 – 6) !

= 12376 combinations

Problem 2.18 A bag contains 2 white balls, 3 black balls and 4 red balls. In how many ways can 3 balls be drawn from the bag, if at least one black ball is to be included in the draw?

Solution:

From 2 white balls, 3 black balls and 4 red balls, 3 balls are to be selected such that at least one black ball should be there. Hence, we have 3 choices as given below

Option 1

We can select 3 black balls

$3C_3$

Option 2

We can select 2 black balls and 1 non-black ball

$3C_2 \times 6C_1$

Option 3

We can select 1 black ball and 2 non-black balls

$3C_1 \times 6C_2$

$nC_n = 1$

$nC_1 = n$

Total number of ways

$= 3C_3 + (3C_2 \times 6C_1) + (3C_1 \times 6C_2)$

$= 1 + (3C_1 \times 6C_1) + (3C_1 \times 6C_2)$

=1+[3×6]+[3×15]

= 1 + 18 + 45

= 64

Ans: 64

Problem 2.19 From a group of 7 men and 6 women, five persons are to be selected to form a committee so that at least 3 men are there on the committee. In how many ways can it be done?

Solution:

We may have (3 men and 2 women) or (4 men and 1 woman) or (5 men only).

$\therefore$ Required number of ways $= (7C_3 \times 6C_2) + (7C_4 \times 6C_1) + (7C_5)$

Problem 2.20 Find the number of ways of choosing 4 cards from a pack of 52 playing cards. In how many of these

1. four cards are of the same suit,
2. four cards belong to four different suits,
3. face cards,
4. two are red and two are black cards,
5. Cards are of the same color?

Solution: There will be a number of possible ways for choosing 4 cards from 52 cards as there are combinations of 52 different things when we take 4 at a time.

Therefore, the required number of ways = $52C_4$

= 52! / (4! 48!)

= 48! x 49 x 50 x 51 x 52) / (2 x 3 x 4 x 48!)

= 270725

(1) Four cards of the same suit:

There are four suits: Spade, heart, Club, diamond. Totally, there are 13 cards of each suit

Therefore, the required numbers of ways are given by $13C_4 + 13C_4 + 13C_4 + 13C_4$

= 4(13! / (4! 9!))

= 2860

Ans: 2860

(2) Four cards belong to four different suits:

Since there are 13 cards in each suit. Therefore choosing 1 card from 13 cards of each suit, it becomes

$= 13C_1 + 13C_1 + 13C_1 + 13C_1$

$= 4 * (13C_1)$

= 52

Ans: 52

(3) Face cards:

There are 12 face cards and 4 cards are selected from these 12 cards, it becomes

$= 12C_4$

Therefore, the required number of ways = 12! / (4! 8!) = 495

Ans: 495

(4) Two red cards and two black cards:

There are 26 red and 26 black cards in a pack of 52 cards.

Therefore, the required number of ways = $26C_2 \times 26C_2$

= (325) *2
=650

Ans: 650

(5) Cards of the same color:

Out of 26 red cards and 26 black cards, 4 red and black cards are selected in $26C_4$ ways. So, the required number of ways = $26C_4 + 26C_4$
= 2 (26! / 4! 22!)
=2x 14950

Ans: 29900

2.4 Relationship between Permutations and Combinations

The concepts of and differences between permutations and combinations can be illustrated by examination of all the different ways in which a pair of objects can be selected from five distinguishable objects—such as the letters A, B, C, D, and E. If both the letters selected and the orders of selection are considered, then the following 20 outcomes are possible:

AB	BA	AC	CA	AD
DA	AE	EA	BC	CB
BD	DB	BE	EB	CD
DC	CE	EC	DE	ED

Each of these 20 different possible selections is called a permutation. They are called the permutations of five objects taken two at a time, and the number of such permutations possible is denoted by the symbol $5P_2$

$nP_r = (n!) / (n-r)!$
$5P2 = (5!) / (3)!$
$= 20$
For combinations, $nCr = (n!) /r ! (n-r)!$
$5C2 = 5! / (3)! 2!$
$= 10$

n objects are selected from a set of r objects to produce subsets without ordering. Contrasting the previous permutation example with the corresponding combination, the AB and BA subsets are no longer distinct selections; by eliminating such cases there remain only 10 different possible subsets
—AB, AC, AD, AE, BC, BD, BE, CD, CE, and DE.

In permutation and combination, the relationship between the two concepts is given by two theorems. They are

$$nC_r + nC_{r-1} = (n+1)C_r$$

$$nP_r = (r!) * nC_r 0 \leq r \leq n$$

The differences between the Permutation and combination are listed in the table 2.1

Table 2.1 Permutation Vs Combination

Permutation	Combination
Arranging people, digits, numbers, alphabets, letters, and colors	Selection of menu, food, clothes, subjects, team.
Picking a team captain, pitcher, and shortstop from a group.	Picking three team members from a group.
Picking two favorite colors, in order, from a color brochure.	Picking two colors from a color brochure.
Picking first, second and third place winners.	Picking three winners.

You simply have to choose 3 of the 8 teachers, and you know that you can do that in 8C3 ways. That is all that is required.**$8C_3 = (8*7*6)/(3*2*1) = 56$ ways**

There are 8 teachers in a school of which 3 need to give a presentation each. In how many ways can all three presentations be done? This question is a little different. You need to find the ways in which the presentations can be done. Here the presentations will be different if the same three teachers give presentations in different order. Say Teacher 1 presents, then Teacher 2 and finally Teacher 3 — this will be different from Teacher 2 presenting first, then Teacher 3 and finally Teacher 1. So, not only do we need to select the three teachers, but we also need to arrange them in an order. Select 3 teachers out of 8 in $8C_3$ ways and then arrange them in 3! ways:

$nP_r = (r!) * nC_r$

$8P_3 = 8C_3 * 3! = 56 * 6 = 336$ ways

Case Study 1: Relationship Study - nP_r and nC_r

2.5 Binomial Theorem

The Binomial Theorem is the method of expanding an expression which has been raised to any finite power. A binomial Theorem is a powerful tool of expansion, which has application in Algebra, probability, etc. Binomial Expression: A binomial expression is an algebraic expression which contains two dissimilar terms. Ex: a + b, $a^3 + b^3$, etc. A binomial is an algebraic expression containing 2 terms. For example, (x + y) is a binomial.

We sometimes need to expand binomials as follows:

$(a + b)^0 = 1$
$(a + b)^1 = a + b$
$(a + b)^2 = a^2 + 2ab + b^2$
$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$
$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$
$(a + b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$

Points to Remember

As the power increases the expansion becomes lengthy and tedious to calculate. A binomial expression that has been raised to a very large power can be easily calculated with the help of Binomial Theorem. Clearly, doing this by direct multiplication gets quite tedious and can be rather difficult for larger powers or more complicated expressions.

Pascal's Triangle

We note that the coefficients (the numbers in front of each term) following pattern and as shown in figure 2.1

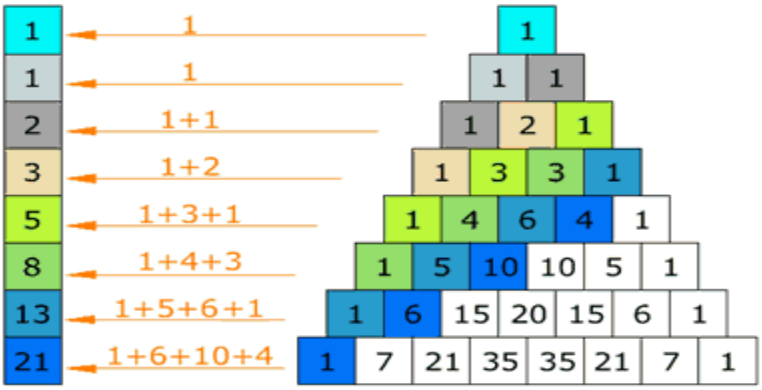


Fig. 2.1 Pascals' Triangle

We use the binomial theorem to help us expand binomials to any given power without direct multiplication. As we have seen, multiplication can be time-consuming or even not possible in some cases.

Properties of the Binomial Expansion $(a + b)^n$

[1]. There are n+1 term.

[2]. The first term is a^n and the final term is b^n.

[3]. Progressing from the first term to the last, the exponent of decreases by 1 from term to term while the exponent of b increases by 1. In addition, the sum of the exponents of a and b in each term is n.

[4]. If the coefficient of each term is multiplied by the exponent of a in that term, and the product is divided by the number of that term, we obtain the coefficient of the next term.

Let $n \in N, x, y, \in R$ then

i.e.$(x + y)^n = \Sigma\, {}^nC_r\, x^{n-r} \cdot y^r$ where, Limit r = 0 to r = n

$(x + y)^n = {}^nC_0\, x^n\, y^0 + {}^nC_1\, x^{n-1} .\, y^1 + {}^nC_2\, x^{n-2}. y^2 + \ldots + {}^nC_n\, y^n$

Problem 2.21: Expand $(x^2 + 3)^6$

Solution:

Students trying to do this expansion in their heads tend to mess up the powers. But this isn't the time to worry about that square on the x. I need to start my answer by plugging the terms and power into the Theorem. The first term in the binomial is "x^2", the second term in "3", and the power n is 6, so, counting from 0 to 6, the Binomial Theorem gives me:

$(x^2 + 3)^6 =$

$6C_0\,(x^2)^6(3)^0 + 6C_1\,(x^2)^5(3)^1 + 6C_2\,(x^2)^4(3)^2 + 6C_3\,(x^2)^3(3)^3 + 6C_4\,(x2)^2(3)^4 + 6C_5\,(x2)^1(3)^5 + 6C6\,(x^2)^0(3)^6$

Then simplifying gives me

$(1)(x^{12})(1) + (6)(x^{10})(3) + (15)(x^8)(9) + (20)(x^6)(27) + (15)(x^4)(81) + (6)(x^2)(243) + (1)(1)(729)$

$= x^{12} + 18\,x^{10} + 135x^8 + 540x^6 + 1215x^4 + 1458x^2 + 729$

Problem 2.22: Find the tenth term in the expansion of $(x + 3)^{12}$

Solution:

To find the tenth term, I plug x, 3, and 12 into the Binomial Theorem, using the number 10 – 1 = 9 as my counter:

$12C_9\,(x)^{12-9}(3)^9$

$= (220)\, x^3(19683)$

$= 4330260x^3$

Problem 2.23: What is the fourth term in the expansion of $(3x - 2)^{10}$

Solution:

To find the fourth term, I plug 3x, - 2, and 10 into the Binomial Theorem, using the number 4 – 1 = 3 as my counter:

$$10C_3\,(3x)^{10-3}(-2)^3$$

$$= (120)\,(2187)\,x^7(-8)$$

$$= -2099520x^7$$

Problem 2.24: Expand $(a + b)^3$

Solution:

The first term in the binomial is "a", the second term in "b", and the power n is 3, so, counting from 0 to 3, the Binomial Theorem gives me:

$$(a + b)^3 = 3C_0\,(a)^3(b)^0 + 3C_1\,(a)^2(b)^1 + 3C_2\,(a)^1(b)^2 + 3C_3\,(a)^0(b)^3$$

$$= (a)^3 + 3\,(a)^2(b)^1 + 3\,(a)^1(b)^2 + (b)^3$$

Problem 2.25: Expand $(a + b)^8$ What is the coefficient of a^4

Solution:

The first term in the binomial is "a", the second term in "b", and the power n is 8, so, counting from 0 to 3, the Binomial Theorem gives me:

$(a + b)^8 =$

$3C_0\,(a)^8(b)^0 + 3C_1\,(a)^7(b)^1 + 3C_2\,(a)^6(b)^2 + 3C_3\,(a)^5(b)^3 + 3C_4\,(a)^4(b)^4 + 3C_5\,(a)^3(b)^5 + 3C_6\,(a)^2(b)^6 + 3C_7\,(a)^1(b)^7 + 3C_8\,(a)^0(b)^8$

The coefficient of $a^4 = 3C_4\, b^4$

Problem 2.26: What is the sixth term in the expansion of $(7x + 2)^7$

Solution:

To find the fourth term, I plug 7x, 2, and 7 into the Binomial Theorem, using the number 6– 1 = 5 as my counter:

$$7C_5\,(7x)^{7-5}(2)^5$$

$$= (21)\,(49)\,x^2(32)$$

$$= 32928x^2$$

Problem 2.27

Find the number of permutations if n = 10 and r = 3?

Given Data:

n = 10

r = 3

Solution:

Permutation:

$nP_r = (n!) / (n-r)!$

$10P_3 = (10!) / (10-3)!$

$= 10! / 7!$

$= (10 \times 9 \times 8 \times 7!)/ 7!$

$= 720$

Ans: 720

Problem 2.28 Agriculture officer asks a former to choose 3 varieties of the seed from the box. If the box has 10 varieties, how many ways could the former choose the seed variety?

Given Data:

Here, student has to choose 3 variety from 10 items. Here, r= 3 and n= 10

Solution:

Combination,

$nC_r = (n!) /r ! (n-r)!$

$10C_3 = 10! / (10-3)! 3!$

Ans: 120

Summary of the Chapter

1. **Probability is a numerical description of how likely an event is to occur or how likely it is that a proposition is true. Probability is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility and 1 indicates certainty.**
2. **A permutation is defined as an arrangement in a definite order of a number of objects taken some or all at a time. Counting permutations are merely counting the number of ways in which some or all objects at a time are rearranged.**
3. **A permutation is the choice of r things from a set of n things without replacement and where the order matters.** $nP_r = (n!) / (n-r)!$

4. **Cor. number of all permutations of n things, taken all at a time = n!**
5. **The number of permutations of different objects "n" taken "r" at a time, where repetition is allowed and is given by n^r**
6. **The combination is a way of selecting items from a collection, such that (unlike permutations) the order of selection does not matter.**
7. **The number of combinations of n objects taken r at a time and the combination formula is given by $nC_r = n!/r!(n-r)!$**

Model Questions

Part A

[1]. Define: Permutation
[2]. List out the difference between Permutation and Combinations
[3]. Remember: Combinations
[4]. Draw the Pascals' Triangle
[5]. Write the properties of Binomial Expansion $(a + b)^n$
[6]. Write down the formula for nC_r
[7]. Write down the formula for nP_r
[8]. Draw the relationship between the nC_r and nP_r
[9]. Write down the formula for Binomial Expansion

Part B

[1]. Evaluate: 9! / 5!
[2]. It is needed to seat 3 boys and 7 girls in a row so that the girl gets the even places. How many such arrangements are possible?
[3]. How many words can be formed using all the letters of the word Analytics, using each letter exactly one time?
[4]. Evaluate $10C_3$
[5]. Teacher asks a student to choose 7 items from the table. If the table has 25 items to choose, how many ways could the students choose the things?
[6]. In how many ways a committee consisting of 3 engineers and 3 doctors, can be chosen from 7 engineers and 8 doctors.
[7]. Find the tenth term in the expansion of $(x + 8)^{12}$
[8]. Expand $(x^2 + 3)^7$
[9]. What is the coefficient for x^3 in $(2x+4)^8$

Chapter 3 Data and Data Visualization

Introduction

Data is the raw form of information, at least the way we use the words “data” and “information”. Information is something that useful to us, aiding the decision-making process. Data is where information comes from, but the important part is **how** to turn data into information?

Transforming data into information is a challenge to all professionals working in the ***rural and agriculture arena***. Reams of data do not provide the answers being sought to challenging questions. Several tools including ***charts, graphs and tables*** can be used to visually display data and provide further ideas for analysis. This chapter describes in detail when to use these tools and how to construct and use them.

Objectives

- To understand the Data into Information for decision makers
- To visualize the Categorical/ Quantitative data
- To understand the importance and elements of data visualization
- To become familiar with different types of charts

Chapter Structure

3.1 Data into Information

3.2 Different Types of Data

3.3 Scales of Measurements

3.4 Introduction to Data Visualization

3.5 Importance of Data Visualization

3.6 Line Chart, Pie chart, Bar Chart and Histogram

"

Data! Data! Data! I can't make bricks without clay"

— Sherlock Holmes, in Arthur Conan Doyle's The Adventure of the Copper Beeches

3.1 Data into Information

Data is a collective name for information recorded for statistical purposes. Data is defined as a systematic record of a particular quantity. Data is a collection of facts such as numbers, words, measurements, observations or even just descriptions of things. It is a collection of facts and figures to be used for a specific purpose such as a survey or analysis. When arranged in an organized form, can be called information.

Data is raw, unorganized facts that need to be processed. Data can be something simple and seemingly random and useless until it is organized. When data is processed, organized, structured or presented in a given context so as to make it useful, it is called information.

Data are the facts or details from which information is derived. Individual pieces of data are rarely useful alone.

The following table 3.1 is shown very clear picture about the difference between data and information

Table 3.1 Data Vs Information

Nature	Data	Information
Definition	Data means raw facts gathered about someone or something, which is bare and random.	When data is processed, organized, structured or presented in a given context so as to make it useful, it is called information
Based on	Records and observation	Analysis
Form	Unorganized	Organized
Usefulness	May or May be	Yes
Dependency	Does not depend on information	Without data, information cannot be processes

Data and Information
Each year rain fall data for Hyderabad is one piece of data
The average rain fall for Hyderabad is information that can be derived from the given data

Data and Information
Each Tree height in a Plant is one piece of data
The average height of the tree in a plant is information that can be derived from the given data

3.2 Different types of Data

Data Types are an important concept of statistics and business analytics which needs to be understood, to correctly apply statistical measurements to your data and therefore to correctly conclude certain assumptions about it.

According to source of data, data are classified into two types. Table 3.2 represents different types of data and its nature.

Table 3.2 Primary Data Vs Secondary Data

Nature	Primary Data	Secondary Data
Definition	Primary data are collected for the first time by an investigator for a specific purpose	Secondary data are the data that are sourced from someplace that has originally collected it. This means that this kind of data has already been collected by some researchers or investigators in the past and is available either in published or unpublished form.
Statistical Operation	Primary data are 'pure' in the sense that no statistical operations have been performed in this data	statistical operations may have been performed on them already
Example	Census of India	Ministry of Rural Development and Ministry of Agriculture Statistics Handbook Data

Data Portal – Government of India

Open Government Data (OGD) Platform India - data.gov.in - is a platform for supporting Open Data initiative of Government of India. The portal is intended to be used by Government of India Ministries/ Departments their organizations to publish datasets, documents, services, tools and applications collected by them for public use. It intends to increase transparency in the functioning of Government and also open avenues for many more innovative uses of Government Data to give different perspective. The base Open Government Data Platform India is a joint initiative of Government of India and US Government. Open Government Data Platform India is also packaged as a product and made available in open source for implementation by countries globally. The entire product is available for download at the Open Source Code Sharing Platform "GitHub".

Points to Remember

At high end, Data can be classified into qualitative or quantitative.

Qualitative Data

Qualitative data is descriptive information (it describes something). Qualitative data deals with characteristics and descriptors that can't be easily measured, but can be observed subjectively—such as smells, tastes, textures, attractiveness, and color.

Quantitative

Quantitative data is numerical information (numbers). Quantitative data deals with numbers and things you can measure objectively: dimensions such as height, width, length, temperature, humidity, prices, area and volume. There are two types of quantitative data, which is also referred to continuous data and discrete data.

Discrete Data

Discrete data can only take certain values (like whole numbers). Discrete data is numerical data that cannot be shown in decimals. Discrete data is a count that can't be made more precise. Typically, it involves integers. For instance, the number of trees in your plant is discrete data, because you are counting whole, indivisible entities: you can't have 2.5 coconut tree, or 1.3 bananas.

Continuous data

Continuous data is measurable numerical data that can be shown in decimals and could be divided and reduced to finer and finer levels. For example, height of the coconut tree is 250.35 cm. For example, we can measure the height of the tree at progressively more precise scales—meters, centimeters, millimeters, and beyond—so height is continuous data. The different types of data are represented in figure 3.1

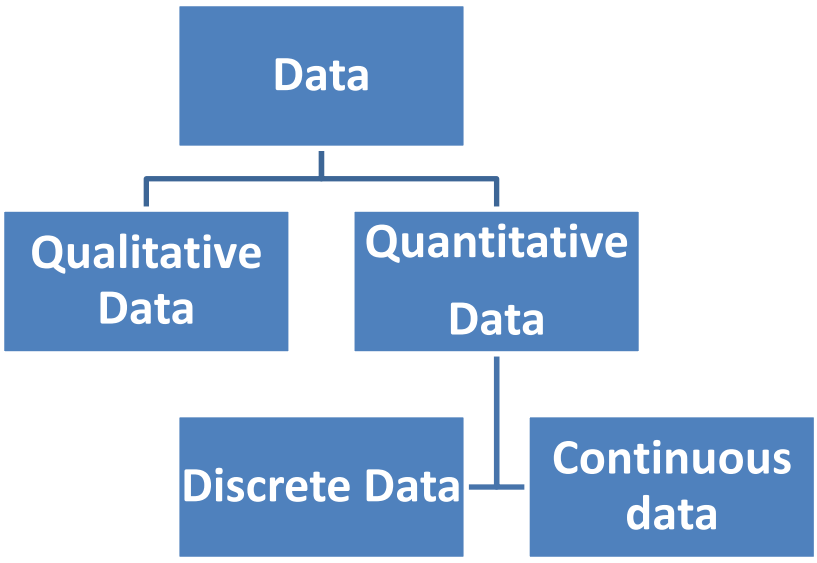


Fig. 3.1 Different Types of Data

Difference between Discrete Data and Continuous Data

Discrete data involves round, concrete numbers that are determined by counting. Continuous data involves complex numbers that are measured across a specific time interval.

What do we know about Banana

Types of Data in Banana

Source : https://www.kadalys.com/en/banana-tree/

Banana (Musa sp.) is the second most important fruit crop in India next to mango. Its yearround availability, affordability, varietal range, taste, nutritive and medicinal value makes it the favorite fruit among all classes of people. It has also good export potential.

Scales of Measurements

Qualitative Data:	Quantitative Data:	Quantitative Data:
It is Poovan Variety It has long leaf	**Discrete Data:** The tree has 10 leaf and 5 fruits	**Continuous Data:** The height of the banana Plant 153.2 cm Banana Yield is 45.4 (tones/ha.) (Flood/furrow irrigation)

3.3 Scales of Measurement

Different types of Scales of Measurement

There are four different types of scales of measurement in the data aspects. The listed different types of scales of measurement and examples are represented in table 3.3.

1. **Nominal Scale**

   Nominal variables (also called categorical variables) can be placed into categories. They don't have a numeric value and so cannot be added, subtracted, divided or multiplied. They also have no order; if they appear to have an order then you probably have ordinal variables instead.

2. **Ordinal Scale**
   The ordinal scale contains things that you can place in order. Ordinary scale values have both an identity and a magnitude.
3. **Interval Scale**
   An interval scale has ordered numbers with meaningful divisions. An interval scale is one where there is order and the difference between two values is meaningful.
4. **Ratio Scale**
   The ratio scale is exactly the same as the interval scale with one major difference: zero is meaningful. For example, a height of zero is meaningful (it means you don't exist). Different types of scale measurements and representation are represented in the figure 3.2

Table 3.3 Types of Scale of Measurement and Examples

Nominal Scale blood type, gender, race, eye color	**Ordinal Scale** For example, hottest to coldest, lightest to heaviest, richest to poorest. Basically, if you can rank data by 1st, 2nd, 3rd place (and so on), then you have data that's on an ordinal scale. socio economic status ("low income", "middleincome", "high income"), education level ("high school","BBA","MBA,"PhD"). income level ("less than 50K", "50K-100K", "over 100K"), satisfaction rating ("extremely dislike", "dislike", "neutral", "like", "extremely like").
Ratio Scale enzyme activity, dose amount, reaction rate, flow rate, concentration, pulse, weight, length, temperature	**Interval Scale** Temperature is on the interval scale a difference of 10 degrees between 90 and 100 means the same as 10 degrees between 150 and 160. Compare that to high school ranking (which is ordinal), where the difference between 1^{st} and 2^{nd} might be .01 and between 10^{th} and 11^{th} .5. If you have meaningful divisions, you have something on the interval scale.

Scale of Measurement	Nominal Scale	Ordinal Scale	Interval scale	Ratio scale
Identity	√	√	√	√
Magnitude		√	√	√
Equal Intervals			√	√
Minimum Value as Zero				√

Fig. 3.2. Scale of Measurement and Representation

3.4 Introduction to Data Visualization

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and

patterns in data. In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.

Data visualization is the graphic representation of data.

Data visualization is the act of taking information (data) and placing it into a visual context, such as a map or graph.

Common general types of data visualization:

- ✓ Charts
- ✓ Tables
- ✓ Graphs
- ✓ Maps
- ✓ Infographics
- ✓ Dashboards

A **table** contains quantitative data organized into rows and columns with categorical labels. It is primarily used to look up specific values. In the example above, the table might have categorical column labels representing the name (a qualitative variable) and age (a quantitative variable), with each row of data representing one person (the sampled experimental unit or category subdivision).

A **graph** is primarily used to show relationships among data and portrays values encoded as visual objects (e.g., lines, bars, or points). Numerical values are displayed within an area delineated by one or more axes. These axes provide scales (quantitative and categorical) used to label and assign values to the visual objects. Many graphs are also referred to as charts.

3.5 Importance of Data Visualization

Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

Data visualization can also:

- Identify areas that need attention or improvement.
- Clarify which factors influence customer behavior.
- Help you understand which products to place where.
- Predict sales volumes.

A picture is worth a thousand words – especially when you are trying to find relationships and understand your data – which could include thousands or even millions of variables. To create meaningful visuals of your data, there are

some basic tips and techniques you should consider. Data size and composition play an important role when selecting graphs to represent your data.

Comprehend Information Quickly

By using graphical representations of business information, businesses can see large amounts of data in clear, cohesive ways – and draw conclusions from that information. And since it's significantly faster to analyze information in graphical format (as opposed to analyzing information in spreadsheets), businesses can address problems or answer questions in a more timely manner.

Identify Relationships and Patterns

Even extensive amounts of complicated data start to make sense when presented graphically; businesses can recognize parameters that are highly correlated. Some of the correlations will be obvious, but others won't. Identifying those relationships helps organizations focus on areas most likely to influence their most important goals.

Pinpoint Emerging Trends

Using data visualization to discover trends – both in the business and in the market – can give businesses an edge over the competition, and ultimately affect the bottom line. It's easy to spot outliers that affect product quality or customer churn, and address issues before they become bigger problems.

Communicate the Story to Others

Once a business has uncovered new insights from visual analytics, the next step is to communicate those insights to others. Using charts, graphs or other visually impactful representations of data is important in this step because it's engaging and gets the message across quickly.

According to Colin Ware's Information Visualization: Perception for Design, he defines four pre-attentive visual properties:

- Color
- Form
- Movement
- Spatial positioning

These four components make up the composition of each data visualization and should be carefully considered for presentation.

Points to Remember

3.6 Different Types of Chart

The following charts and examples of methods to visualize data:

Line Chart

Line charts are used to show trend of a measure (or a variable) over time. A line chart or line plot or line graph or curve chart is a type of chart which displays information as a series of data points called 'markers'

connected by straight line segments. Line chart displays series of data points connected by straight line segments. Line graphs are often used to display time series chronologically with category axis (usually horizontal x-axis) serving as an evenly spaced date-time scale.

Case Study – Practice Class

Practice all the graphs/charts (problems) in chapter 3 and 4 in Excel Spread Sheet

In Excel Spread sheet ⟶ Select the Entered Numerical Data ⟶ Select Insert - Charts

Ref Book :Bill Jelen (2010). Charts and Graphs: Microsoft Excel 2010 (MrExcel Library), Que Publishing

- Step 1: Identify the variables
- Step 2: Determine the variable range
- Step 3: The scale of the graph
- Step 4: Number and label each axis and title the graph
- Step 5: Determine the data points and plot on the graph
- Step 6: Draw the graph
- Step 7: Inference

STEPS FOR LINE CHART

Go Through, Ministry of Rural Development
Website https://rural.nic.in/ and refer
Key Data on Rural Development from IDFC India Rural Development Report
Workout state wise share of non-agriculture in total employment rural (%) in the year 2009-10 with help of line chart

Practice

Problem 3.1: Draw the line chart for umbrella sales with respect to the Months. What you infer from the line chart

Month	Umbrella Sales (Nos)	Month	Umbrella Sales (Nos)
Jan	2000	July	6000
Feb	4500	August	4000
March	5000	September	6000
April	6500	October	5000
May	6000	November	7500
June	4000	December	8000

Solution: X axis – Month; Y axis - umbrella Sales (Nos); Variable Range - umbrella Sales (Nos) – 0 – 9000; Scale of the Graph – 1000

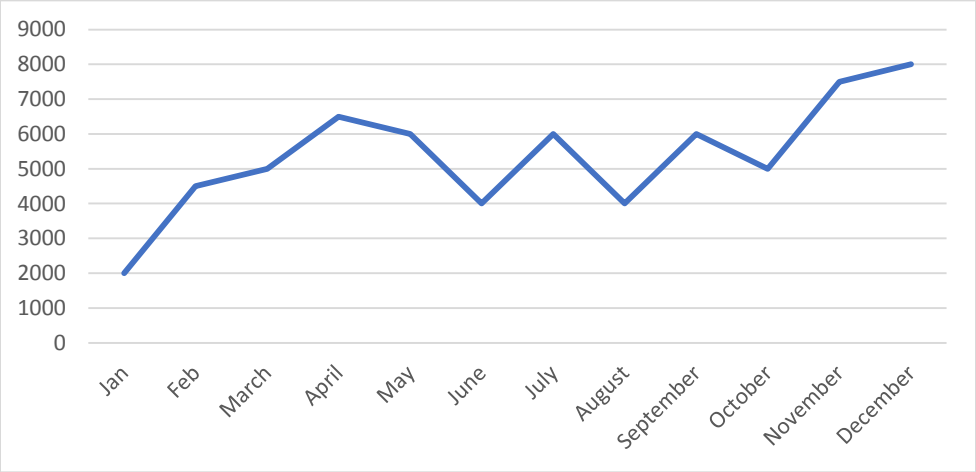


Fig. 3.3 Line Chart for Umbrella Sales

Interference from the Line Chart figure 3.3: November and December month Umbrella Sales are in Peak

Problem 3.2:

Draw the line chart for wind speed at different height at 50 m, 32 m, 16 m, 8 m and 2 m from the ground level at different time period. What you infer from the line chart

Solution: X axis – Time; Y axis – Wind speed (m/sec); Variable Range - Wind speed (m/sec) – 0 – 6 m/sec; Scale of the Graph – 1

Time/Wind Speed (m/sec)	@2011WS50	@2011WS32	@2011WS16	@2011WS08	@2011WS02
12:00:00 AM	5.1	4.4	3.6	2.7	1.8
12:09:56 AM	5	4.3	3.5	2.6	1.5
12:20:01 AM	4	3.3	2.7	1.9	1.1
12:29:57 AM	4.3	3.6	3.2	2.5	1.6
12:40:02 AM	4.4	3.7	3.4	2.7	1.8
12:49:58 AM	4.8	4.1	3.8	3.1	2.1
1:00:03 AM	4.4	3.7	3.4	2.7	1.7
1:09:59 AM	3.7	3	2.6	1.9	1
1:20:04 AM	4.1	3.4	3	2.3	1.5
1:30:00 AM	3.8	3.1	2.5	1.7	0.9
1:39:56 AM	3.8	3.1	2.8	1.8	1
1:50:01 AM	3.1	2.4	1.9	1.1	0.5
1:59:57 AM	3.6	2.9	2.5	1.7	0.7
2:10:02 AM	4.3	3.6	3.3	2.5	1.6
2:19:58 AM	3.8	3.1	3	2.4	1.5

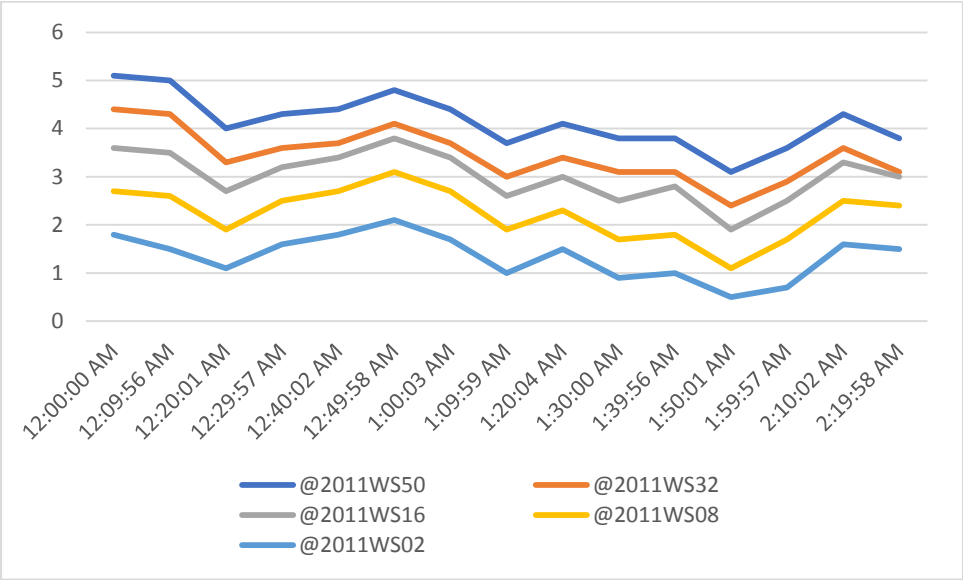


Fig. 3.4 Line Chart for Wind Speed

Interference from the Line Chart figure 3.4: Measurement Height for wind speed is major factor. Wind speed is not same at level; If level (height) increase; the wind speed will also increase

Bar Chart

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column chart. The classic Bar Chart uses either horizontal or vertical bars (column chart) to show discrete, numerical comparisons across categories. One axis of the chart shows the specific categories being compared and the other axis represents a discrete value scale.

Examples of Bar Chart

Comparison of values, such as sales performance for several persons or businesses in a single time period. For a single variable measured over time (trend) a line chart is preferable.

Problem 3.3:

A survey of 145 people asked them "Which is the nicest fruit? Draw the bar chart and conclude:

Fruit:	Apple	Orange	Banana	Kiwifruit	Blueberry	Grapes
People:	35	30	10	25	40	5

Steps in the Bar Chart Process

[1]. Decide on a title for your graph (from the table).
[2]. Draw the vertical and horizontal axes.
[3]. Label the horizontal axes (Types of fruit).
[4]. Write the names of pets where the bars will be (Apple, Orange, Banana and so on).
[5]. Label the vertical axes (Number of people).
[6]. Decide on the scale. Explain that you should consider the least and the greatest number shown on the graph. Discuss what range of numbers should be shown on this bar graph (Begin at 0 and count by 5 to 50).
[7]. Draw a bar to show the total for each item.

Solution: X axis – Fruits; Y axis – People No; Variable Range - People No (Nos) – 0 – 45; Scale of the Graph – 5

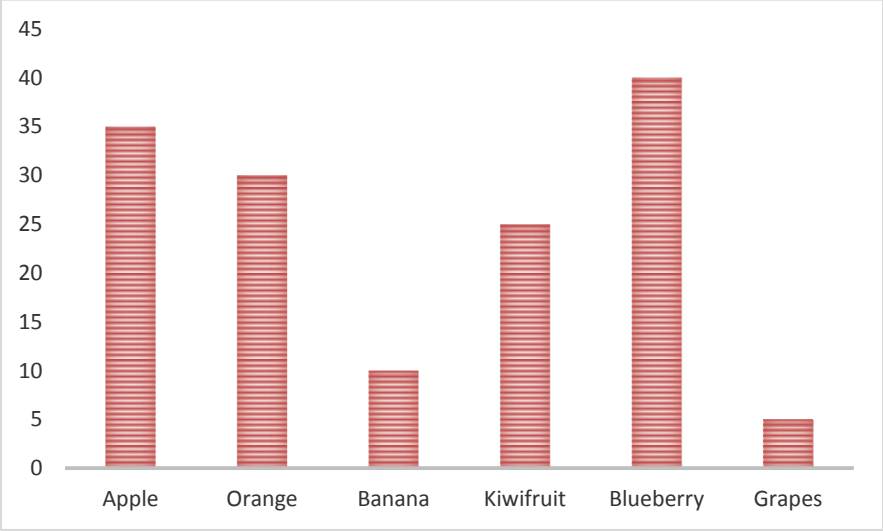


Fig. 3.5a. Bar chart for Fruits Survey (Vertical)

Bar Graphs can also be Horizontal, like this:

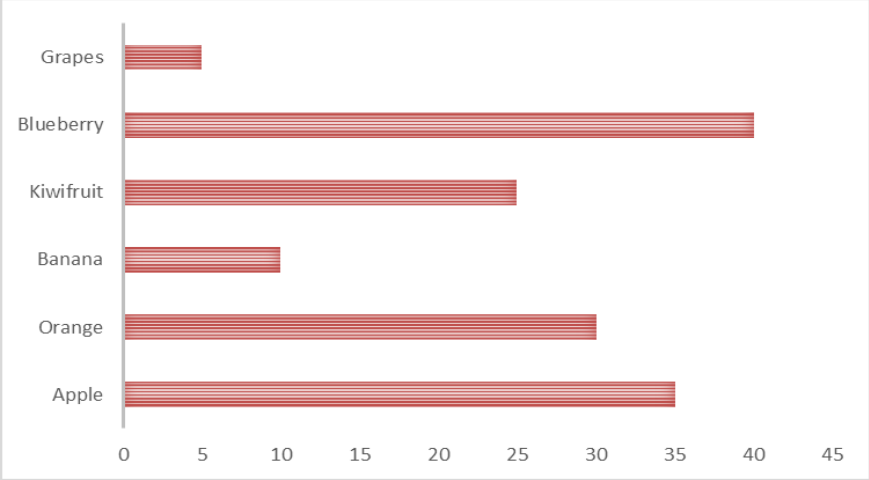
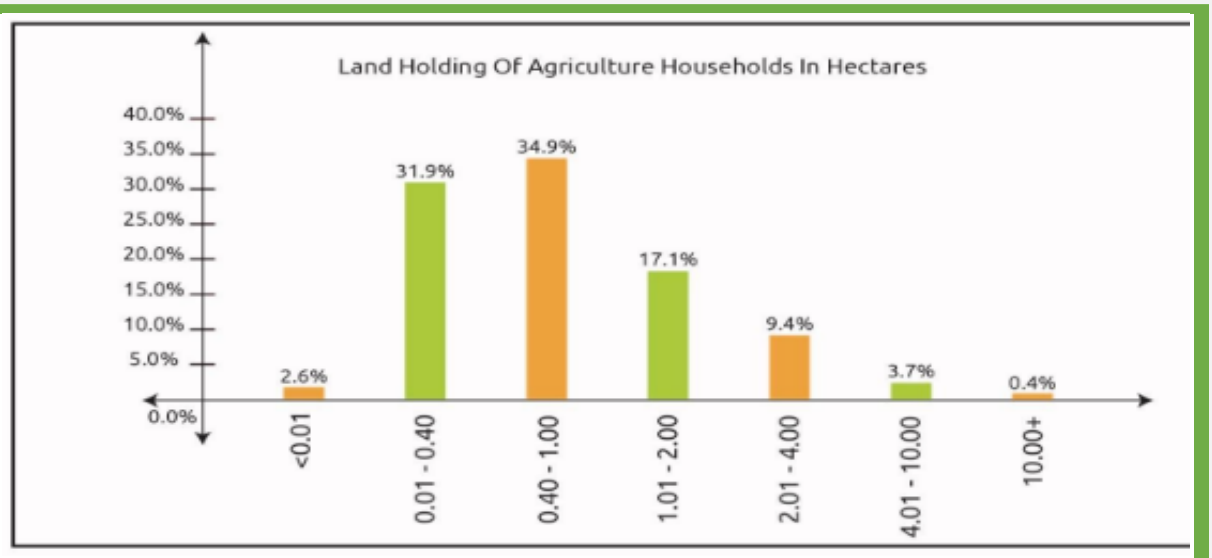


Fig. 3.5b. Bar chart for Fruits Survey (Horizontal)

Interference from the Bar Chart figure 3.5 a and 3.5 b: Those groups of people think Blueberries are the nicest.



Source: https://www.saddahaq.com/58-of-the-rural-households-are-agricultural-households-nsso-report-on-agricultural-households-part-1

[1]. What you infer from the bar chart/graph.

[2]. Produce chart to table

Problem 3.4:

The cropping pattern in Tamil Nadu in the year 1974-75 was as follows. Draw the Bar Chart

Crops	Area in 1,000 hectares
Cereals	3940
Oilseeds	1165
Pulses	464
Cotton	249
Others	822

Solution: X axis – Crops; Y axis – Area; Variable Range - Area – 0 – 4500; Scale of the Graph – 500

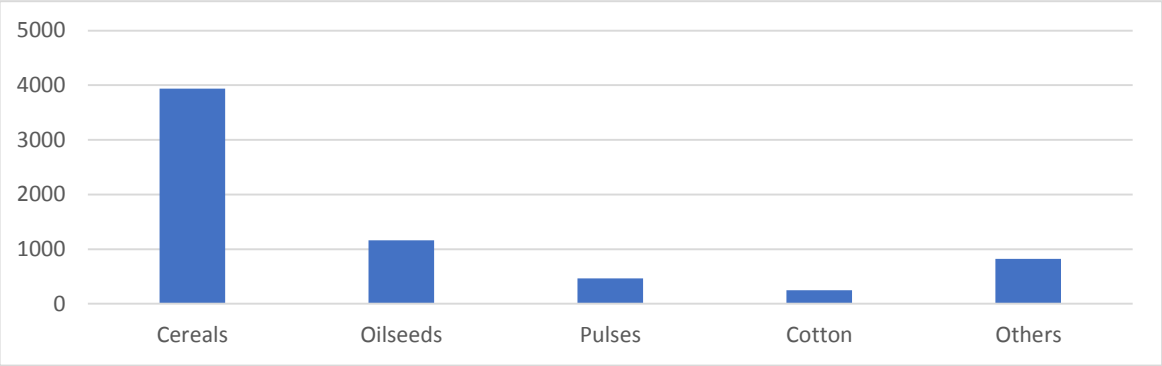
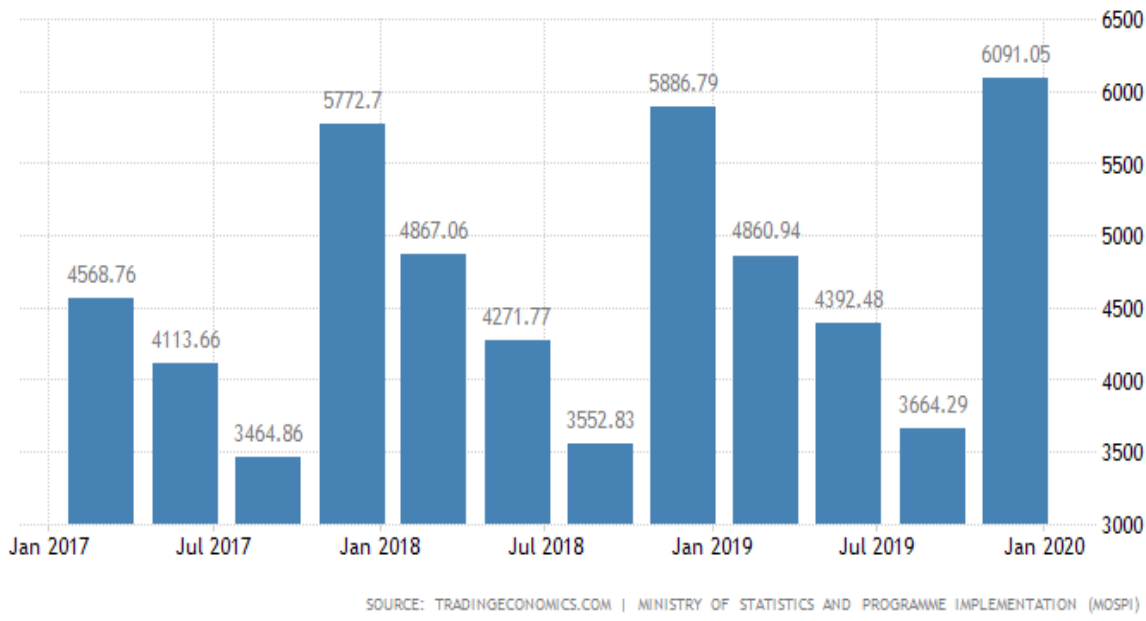


Fig. 3.6. Bar chart for Cropping Pattern

Interference from the Bar Chart figure 3.6: Cereals are major produce in Tamil Nadu compared to oilseeds, pulses and cotton according to the cropping pattern in the year 1974-75



[1]. What you infer from the bar chart/graph.

[2]. Produce: Chart to Table

Case Study 2: Practice

Multiple bar Diagram

If the data is classified by attributes and if two or more characters or groups are to be compared within each attribute, we use multiple bar diagrams. If only two characters are to be compared within each attribute, then the

resultant bar diagram used is known as **double bar diagram**.The multiple bar diagram is simply the extension of simple bar diagram. For each attribute two or more bars representing separate characters or groups are to be placed side by side. **Component Bar Diagram** This is also called sub – divided bar diagram. Instead of placing the bars for each component side by side we may place these one on top of the other. This will result in a component bar diagram.

Problem 3.5:

Draw a multiple bar diagram for the following data which represented agricultural production for the period from 2000-2003

Solution: X axis – Year (2000 – 2003) ; Y axis – Agricultural Production; Variable Range - Agricultural Production – 0 – 160; Scale of the Graph – 20

Year	Food grains (tones)	Vegetables (tones)	Others (tones)
2000	150	30	10
2001	110	40	35
2002	100	35	25
2003	120	50	28

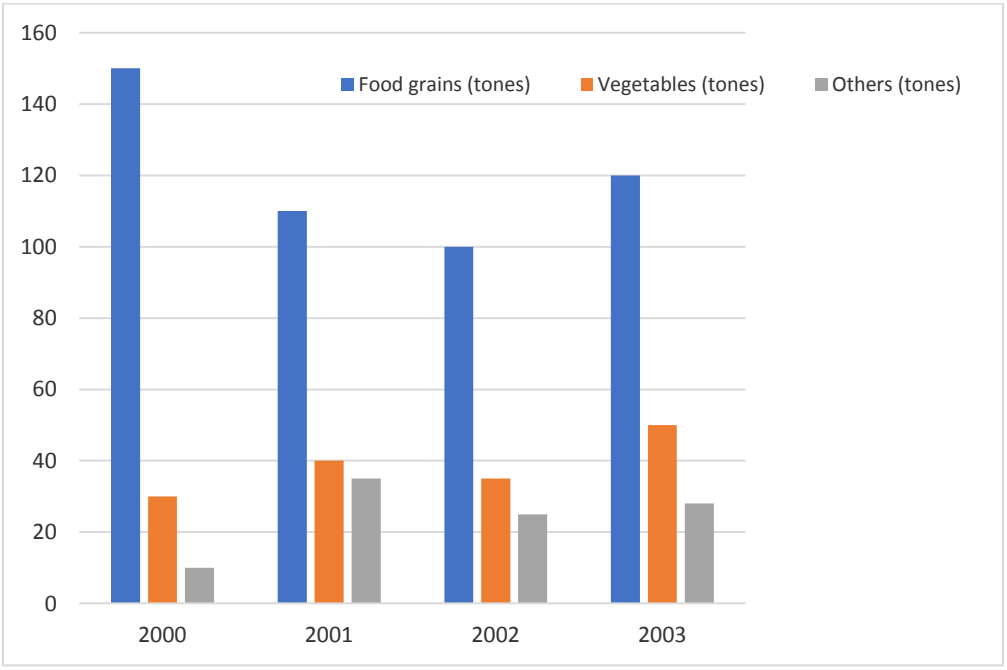


Fig. 3.7. Multiple Bar Chart for Cropping Pattern

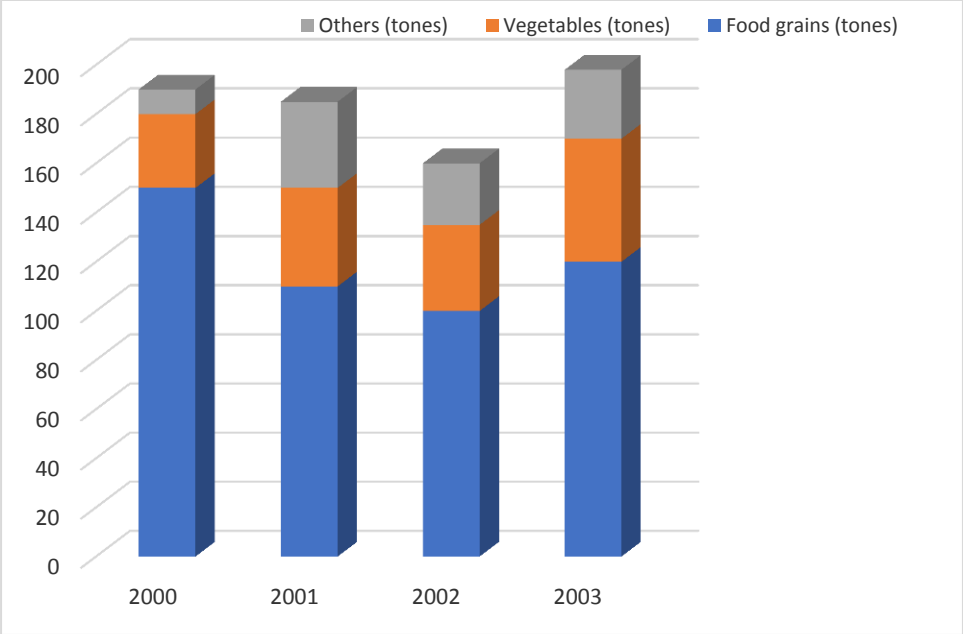


Fig. 3.8. Component Bar chart for Cropping Pattern

Interference from the Multiple Bar Chart/ Component Bar Diagram figure 3.7 and figure 3.8: Food grains (tones) are the most, followed by Vegetables (tones)

The cropping pattern of Tamil Nadu in 3 different years was as follows.

	Area		
Crops	2001	2002	2003
Cereals	3600	3650	3950
Oilseeds	1000	1150	1100
Pulses	400	450	460
Cotton	200	230	240
Others	800	820	820

[1]. Draw multiple bar diagram, Component bar diagram, Percentage bar diagram and Pie chart.

[2]. What do infer from the chart

To Do Activity

Pie chart

Pie chart is a special chart that uses "pie slices" to show relative sizes of data. Pie charts are interesting graph visualization. At a high-level, they're easy to read and understand because the parts-of-a-whole relationship is made obvious. But top data visual experts agree that one of their disadvantages is that the percentage of each section isn't obvious without adding numerical values to each slice of the pie.

Problem 3.6:

Imagine you survey your friends to find the kind of movie they like best:

Favorite Type of Movie				
Comedy	Action	Romance	Drama	SciFi
4	5	6	1	4

Draw the pie chart and conclude:

Solution: Step 1:

First, put your data into a table (like above), then add up all the values to get a total:

Favorite Type of Movie					
Comedy	Action	Romance	Drama	SciFi	TOTAL
4	5	6	1	4	20

Step 2:

Next, divide each value by the total and multiply by 100 to get a percent:

Comedy	Action	Romance	Drama	SciFi	TOTAL
4	5	6	1	4	20
4/20 = 20%	5/20 = 25%	6/20 = 30%	1/20 = 5%	4/20 = 20%	100%

Step 3:

Now to figure out how many degrees for each "pie slice" (correctly called a sector). A Full Circle has 360 degrees, so we do this calculation:

Comedy	Action	Romance	Drama	SciFi	TOTAL
4	5	6	1	4	20
20%	25%	30%	5%	20%	100%
4/20 × 360° = 72°	5/20 × 360° = 90°	6/20 × 360° = 108°	1/20 × 360° = 18°	4/20 × 360° = 72°	360°

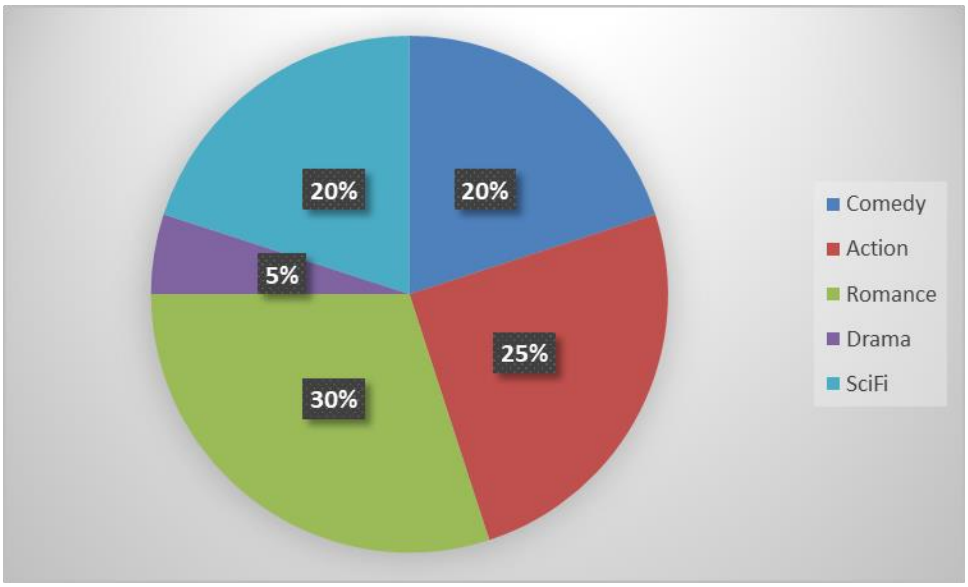


Fig. 3.9. Pie chart for Favorite Types of Movie

Interference from the Pie Chart figure 3.9: It is a really good way to show relative sizes: it is easy to see which movie types are most liked, and which are least liked, at a glance.

Problem 3.7:

The pie-chart given in figure.3.10shows the annual agricultural production of an Indian state. If the total production of all the commodities is 81000 tonnes, find the production (in tonnes) of (i) Wheat (ii) Sugar (iii)Rice (iv)Maize (v)Gram

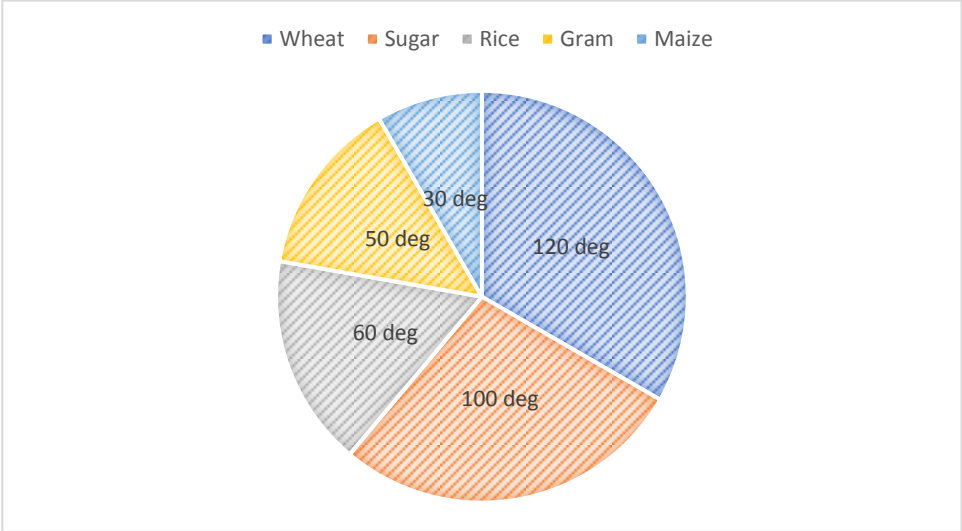


Fig. 3.10. Pie chart for Annual Agriculture Production

Solution: Total Production = 81000 Tonnes.

1) Production of wheat

= Central angle for wheat x Total Production
  360°

= 120°x 8100
  360°

= 27000 tonnes

2) Production of sugar

= Central angle for sugar x Total Production
  360°

= 100°x 8100
  360°

= 22500 tonnes

3) Production of rice

= Central angle for Rice x Total Production
  360°

= 60°x 8100
  360°

=13500 tonnes

4) Production of maize

= Central angle for Maize x Total Production
  360°

= 30°x 8100
  360°

=6750 tonnes

5) Production of rice

= Central angle for gram x Total Production
  360°

= 50°x 8100
  360°

=11250 tonnes

To Do Activity:

Causes of worldwide land degradation are listed. Below is a frequency table of the results:

Overgrazing	Over cultivation	Deforestation	other
35	28	30	7

Draw the Pie Chart

Problem 8: Given the cultivable land area in four southern states of India. Construct a pie diagram for the following data.

Cultivable area (in hectares)				
Andhra Pradesh	Karnataka	Kerala	Tamil Nadu	Total
663	448	290	556	1957

Solution:

To Calculate Percentage				
Andhra Pradesh	Karnataka	Kerala	Tamil Nadu	Total
(663/1957) * 100 = 34 %	(448/1957) * 100 = 23 %	(290/1957) * 100 = 15 %	(556/1957) * 100 = 28 %	100

Degree for Each				
Andhra Pradesh	Karnataka	Kerala	Tamil Nadu	Total
(663/1957) * 360 = 122 deg	(448/1957) * 360 = 83 deg	(290/1957) * 360 = 53 deg	(556/1957) * 360 = 102 deg	360

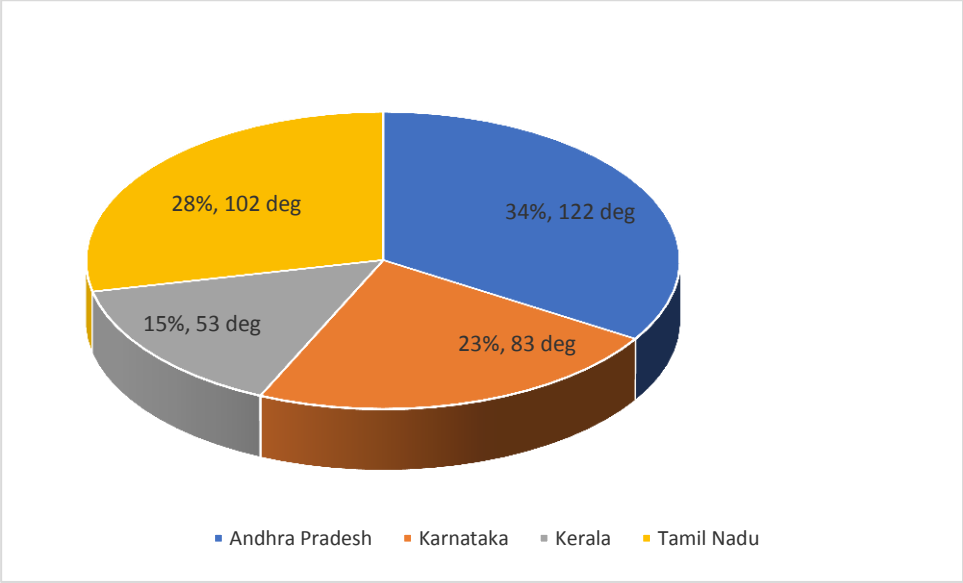


Fig. 3.11. Pie chart for Cultivable Land Area in Four Southern States of India

Interference from the Pie Chart figure 3.11: In southern states of India, Andhra Pradesh and Tamil Nadu have most Cultivable Land Area.

To Do Activity:

What you infer from the chart

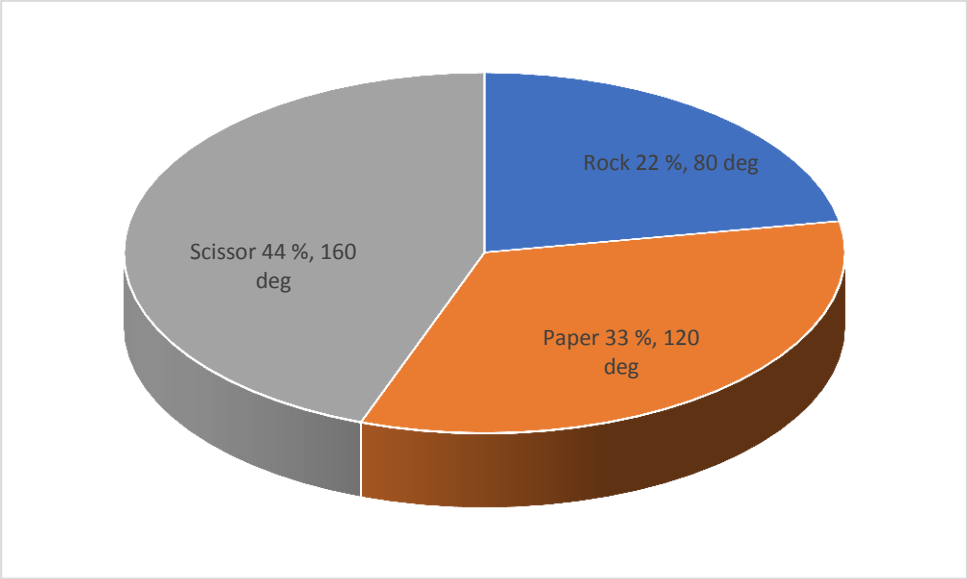


Fig. 3.12 Pie chart for Outcomes for Paper, Scissor and Rock Games

Data			
Rock	Paper	Scissor	Total
2	3	4	9

To Calculate Percentage			
Rock	Paper	Scissor	Total
(2/9) * 100 = 22 %	(3/9) * 100 = 33 %	(4/9) * 100 = 44 %	100 %

Degree for Each			
Rock	Paper	Scissor	Total
(2/9) * 360 = 80 deg	(3/9) *100 = 120 %	(4/9) * 360 = 160 deg	360 deg

Histogram

A histogram is a data visualization that shows the distribution of data over a continuous interval or certain time period. It's basically a combination of a vertical bar chart and a line chart. The continuous variable shown on the X-axis is broken into discrete intervals and the number of data you have in that discrete interval determines the height of the bar. Histograms give an estimate as to where values are concentrated, what the extremes are and whether there are any gaps or unusual values throughout your data set. It differs from a bar chart, in the sense that a bar graph relates two variables, but a histogram relates only one.

Histogram is a plot that lets you discover, and show, the underlying frequency distribution (shape) of a set of continuous data.

Problem 3.9:

The following is a list of prices (in rupees) of apple seeds found in various organic stores. Draw the Histogram

1.45	2.20	0.75	1.23	1.25
1.25	3.09	1.99	2.00	0.78
1.32	2.25	3.15	3.85	0.52
0.99	1.38	1.75	1.22	1.75

Solution: Frequency Table

The values go from 0.52 to 3.85, which is roughly 0.50 to 4.00. We can divide this into 7 intervals of equal length: 0.50 - 0.99, 1.00 - 1.49, 1.50 - 1.99, 2.00 - 2.49, 2.50 - 2.99, 3.00 - 3.49, and 3.50 - 3.99. Then we can count the number of data points which fall into each interval--for example, 4 points fall into the first interval: 0.75, 0.78, 0.55, and 0.99--and make a frequency distribution table:

Intervals (in rupees)	Frequency
0.50 - 0.99	4
1.00 - 1.49	7
1.50 - 1.99	3
2.00 - 2.49	3
2.50 - 2.99	0
3.00 - 3.49	2
3.50 - 3.99	1
Total	**20**

Making a Histogram from the Frequency Distribution Table

To make a histogram, follow these steps:

- On the vertical axis, place frequencies. Label this axis "Frequency".
- On the horizontal axis, place the lower value of each interval. Label this axis with the type of data shown (price of apple seeds.)
- Draw a bar extending from the lower value of each interval to the lower value of the next interval. The height of each bar should be equal to the frequency of its corresponding interval.

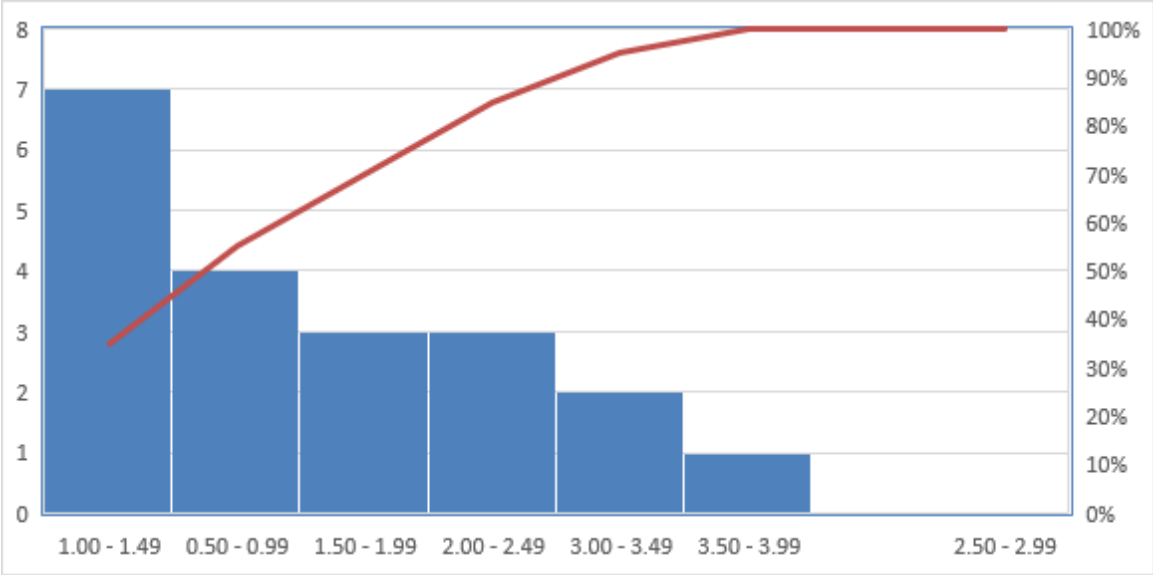


Fig. 3.13. Histogram for the Price of Apple Seeds

Interference from the Histogram figure 3.13: it is easy to Histograms are useful because they allow us to glean certain information at a glance. The previous example shows that more seed cost between Rs 1.00 and Rs 1.49 than any other price, because the bar which corresponds to those values is highest.

Bars Charts are distinguished from Histograms, as they do not display continuous developments over an interval. Bar Chart's discrete data is categorical data and therefore answers the question of "how many?" in each category.

Points to Remember

Problem 3.10:

To construct a histogram from a continuous variable you first need to split the data into intervals, called bins. In the example above, age has been split into bins, with each bin representing a 10-year period starting at 20 years. Each bin contains the number of occurrences of scores in the data set that are contained within that bin. For the above data set, the frequencies in each bin have been tabulated along with the scores that contributed to the frequency in each bin (see below):

Bin	Frequency	Scores Included in Bin
20-30	2	25,22
30-40	4	36,38,36,38
40-50	4	46,45,48,46

50-60	5	55,55,52,58,55
60-70	3	68,67,61
70-80	1	72
80-90	0	-
90-100	1	91

Solution: Notice that, unlike a bar chart, there are no "gaps" between the bars (although some bars might be "absent" reflecting no frequencies). This is because a histogram represents a continuous data set, and as such, there are no gaps in the data (although you will have to decide whether you round up or round down scores on the boundaries of bins).

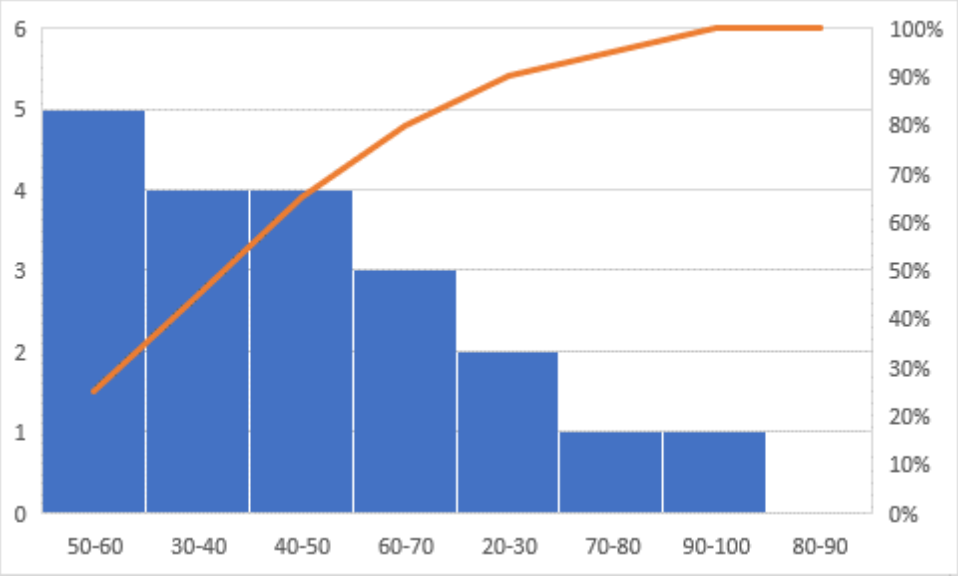


Fig. 3.14. Histogram for the Continuous Variables

To Do Activity

We shall illustrate the construction of a histogram with the following example. In a plot of 165 tillers, the following was obtained

No of Plants	1	2	3	4	5	Total
No of Tillers	12	30	28	42	53	165

In this case,

i. which of them is variables
ii. Which of them is frequency

Summary of the Chapter

1. **Data visualization is the graphical representation of information and data.**
2. **Data are classified into two types i) Quantitative data ii) Qualitative data**
3. **Quantitative Data are further classified into two types i) Discrete data ii) Continuous data**
4. **Categorical and Quantitative types of data, which are used in combination to support a meaningful analysis or visualization**
5. **There are four different types of Measurement Scales i) Nominal ii) Ordinal iii) Ratio iv) Interval**
6. **A table contains quantitative data organized into rows and columns with categorical labels**
7. **A graph is primarily used to show relationships among data and portrays values encoded as visual objects (e.g., lines, bars, or points).**
8. **A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent**
9. **Pie chart is a special chart that uses "pie slices" to show relative sizes of data.**
10. **Histogram is a plot that lets you discover, and show, the underlying frequency distribution (shape) of a set of continuous data.**

Model Questions

Part A

[1]. Define: Tables and Graph
[2]. List out the different types of Data
[3]. List out the difference between Quantitative and Qualitative Data chart
[4]. List out the difference between discrete data and continuous data
[5]. Give one example for Interval Scale Measurement
[6]. Give one example for Ratio Scale Measurement
[7]. List out the importance of Data Visualization
[8]. Remember: Histogram
[9]. List out the difference between pie chart and bar chart
[10]. List out the eight types quantitative messages in data visualization
[11]. Bar graph can be drawn only vertically (True/False)

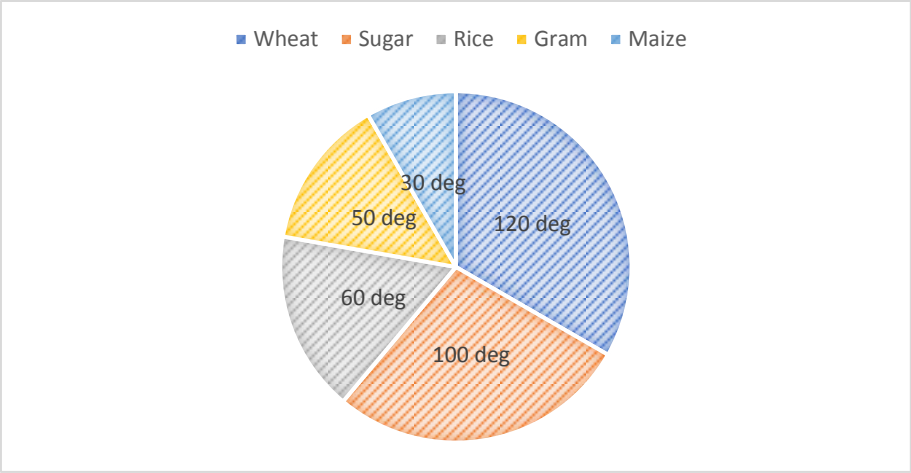
Part B

[1]. A survey of 100 people asked them "Which is your favorite vegetable?

Fruit:	Carrot	Peppers	bitter melon	Cabbage	Cauliflower	Baby corn
People:	30	20	15	10	20	5

Draw the Pie chart and bar chart and conclude:

[2]. Fig. shows the annual agricultural production of an Indian state. If the total production of all the commodities is 90000 tonnes, find the production (in tonnes) of (i) Wheat (ii) Sugar (iii)Rice (iv)Maize (v)Gram



[3]. Former Society where each 15 formers spends these money for fertilizer per day : Rs 100, Rs 50, Rs 150, Rs 230, Rs 70, Rs 90, Rs 110, Rs 180, Rs 130, Rs 40, Rs 120, Rs 80, Rs 150, Rs 30, Rs 85. Draw the histogram

[4]. Here are their figures for the five persons age and salary (Monthly): Draw the Scatter Plot:

Age	20	25	30	40	50
Monthly Salary (Rs)	10000	20000	30000	40000	50000

Chapter 4 Graphs and Data Analysis

Introduction

According to the World Economic Forum, the world produces 2.5 quintillion bytes of data every day, and 90% of all data has been created in the last two years. With so much data, it's become increasingly difficult to manage and make sense of it all. It would be impossible for any single person to wade through data line-by-line and see distinct patterns and make observations. Data proliferation can be managed as part of the data science process, which includes data visualization.

Data visualization is the act of taking information (data) and placing it into a visual context, such as a map or graph. It makes big and small data easier for the human brain to understand, and visualization also makes it easier to detect patterns, trends, and outliers in groups of data. Good data visualizations should place meaning into complicated datasets so that their message is clear and concise. Sometimes a simple graph is the most effective, whereas other times a more complex visualization is needed to get the job done.

Objectives

To understand the different types of graphs for data visualization
To become familiar with problem solving and decision-making phases
To visualize the challenges in Business Analytics

Chapter Structure

4.1 Eight Types of Quantitative Messages

4.2 Basic of Data Analysis

4.3 Different Types of Graphs

4.4 Real Outcomes of the Business Analytics

4.5 Problem Solving and Decision-Making Phases

4.6 Different types of Data Analysis

4.7 Challenges in Data Analytics

A Picture is worth a Thousand Words

4.1 Eight Types of Quantitative Messages

Author Stephen Few described ***eight types of quantitative messages*** that users may attempt to understand or communicate from a set of data and the associated graphs used to help communicate the message:

Time-series

A single variable is captured over a period, such as the unemployment rate over a 10-year period. A line chart may be used to demonstrate the trend.

Ranking

Categorical subdivisions are ranked in ascending or descending order, such as a ranking of sales performance (the measure) by salespersons (the category, with each salesperson a categorical subdivision) during a single period. A bar chart may be used to show the comparison across the sales persons.

Part-to-whole

Categorical subdivisions are measured as a ratio to the whole (i.e., a percentage out of 100%). A pie chart or bar chart can show the comparison of ratios, such as the market share represented by competitors in a market.

Deviation

Categorical subdivisions are compared against a reference, such as a comparison of actual vs. budget expenses for several departments of a business for a given time period. A bar chart can show comparison of the actual versus the reference amount.

Frequency distribution

Shows the number of observations of a variable for given interval, such as the number of years in which the stock market return is between intervals such as 0-10%, 11-20%, etc. A histogram, a type of bar chart, may be used for this analysis. A boxplot helps visualize key statistics about the distribution, such as median, quartiles, outliers, etc.

Correlation

Comparison between observations represented by two variables (X, Y) to determine if they tend to move in the same or opposite directions. For example, plotting unemployment (X) and inflation (Y) for a sample of months. A scatter plot is typically used for this message.

Nominal Comparison

Comparing categorical subdivisions in no order, such as the sales volume by product code. A bar chart may be used for this comparison.

Geographic or Geospatial

Comparison of a variable across a map or layout, such as the unemployment rate by state or the number of persons on the various floors of a building. A cartogram is a typical graphic used.

4.2 Basic of Data Analysis

Mean, median, and mode are three different statistical parameters. There are many statistical parameters in analytics, but these are the three most common, and are certainly the three you are most likely to encounter in our pre-business analytics courses. Each set has a unique mean, mode, median values.

Mean

Description	How to find	Features
"Average" value of the set of numbers.	Add all numbers to get a total, and then divide by the number of entries (number count of values you added).	Easy and quick way to represent the entire data values by a single or unique number due to its straightforward method of calculation.

Median

Description	How to find	Features
Middle value of the set of numbers	Organize the numbers in increasing order,the median is the middle or centermost number. If there are two middle numbers, add them and divide by 2 to get the median.	The concept of the median is intuitive thus can easily be explained as the center value.

Mode

Description	How to find	Features
Most common or frequent value or item of the set	Tally or count how many times a number appears in the list of data. The mode is the one that shows the most.	Useful to find the most "popular" or common item. This includes data sets that do not involve numbers.

Range

Range is the difference between the maximum and minimum value of the set. What the range provides is a quick and rough estimate of the spread of data values within a set.

Relationship between Mean, Mode and Median

Mode = mean - 3 [mean - median]

Mode = 3 median - 2 mean

Median = mode + 2/3 [mean - mode]

Points to Remember

Problem 4.1: Find the Mean, Median, Mode and Range of the given data

Data Set: 3, 8, 3, 4, 3, 6, 4, 2, 3

Solution:

Mean:	**Median:**
The mean is the usual average, so I'll add and then divide: (3 + 8 + 3 + 4 + 3 + 6 + 4 + 2 + 3) ÷ 9 = 4	The median is the middle value, so first I'll have to rewrite the list in numerical order: 2, 3, 3, 3, 3, 4, 4, 6, 8 The formula for the place to find the median is "([the number of data points] + 1) ÷ 2", There are nine numbers in the list, so the middle one will be the (9 + 1) ÷ 2 = 10 ÷ 2 = 5th number: 2, 3, 3, 3, 3, 4, 4, 6, 8 So the median is 3.

Mode:	**Range:**
The mode is the number that is repeated more often than any other, so 3 is the mode.	The largest value in the list is 8, and the smallest is 2, so the range is 8 – 2 = 6.

Ans:

Mean: 4

Median: 3

mode: 3

range: 6

Problem 4.2: Find the median of the data set below.

{102, 403, 729, 843, 920, 360, 842, 941, 357, 483, 207, 670, 471, 109}

Solution:

First, order the data. Note that because the set has 14 members, the median is the mean of two central values. These values are underlined in the ordered set below.

{102, 109, 207, 357, 360, 403, **471, 483**, 670, 729, 842, 843, 920, 941}

Now, calculate the median M by finding the mean of 471 and 483.

The median of this data set is thus 477.

Ans: 477

Practices

Let us take the body weight (Kg) of 10 poultry birds in a village.

3, 3.5, 2, 1.5, 1.2, 3, 4, 3, 3, 1.8

Find out the mean, median, mode, range value and what you infer from the statistical parameters

4.3 Different Types of Graphs

Scatter Graph

A scatter plot (also called a scatterplot, scatter graph, scatter chart, scatter gram, or scatter diagram) is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. It is used for Determine the relationship (e.g., correlation) between the variables. For example, unemployment (x) and inflation (y) for multiple time periods. A Scatter (XY) Plot has points that show the relationship between two sets of data as shown in figure 4.1.

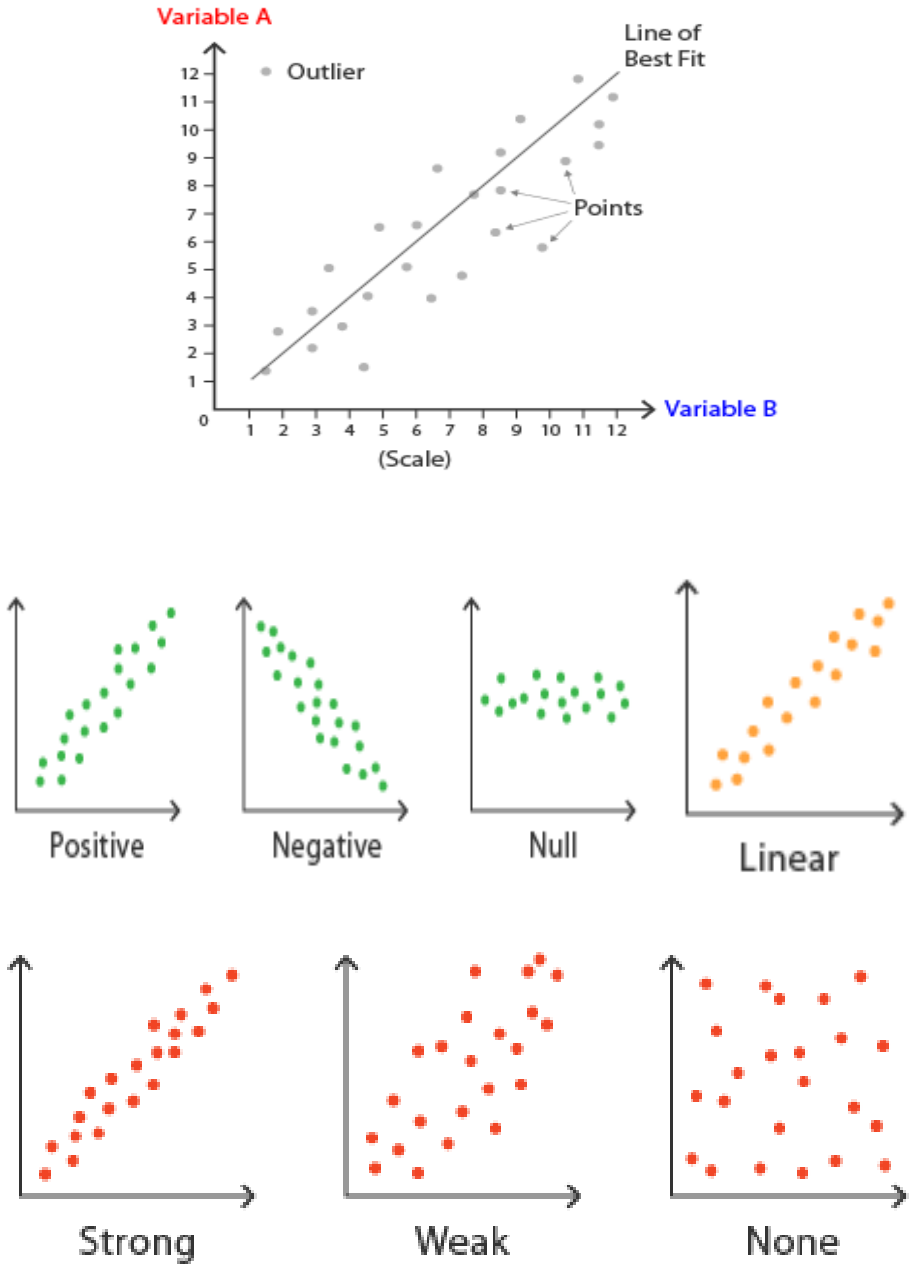


Fig. 4.1. Correlation Strength

Problem 4.3: The local ice cream shop keeps track of how much ice cream they sell versus the noon temperature on that day. Here are their figures for the last 12 days: Draw the Scatter Plot:

Solution:

Ice Cream Sales Vs Temperature	
Temperature °C	Ice Cream Sales (Rs)
14.2°	215
16.4°	325
11.9°	185
15.2°	332
18.5°	406
22.1°	522
19.4°	412
25.1°	614
23.4°	544
18.1°	421
22.6°	445
17.2°	408

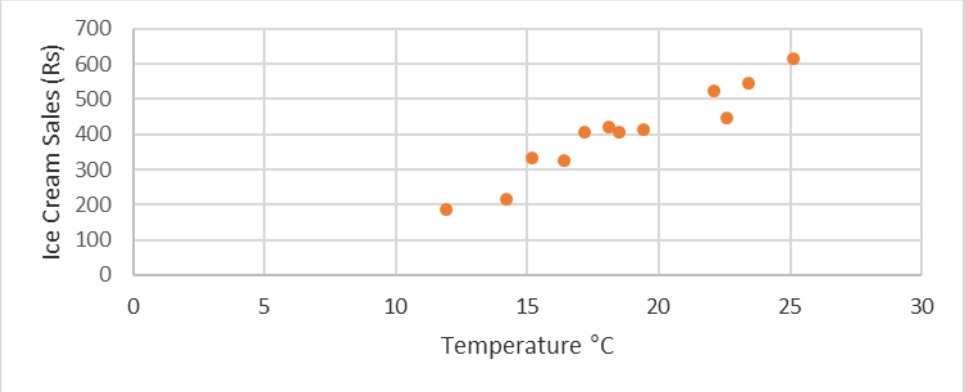


Fig. 4.2. Scatter Graph for Ice Cream Sales Vs Temperature

From the scatter graph figure 4.2, It is now easy to see that warmer weather leads to more sales, but the relationship is not perfect.

Two variables have a **positive association** when above-average values of one tend to accompany above-average values of the other, and when below-average values also tend to occur together.

Points to Remember

Two variables have a **negative association** when above-average values of one tend to accompany below-average values of the other.

Points to Remember

Problem 4.4: Here are their figures for the student's achievement motivation and GPA: Draw the Scatter Plot:

Students' Achievement Motivation	1	2	3	4	5
GPA	0.5	1.8	2.5	2.9	3.2

Solution:

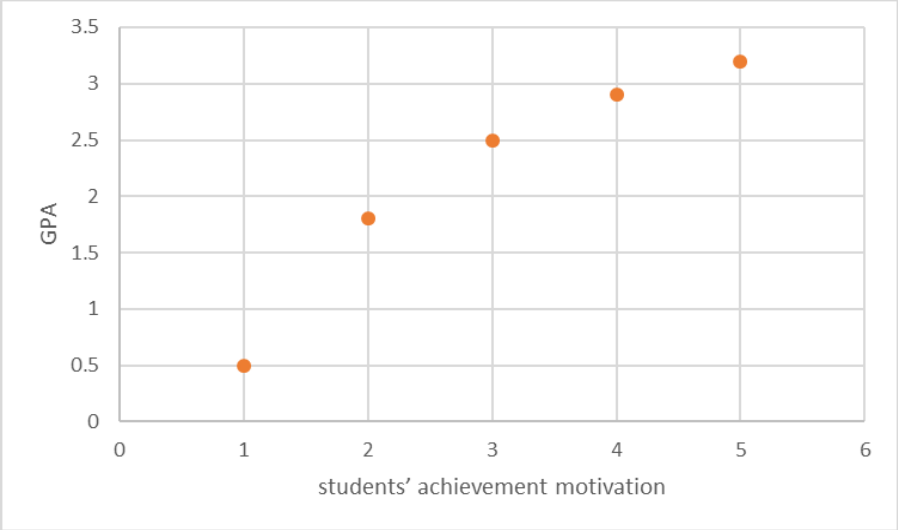


Fig. 4.3. Scatter Graph for Students Achievement Motivation Vs GPA (Positive Correlation)

This Scatter Graph figure 4.3 compares students' achievement motivation and their GPA. These two variables have a positive association because as GPA increases, so does motivation.

Interference

Problem 4.5: Here are their Sample height and thickness: Draw the Scatter Plot:

Height	0.4	1.2	2	3.1	4.5	5.7	7.1	8.4	9.3	9.8
Thickness	7.8	7.1	6.8	6	5.2	4.3	3.4	2.3	1.1	0.5

Solution:

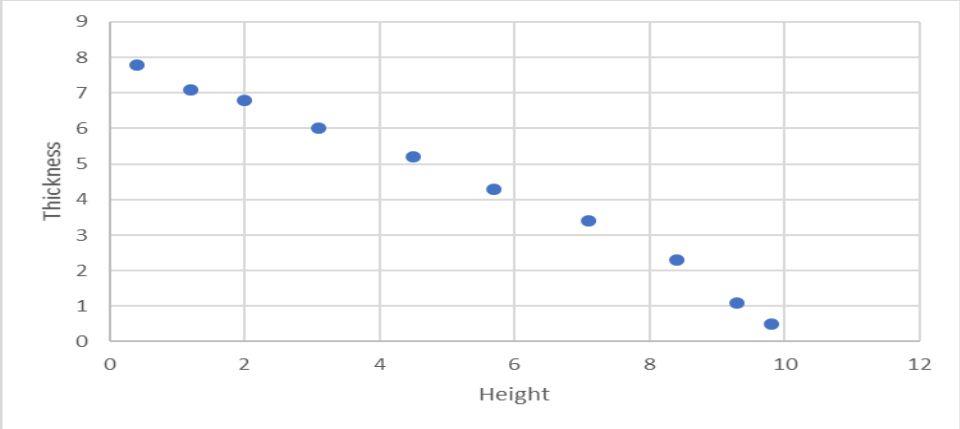


Fig. 4.4. Scatter Graph for Samples Height Vs Thickness (Negative Correlation)

This Scatter Graph figure 4.4 compares samples height and their thickness. These two variables have a negative association because height decrease, so does thickness.

Interference

Area Graphs

Area Graphs are Line Graphs but with the area below the line filled in with a certain color or texture. Area Graphs are drawn by first plotting data points on a Cartesian coordinate grid, joining a line between the points and finally filling in the space below the completed line. It likes Line Graphs; Area Graphs are used to display the development of quantitative values over an interval or time period. They are most used to show trends, rather than convey specific values.

Two popular variations of Area Graphs are: grouped and Stacked Area Graphs. Grouped Area Graphs start from the same zero axis, while Stacked Area Graphs have each data series start from the point left by the previous data series.

Points to Remember

Problem 4.6: Here are their elementary, middle and High school budget with different years: Draw the Area Graph:

Quartrant-Average Salary in Rs	Elementary	Middle School	High school
I	8,000	10,000	12,000
II	9,000	11,000	11,000
III	10,000	12,000	13,000
IV	10,000	13,000	14,000

Solution:

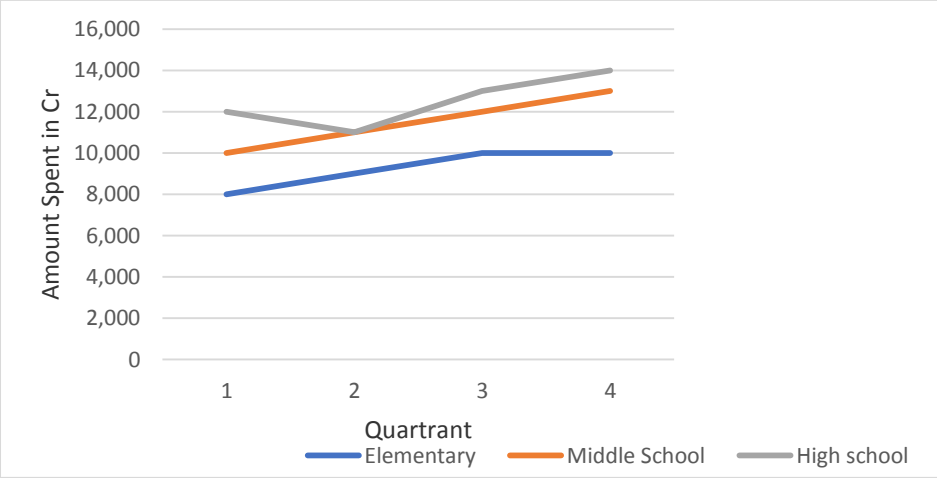


Fig. 4.5. Line Graph for Average Budget Value for Schools

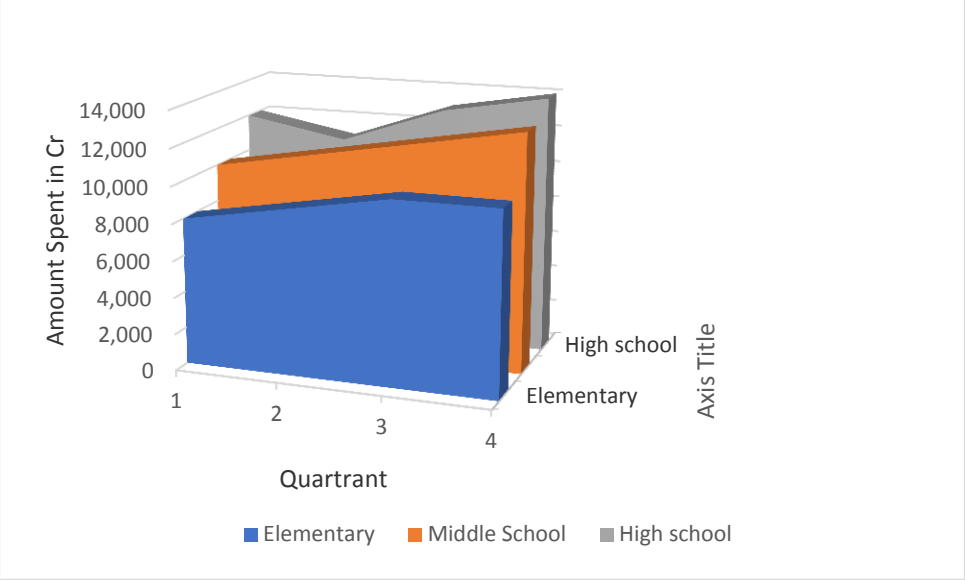


Fig. 4.6. Area Graph for Average Budget Value for Schools

From the line graph figure 4.5 and area graph figure 4.6, the Average Budget Value of Middle Schools' has been increasing gradually, but there are variations in Average Budget Value of High School and Elementary School.

Interference

Problem 4.7: Here is their Rajas' family Monthly Budget in terms of Rupees: Draw the Area Graph:

Monthly budget in Rupees				
Month	Gas	Rent	Food	Utilities
January	1000	6000	3000	5000
February	1500	6000	2500	7000
March	1000	6000	4500	4000

Solution:

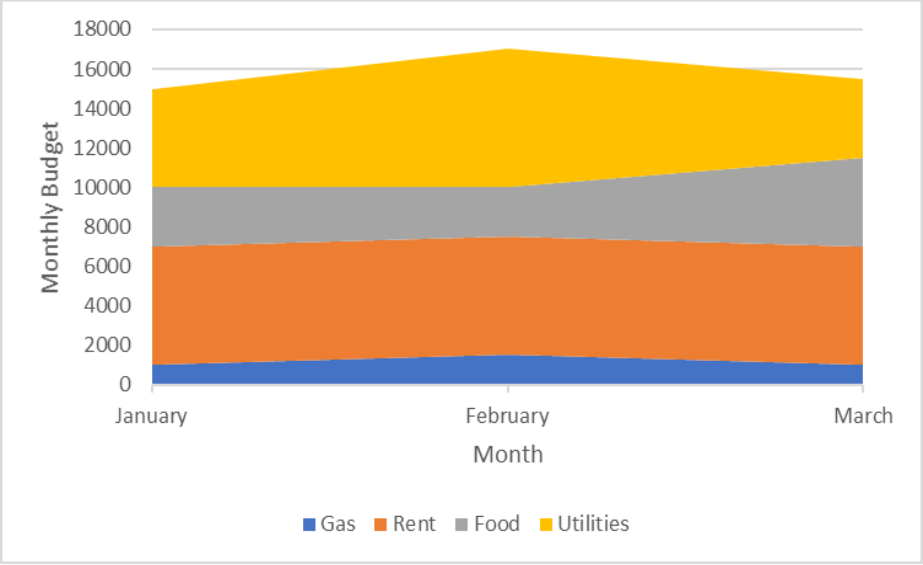


Fig. 4.7 Area Graph for Monthly Budget

From the area graph figure 4.7, the Monthly Budget for Gas and Rent are uniform, but there are variations in Monthly Budget for food and utilities

Interference

Funnel chart

A funnel chart is your data visualization of choice if you want to display a series of steps and the completion rate for each step. This can be used to track the sales process, a marketing funnel or the conversion rate across a series of pages or steps. Funnel charts are most often used to represent how something moves through different stages in a process. A funnel chart displays values as progressively decreasing proportions amounting to 100 percent in total.

Ideally the funnel chart shows a process that starts at 100% and ends with a lower percentage where it is noticeable in what stages the fall out happens and at what rate. If the chart is also combined with research data, meaning quantified measurements of just how many items are lost at each step of the sales or order fulfillment process, then the funnel chart illustrates where the biggest bottlenecks are in the process.

Points to Remember

Problem 4.8: Here are their one company's' customer profile count: Draw the Funnel Diagram:

No of Counts	
website Visit	15600
Catalog download	4200
Requested Price list	1900
Invoice Sent	800
Product Finalized	790

Solution:

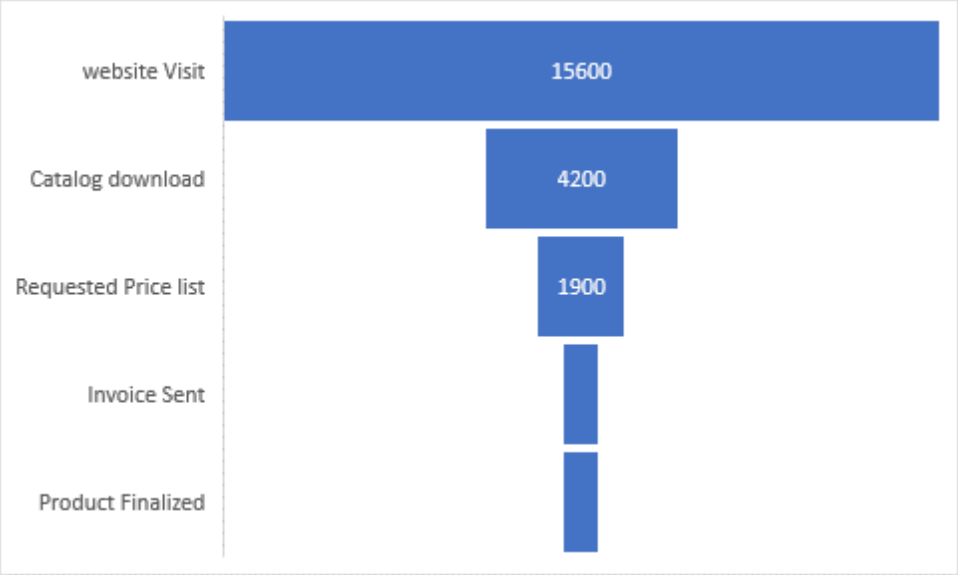


Fig. 4.8.Funnel Chart for Customer Profile Court

4.4 Real Outcomes of the Business Analytics

In other words, whatever the form of business analytics may be, it would help us answer the following fundamental questions critical for business analytics as shown in figure 4.9.

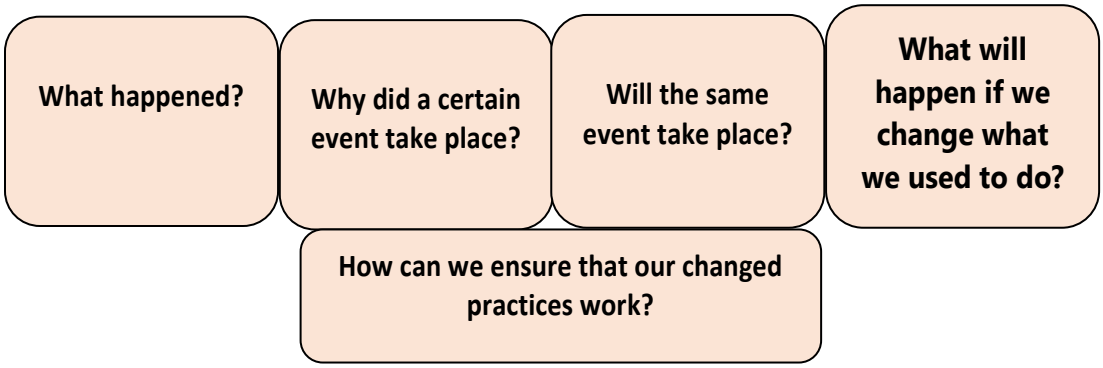


Fig. 4.9. Fundamental Questions Chart for Business Analytics

The following questions are listed:

[1]. What did the data tell us?
[2]. Why did a certain event take place?
[3]. What are the sources of problems?
[4]. Will the same event take place?
[5]. Are there any noticeable patterns of the problem?
[6]. How can we deal with the recurring problem?
[7]. What is the value the change will bring?
[8]. **Is there scientific evidence indicating the validity and usefulness of our changed practices?**

However, organizational titles vary such as analyst, business analyst, business systems analyst or maybe systems analyst. Business analytics makes extensive use of analytical modeling and numerical analysis, including explanatory and predictive modeling and fact-based management to drive decision making. It is therefore closely related to management science. Analytics may be used as input for human decisions or may drive fully automated decisions.

Points to Remember

Data Science uses both structured and unstructured data whereas Business Analytics uses mostly structured data. Data Science is the science of data study using statistics, algorithms, and technology whereas Business Analytics is the Statistical study of business data. A Data Analyst interprets data and turns it into information which can offer ways to improve a business, thus affecting business decisions. Data Analysts gather information from various sources and interpret patterns and trends – as such a Data Analyst job description should highlight the analytical nature of the role.

Points to Remember

Here are their one company's' investment plan (Cr): Draw the Funnel Diagram:

Investment Category	Livestock	Agriculture	Metals	Energy
Investment (Cr)	78	54	30	13

To Do Activity

4.5 Problem Solving and Decision-Making Phases

Business Analytics represents only a portion of the overall problem solving and decision-making process. The following six steps in the problem-solving process as shown in figure 4.10.

1. Recognizing the problem
2. Defining the problem
3. Structuring the problem
4. Analyzing the problem
5. Interpreting results and plan
6. Implementing the solution

1. Recognizing the Problem

Problems exist when there is a gap between what is happening and what we think should be happening. For example, costs are too high compared with competitors.

2. Defining the Problem

Clearly defining the problem is not a trivial task. Complexity increases when the following occur:

- Large number of courses of action
- Several competing objectives
- External groups are affected
- Problem owner and problem solver are not the same person
- Time constraints exist

3. Structuring the Problem

Stating goals and objectives
Characterizing the possible decisions
Identifying any constraints or restrictions

4. Analyzing the Problem

Identifying and applying appropriate Business Analytics techniques
Typically involves experimentation, statistical analysis, or a solution process
Much of this course is devoted to learning BA techniques for use in Step 4.

5. Interpreting Results and plan

Managers interpret the results from the analysis phase.
Incorporate subjective judgment as needed.

Understand limitations and model assumptions.
Plan utilizing the above information.

6. Implementing the Solution

Translate the results of the model back to the real world. Make the solution work in the organization by providing adequate training and resources.

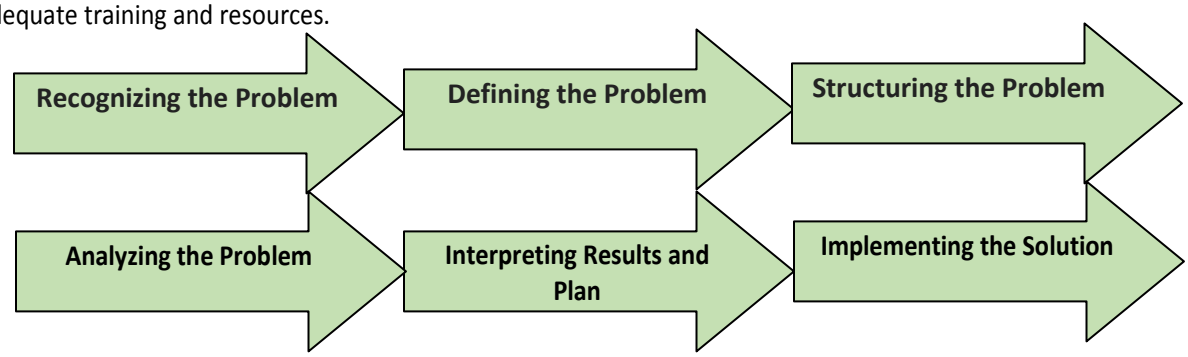


Fig. 4.10.Problem Solving and Decision Making for Business Analytics

Factors to be considered for developing effective Analytical Tools

- Will analytics solve the problem?
- Can they leverage an existing solution?
- Is a decision model really needed?
- Guidelines for successful implementation:
- Build insight, not black boxes.
- Remove unneeded complexity.
- Partner with end users in discovery and design.
- Develop an analytic champion.

Points to Remember

Descriptive analytics: the use of data to understand past and current business performance and make informed decisions;
Predictive analytics: predict the future by examining historical data, detecting patterns or relationships in these data, and then extrapolating these relationships forward in time;
Prescriptive analytics: identify the best alternatives to minimize or maximize some objective

Points to Remember

4.6 Different Types of Data Analysis

Descriptive Analysis

Descriptive analysis performs the function of "Describing" or summarizing raw data to make it easily understandable and interpretable by humans. Descriptive analytics is used to analyze and derive insights from past

data whereas predictive analytics is used to study trends and predict what will happen in the future. Decision analytics supports human decisions with visual analytics that the user models to reflect reasoning. Descriptive analytics gains insight from historical data with reporting, scorecards, clustering etc.

Predictive Analytics

Predictive analytics employs predictive modeling using statistical and machine learning techniques. This type of Business Analytics uses forecasting techniques and statistical models to find out what is going to happen in future. Predictive analysis helps us in predicting the future course of events and taking necessary measures for the same.

Prescriptive Analytics

Prescriptive Analytics is something that came into existence only about a decade ago and has since proven to be a very popular and powerful tool with businesses around the world. It can be used to analyze data in the present as well as predict what might happen go forward. Most significantly, it provides insights into what steps should be taken based on the available data and what the impact of these decisions would be. This tremendous versatility places Prescriptive Analytics at the cutting edge of analytics technology. Prescriptive analytics recommends decisions using optimization, simulation, etc.

Descriptive Analysis: When you visit your local supermarket, there are certain items you buy every single time – staples like milk, bread, fruit etc. Each time you buy an item, say a litre of milk that information is recorded in the store's inventory system. Over a period of a few months, the store can analyze this data to get an idea of how much milk has been sold, which in turn gives the retailer very interesting insight into the buying patterns of customers and supplies.

This is, quite simply, descriptive analytics. All the information is coming from the customer's shopping basket and that data is just being sliced and diced and looked at from different angles in order to draw relevant conclusions. Descriptive analytics is the simplest form of data analysis, as it can only be used to analyze data from the past. Due to this limitation and the ease with which it can be learnt, it is often considered unglamorous. However, it remains a very powerful tool for retail, sales, marketing and much more.

Predictive Analytics: Retailers realized that while walking all the way to pick up your essentials, you might be tempted to buy something else as well. Plus, gum and magazines are cheap impulses buys so you're likely to not think twice to put them in your basket, just before you ring up all your purchases.

This, in a nutshell, is how predictive analytics works – by identifying patterns in historical data and then using statistics to make inferences about the future. We essentially attempt to fit the data into a certain pattern and if the data is following that pattern, we can predict what will happen in the future with some certainty.

If a person buys product A, retailers want to know if he or she is also likely to buy product B or C. Understanding this relationship between products is called product affinity analysis or association analysis. Predictive analytics is widely used across multiple industries such as retail, telecom, pharmaceuticals and more, as a way for companies to optimize their business practices.

Prescriptive Analytics: An excellent example of an industry where this technique is extensively used is aviation. Airlines companies are always looking for ways to optimize their routes for maximum efficiency, which can help them save billions of dollars. The sheer numbers in the industry are staggering – globally, there are over 50 million commercial flights every year; that is effectively a flight every second; even a simple route like San Francisco to Boston has 2000+ possible route options. And so, the industry is constantly using Prescriptive Analytics to identify more efficient ways to operate, which can keep airline costs down and profits up.

Descriptive analytics, Predictive analytics, Prescriptive Analytics: Combined Application

Retail Markdown Decisions

Most department stores clear seasonal inventory by reducing prices.

The Objective is to identify: When to reduce the price and by how much?

Descriptive analytics: examine historical data for similar products (prices, units sold, advertising, …)

Predictive analytics: predict sales based on price

Prescriptive analytics: find the best sets of pricing and advertising to maximize sales revenue

4.7 Challenges in Data Analytics

Big data to refer to massive amounts of business data from a wide variety of sources, much of which is available in real time, and much of which is uncertain or unpredictable. IBM calls these characteristics volume, variety, velocity, and veracity. The term 'Big Data' tends to be the source of confusion for a lot of people, largely because of how it is used interchangeably with 'analytics' in the media. Four Vs are represented as shown in figure 4.11.

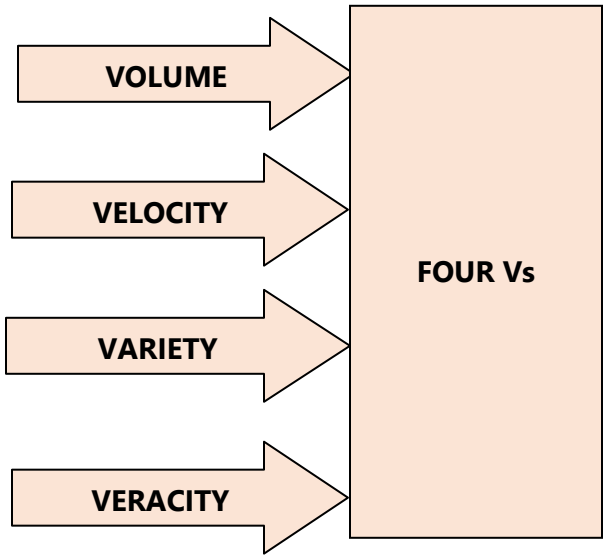


Fig. 4.11. Four Vs in Big Data

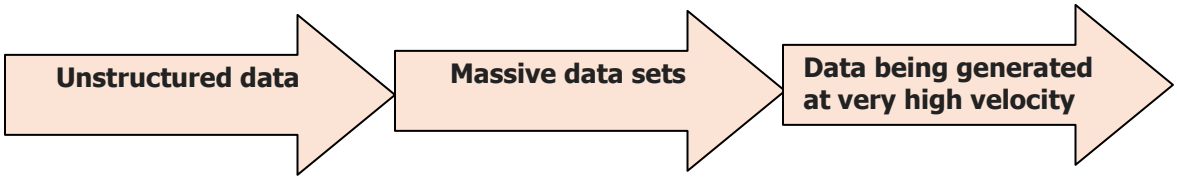


Fig. 4.12. Data in Business Analytics

For a very long time, companies would only work with very structured data – what they generated internally (like customer or transaction data) or the information they purchased from data aggregator businesses. More recently, companies have had to start contending with vast amounts of unstructured data such as audio, video and social media data, which would not be in a standard tabular format. Traditional analytics techniques are not designed to contend with data of this nature. Soon enough, businesses realized that there was a need for alternative tools, databases and platforms in order to effectively analyze the vast amounts of unstructured data that was being generated. This led to the classification of Big Data as an independent concept while still being a part of analytic so subset of analytics that deals with unstructured data, massive data sets, data being generated at very high velocity as shown in figure 4.12.

Interactive Data Corporation (IDC) estimates that Data creation will reach 163 zetta bytes by 2025 as shown in figure 4.13. Big Corporations, Governments, Entrepreneurs and almost everyone else is using Data Science to generate insights by unearthing patterns and by decoding this data. These insights are helping to improve efficiency and to offer innovative solutions to business problems. Analytics is also leveraged widely to plan election campaigns and win elections.

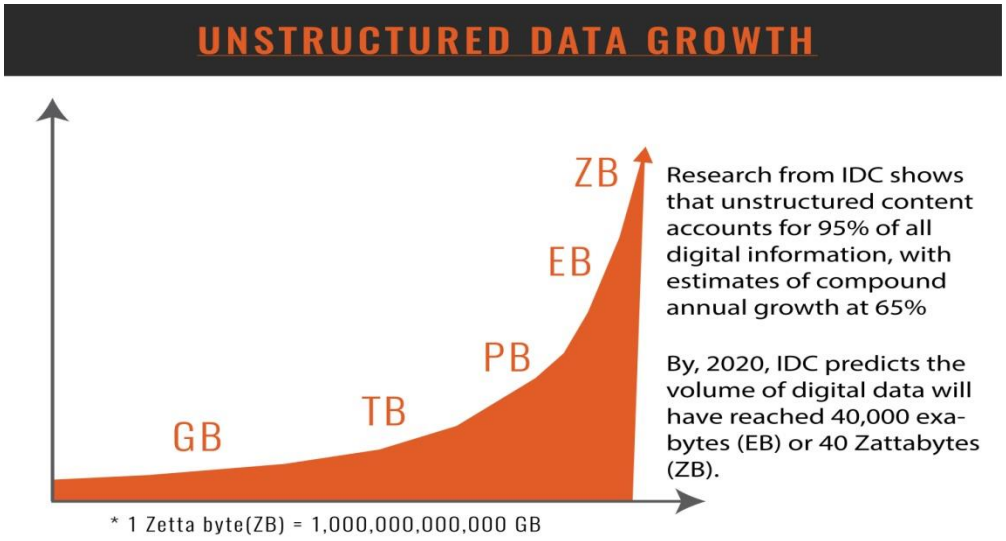


Fig. 4.13. Data Growth in Business Analytics

Source: Interactive Data Corporation (IDC)

Challenges in Business Analytics

- Lack of understanding of how to use analytics
- Competing business priorities
- Insufficient analytical skills
- Difficulty in getting good data and sharing information
- not understanding the benefits versus perceived costs of analytics studies

Summary of the Chapter

1. **Data visualization is the graphical representation of information and data.**
2. **A scatter plot (also called a scatterplot, scatter graph, scatter chart, scatter gram, or scatter diagram) is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. It is used for Determine the relationship (e.g., correlation) between the variables**
3. **Area Graphs are Line Graphs but with the area below the line filled in with a certain color or texture**
4. **A funnel chart is your data visualization of choice if you want to display a series of steps and the completion rate for each step**
5. **Problem Solving and Decision-Making has six phases**
6. **Volume, Velocity, Variety and Veracity are 4 Vs in Business Analytics**
7. **There are three different types of analytics:**
8. **Descriptive analytics: examine historical data for similar products (prices, units sold …)**
9. **Predictive analytics: predict sales based on price**
10. **Prescriptive analytics: find the best sets of pricing and advertising to maximize sales revenue**

Model Questions

Part A

[1]. List out the eight types quantitative messages in data visualization
[2]. Define: Scatter Diagram
[3]. Remember: Funnel Chart
[4]. Remember: Descriptive Analytics
[5]. List out the Points for Data Analytics Vs Business Analytics
[6]. List out the any two data visualization software packages
[7]. List out the different types of data analysis
[8]. Give one example for Descriptive Analysis
[9]. Give one example for Predictive Analysis
[10]. Give one example for Prescriptive Analysis

[11]. Define: four Vs

[12]. How can identify the positive and negative associates between the variables through scatter diagram

[13]. List out the factors for effective analytical tools

[14]. Differentiate the line graph and area graph

Part B

[1]. Here are their figures for the five persons age and salary (Monthly): Draw the Scatter Plot:

Age	20	25	30	40	50
Monthly Salary (Rs)	10000	20000	30000	40000	50000

[2]. Explain about the Decision Making and Problem-Solving Phases

[3]. Explain about the different types data analysis Techniques

[4]. Elaborate the challenges in Data Analysis

[5]. The table below shows the number of absences, x, in a Calculus course and the final exam grade, y, for 7 students. Find the scatter line.

X	1	0	2	6	4	3	3
Y	85	80	70	55	90	90	95

Chapter 5 Analytical Techniques

Introduction

The regression and correlation analysis described on this chapter is used to investigate relationships between two variables (x and y). Is a change in one of these variables associated with a change in the other?

For example,

[1]. If we increase the investment do, we increase the growth rate of productivity?

[2]. **Does an increase in protein content of bird tissues correlate with thinning of the eggshell?**

When it comes to planning out the future of our rural business, the reliable source for prediction is a look at your past. Quantitative sales/demand forecasting is one of the most objective ways that you can predict where your rural business will be in the near future. The amount of analytical attention to the forecasting process should be in direct proportion to the importance of the product/service being forecast.

Objectives

To understand the importance and elements of sales/demand forecasting

To become familiar with different types of forecasting methods

To visualize the regression and correlation analysis for the given data

Chapter Structure

5.1 Correlation Analysis

5.2 Regression Analysis

5.3 Regression Vs Correlation Analysis

5.4 Fundamentals of Forecasting

5.5 Different Types of Forecasting Methods

Those who have knowledge don't predict. Those who predict, don't have knowledge

5.1 Correlation Analysis

Correlation Analysis is a technique of statistical evaluation used to study the strength of a relationship between two, numerically measured, continuous variables (e.g. rain fall and agriculture productivity). This analysis is useful when an analyst wants to establish if there are possible connections between variables. It is often mis understood that correlation analysis determines cause and effect; however, this is not the case because other variables that are not present in the analysis may have impacted on the results.

Correlation analysis deals with relationships among the variables. The correlation coefficient is a measure of linear association between two variables. Values of the correlation coefficient are always between -1 and +1.

Pearson correlation is the most widely used correlation statistic to measure the degree of the relationship between linearly related variables. For example, in the agriculture market, if we want to measure how two agriculture commodity prices are related to each other.Pearson r correlation is used to measure the degree of relationship between the two. The point-biserial correlation is conducted with the Pearson correlation formula except that one of the variables is dichotomous.

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

r_{xy}= Pearson correlation coefficient between x and y

n = No of observations

x_i= Value of x (for the i^{th} observation)

y_i= Value of y (for the i^{th} observation)

Positive correlation exists if one variable increases simultaneously with the other, i.e. the high numerical values of one variable relate to the high numerical values of the other.
Negative correlation exists if one variable decrease when the other increases, i.e. the high numerical values of one variable relate to the low numerical values of the other.

Pearson correlation coefficient is the measurement of correlation and ranges (depending on the correlation) between +1 and -1.

+1 indicates the **strongest positive correlation possible**
-1 indicates the **strongest negative correlation possible**
The closer the coefficient to either of these numbers the stronger the correlation of the data it represents.

On this scale 0 indicates no correlation, hence values closer to zero highlight weaker/poorer correlation than those closer to +1/-1.

Assumptions for the Pearson r correlation,

Normal Distribution: both variables should be normally distributed (normally distributed variables have a bell-shaped curve).

Linearity: Straight-line relationship between each of the two variables

Homoscedasticity: Data is equally distributed about the regression line.

Points to Remember

To Do Activity

Correlation between the Weather and Corp yield

The following table indicates rain fall (mm) and corp yield in different locations. Is there an association between the two variables?

Locations	A	B	C	D	E	F
Rain fall (mm)	10	20	30	40	50	60
Corp Yield (M. tons)	100	200	300	400	500	600

To Do Activity

Correlation between no of absences and grade

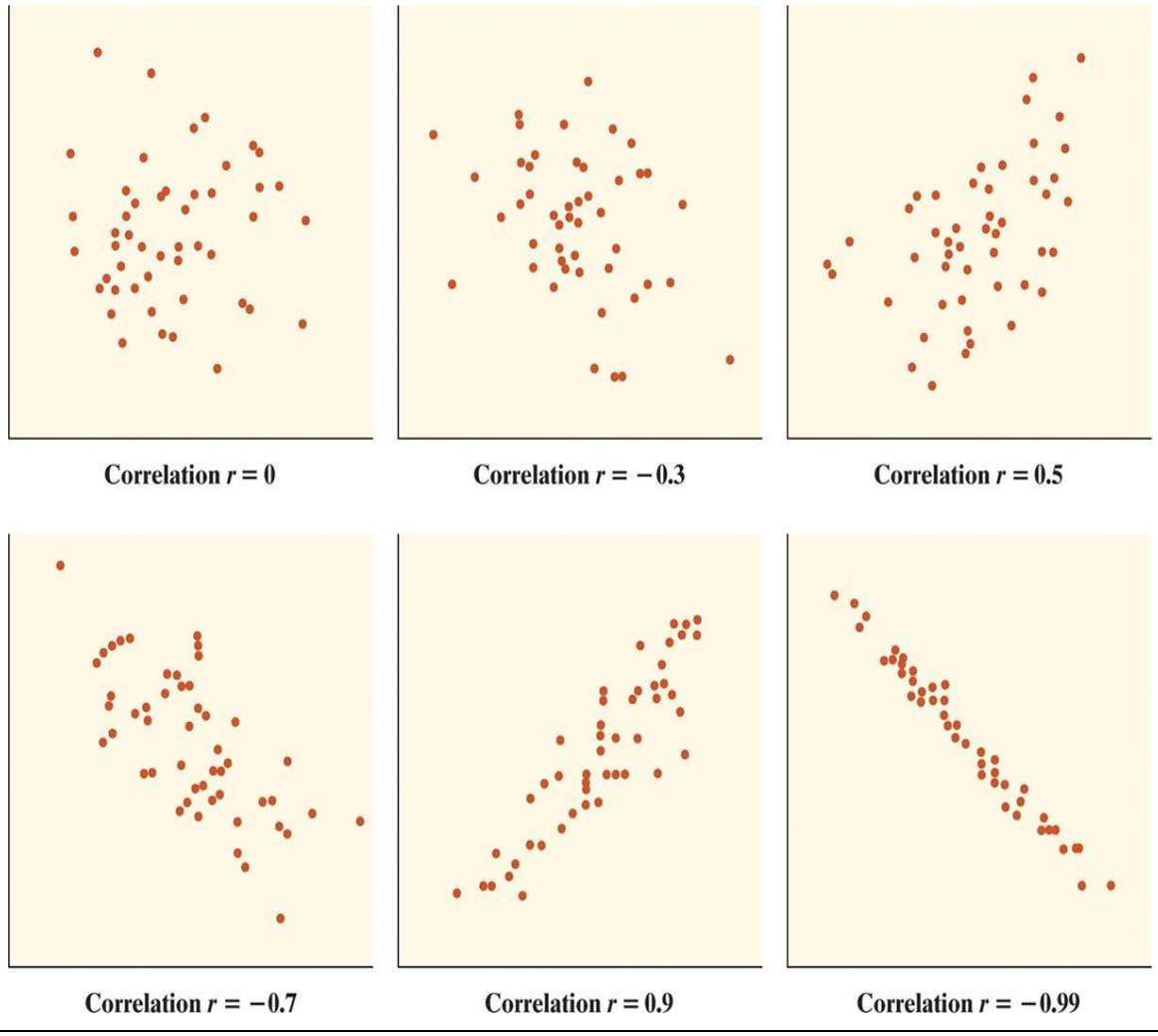
Table shows the numbers of absences in a business analytics course (x) and final exam grade/marks (y) for 6 students. Find the correlations coefficient and interpret the result

No of Absences (x)	1	0	2	7	4	3
Final Exam Grade (Y)	95	95	90	55	70	80

The same problem solve and conclude with regression method also

To Do Activity

What you infer from the following Graphs



Problem 5.1:

The following table indicates No of customers visited in the farm and No of products sold. Is there an association between the two variables?

Farm Location	A	B	C	D	E	F	G	H	I
No of Customer Visited in the farm (x)	56	56	65	65	50	25	87	44	35
No of Product Sold (y)	87	91	85	91	75	28	122	66	58

Solution:

Step 1: Formulate the table

No of Customers Visited in the farm (x)	No of Products Sold (y)	X^2	Y^2	XY
56	87	3136	7569	4872
56	91	3136	8281	5096
65	85	4225	7225	5525
65	91	4225	8281	5915
50	75	2500	5625	3750
25	28	625	784	700
87	122	7569	14884	10614
44	66	1936	4356	2904
35	58	1225	3364	2030
∑ X = 483	∑ y = 703	∑ X^2 = 28577	∑ y^2 = 60369	∑ XY = 41406

Step 2: Substitute the values in the formula

n = no of observation = 9

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

r_{xy} = 0.966

Step 3: Plot the Graph

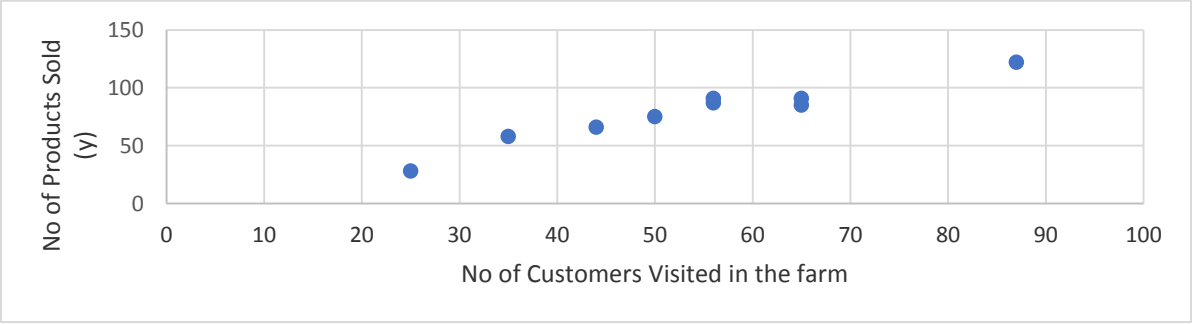


Fig. 5.1 Plot for Correlation Analysis

Interference from the Correlation Analysis figure 5.1: From the Graph and r_{xy} = 0.966, the variables have Positive correlation. if one no of customers in the farm increases simultaneously with the no of products sold

Problem 5.2:

An agriculture research organization tested a particular chemical fertilizer to try to find out the whether an increase in the amount of fertilizer used would lead to corresponding increase in the food supply

Amount of Fertilizer IP (x)	2	1	3	2	4	5	3
Bushel of Beans (y)	4	3	4	3	6	5	5

Solution:

Step 1: Formulate the table

Amount of Fertilizer IP (x)	Bushel of Beans (y)	X^2	y^2	XY
2	4	4	16	8
1	3	1	9	3
3	4	9	16	12
2	3	4	9	6
4	6	16	36	24
5	5	25	25	25
3	5	9	25	15
∑ X = 20	∑ y = 30	∑ X^2 = 68	∑ y^2= 136	∑ XY = 93

Step 2: Substitute the values in the formula

n = no of observation = 7

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

$$r_{xy} = \frac{(7 * 93) - (20 * 30)}{\sqrt{(7*68 - 400))} * \sqrt{(7*136 - 900))}}$$

$$r_{xy} = \frac{651 - (600)}{8.71 * 7.21}$$

$r_{xy} = 0.81$

Step 3: Plot the Graph

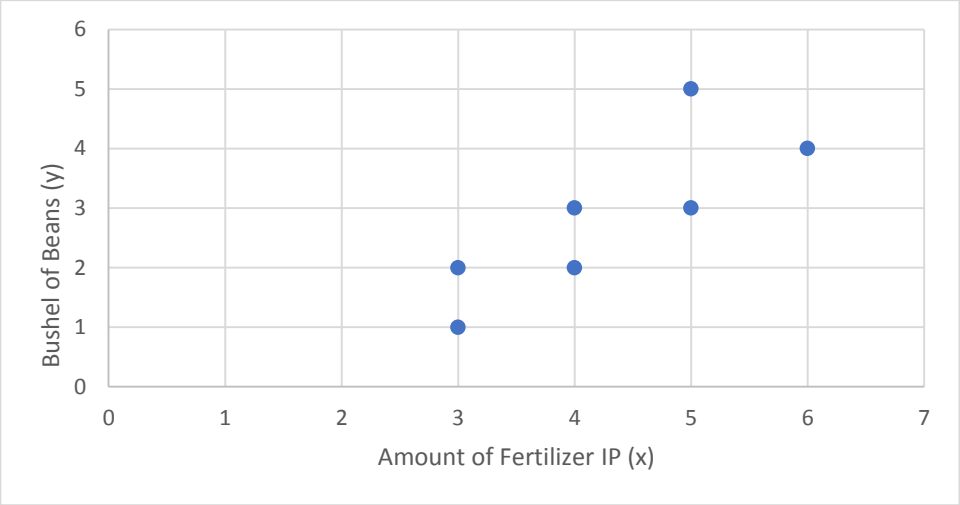


Fig. 5.2. Correlation Analysis for Amount of Fertilizer Vs Bushel of Beans

Interference from the Correlation Analysis figure 5.2: From the Graph and r_{xy} =0.81, the variables have Positive correlation. if amount of fertilizer increases simultaneously with the bushels of beans increase

To Do Activity:

Correlation between the investment and profit per 1 kg milk produced in selected dairy cattle farms

The following table indicates investment (mm) and profit per 1 kg milk produced (Rs). Is there an association between the two variables?

Investment (Rs)	1	2	3	4	5
profit per 1 kg milk produced (Rs)	2	4	6	8	10

Problem 5.2:

The time x in years that a farmer invested at a land and the farmer's return profit, y, for 5 formers are listed in the table below. Calculate and interpret the correlation coefficient r.

Invested Amount in Lakhs (x)	5	3	4	10	15
Profit Return in Lakhs (y)	25	20	21	35	38

Solution:

Step 1: Formulate the table

Invested Amount in Lakhs (x)	Profit Return in Lakhs (y)	X^2	y^2	XY
5	25	25	625	125
3	20	9	400	60
4	21	16	441	84
10	35	100	1225	350
15	38	225	1444	570
∑ X = 37	∑ y = 139	∑ X^2 = 375	∑ y^2 = 4135	∑ XY = 1189

Step 2: Substitute the values in the formula

n = no of observation = 5

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

$$r_{xy} = \frac{(5 * 1189) - (37 * 139)}{\sqrt{(5*375 - 37^2)} * \sqrt{(5*4135 - 139^2)}}$$

$$r_{xy} = \frac{802}{827.72}$$

$$r_{xy} = 0.97$$

r_{xy} =0.97, the variables have Positive correlation. if amount of investment increases simultaneously with the profit return increase

5.2 Regression Analysis

In "simple linear regression" (ordinary least-squares regression with 1 variable), you fit a line

$$y = a\,x + b$$

In the attempt to predict the target variable *y* using the predictor *x*, let's consider a simple example to illustrate how this is related to the linear correlation coefficient, a measure of how two variables are linearly related (or vary together).

y = how far up; x = how far along

a = Slope or Gradient (how steep the line is); b = the Y Intercept (where the line crosses the Y axis)

Step 1:

For each (x, y) point calculate x^2 and xy

Step 2:

Sum of all x, y, x^2 and xy, which gives us Σx, Σy, Σx^2 and Σxy

Step 3:

Calculate Slope a:

$$a = (N \sum (xy) - \sum x \sum y) / (N \sum (x^2) - (\sum x)^2)$$

N is the number of observations

Step 4:

Calculate Intercept b

$$b = ((\sum y) - a\,(\sum x)) / (N)$$

Step 5:

Assemble the equation of a line

$$y = a\,x + b$$

Problem 5.3:

A former found how many hours of sunshine Vs how much corps were sold at the shop from Monday to Friday.

How many kg corp sold expect at 10 hrs of sun tomorrow?

"x" Hours of Sunshine	2	3	5	7	9
"y" Corp Sold (Kg)	4	5	7	10	15

$$y = a\,x + b$$

Solution:

Let us find the best a (slope) and b (y-intercept) that suits that data

Step 1: For each (x, y) calculate x^2 and xy:

x	y	x^2	xy
2	4	4	8
3	5	9	15
5	7	25	35
7	10	49	70
9	15	81	135

Step 2: Sum of all x, y, x^2 and xy (gives us Σx, Σy, $Σx^2$ and Σxy):

x	Y	x^2	xy
2	4	4	8
3	5	9	15
5	7	25	35
7	10	49	70
9	15	81	135
Σx: 26	**Σy: 41**	**$Σx^2$: 168**	**Σxy: 263**

Step 3: Calculate Slope a:

$$a = (N \sum (xy) - \sum x \sum y) / (N \sum (x^2) - (\sum x)^2)$$

also N (number of data values) = 5

$a = (5 \times 263 - 26 \times 41)/(5 \times 168 - 26^2)$

$= (1315 - 1066)/(840 - 676)$

a = 1.52

Step 4: Calculate Intercept b

$$b = ((\sum y) - a\,(\sum x)) / (N)$$

b = (41 − 1.52 x 26)/5

b = 0.30

Step 5: Assemble the equation of a line

y = a x + b

$$y = 1.52\,x + 0.30$$

Interference from the Regression Analysis:

we expect 10 hrs of sun tomorrow, so we can use the above equation to estimate that corp will sell 15.5

To Do Activity:

We'll build on the previous example of trying to forecast next year's sales based on changes in GDP. If sales point 350, then what will happen in GDP?

Sales (Cr)	100	250	275	200	300
GDP (%)	2	2.90	3.40	3.60	3.90

To Do Activity:

Correlation between the Investment and Corp Productivity: The following table indicates Investment (Cr) and corp productivity (M.tons) in different locations. Is there an association between the two variables?

Locations	A	B	C	D	E	F
Investment (Cr)	10	20	30	40	50	60
Corp Productivity (M. tons)	5	6	7	8	9	10

To Do Activity:

The table shows plant height x in inches and the yield per plant (g) for 6 plant, Draw the regression equation

Plant	A	B	C	D	E	F
plant height x	68	72	65	70	62	75
yield per plant (g)	90	85	100	88	105	70

5.3 Correlation Analysis Vs Regression Analysis

Correlation is described as the analysis which lets us know the association or the absence of the relationship between two variables 'x' and 'y'. On the other end, Regression analysis, predicts the value of the dependent variable based on the known value of the independent variable, assuming that average mathematical relationship between two or more variables.The differences between the Correlation Analysis and Regression are listed in the table 5.1.

Table 5.1 Correlation and Regression Analysis

Comparison	Correlation	Regression
Meaning	Correlation is a statistical measure which determines co-relationship or association of two variables.	Regression describes how an independent variable is numerically related to the dependent variable.
Usage	To represent linear relationship between two variables.	To fit a best line and estimate one variable based on another variable.
Indicates	Correlation coefficient indicates the extent to which two variables move together.	Regression indicates the impact of a unit change in the known variable (x) on the estimated variable (y).

Take last five years Budget value Figure. Is there a link between a union budget allocation amount for agriculture (Rs) and agriculture productivity (tones)? Is there a link and Correlation measures these sorts of relationship? You may be examining 2 variables and trying to study the relationship that may exist between the two. A simple example might be the union budget allocation amount for agriculture (Rs) and agriculture productivity (tones).

1. Check, is there strong correlation and relation between the two variables
2. Plot the graph between union budget allocation amount for agriculture (Rs) and agriculture productivity (tones)

Case Study 1

5.4 Fundamentals of Forecasting

Virtually all the management decisions require as input a good estimate of future demand. In this situation, Forecasting is essentially needed one. forecasting is the art or science of predicting future demand by anticipating what consumers are likely to do in a given set of circumstances.

Forecasting:

- Prediction
- Projection
- Estimate of some future activity, event, or occurrence.

Time Series

A time series is a time-ordered sequence of observations taken at regular intervals, e.g. daily, weekly, monthly, quarterly or annually

Demand Forecasting

Demand forecasting refers to the process of determining the amount of agriculture product and related information that consumers will require, either in the short or long term.

Tamil Nadu Agriculture Weather Network
helping people with real time weather

Success of monsoon-based agriculture depends on the climate that prevailed during the particular cropping season. Providing real time weather information to the farmers for making crop management decision can minimize the risk and losses due to extreme climate. In order to improve the accuracy of present weather forecasting, weather information are required at high spatial and temporal resolution. Installation and interlinking of automatic weather stations (AWS) at block level will be helpful to avoid climate risks and increase the productivity by timely weather based agro advisory.

In this context, Government of Tamil Nadu under National Agricultural Development Project (NADP) have funded for establishing Tamil Nadu Agricultural Weather Network (TAWN) by installing 385 AWS, 224 in the first phase and 161 in the second phase. The Agro Climate Research Centre (ACRC), Directorate of Crop Management (DCM), Tamil Nadu Agricultural University (TNAU), Coimbatore in collaboration with Department of Agriculture, Tamil Nadu established the TAWN. In the network, 10 types of agricultural related weather parameters from 385 AWS are collected at hourly interval and hosted in this website. The medium range weather forecast developed by using these weather parameters is also hosted. Using this information, the Agricultural officers will develop weather based agro advisories at block level for the farmers. It is first of its kind in India, such a denser weather network has been established and is expected to help in monitoring the global warming and climate change impacts on Tamil Nadu Agriculture.

Point to Remember – Agriculture Weather Forecasting; Courtesy: Tamil Nadu Agriculture University

MEGHDOOT App

(A Mobile App to Assist Farmers for Weather Based Farm Management)

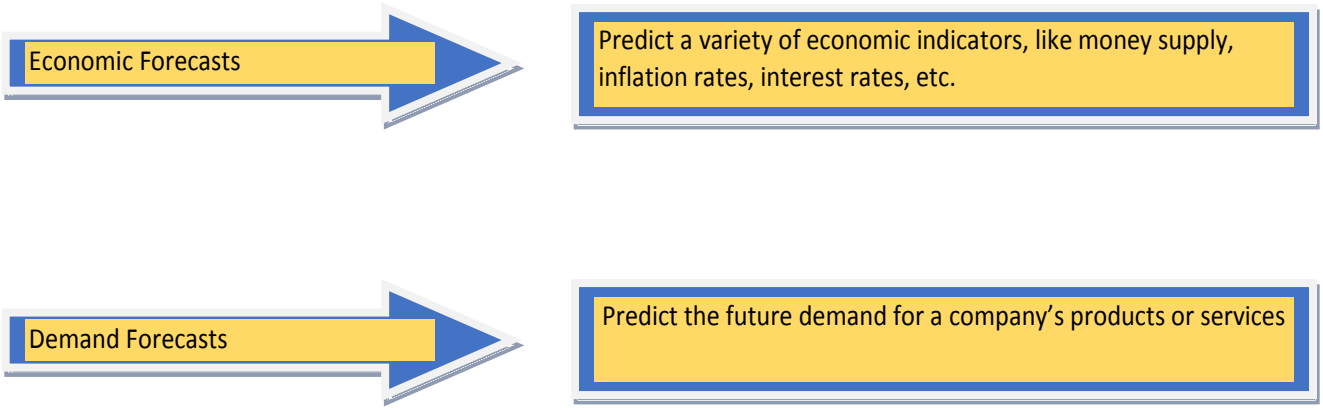
MEGHDOOT mobile app, a joint initiative of IMD, IITM and ICAR aims to deliver Medium range Weather Forecast and weather-based farm advisories to the farmers through a userfriendly mobile application developed by ICRIST. The app seamlessly aggregates contextualized crop wise advisories at district level issued by Agro-Met Field Units (AMFUs) every Tuesday and Friday with forecast and historical weather data for the farmer's benefits. The advisories are also issued in vernacular languages, wherever available.

https://www.imdagrimet.gov.in/node/531793http://www.imdagrimet.gov.in/

Points to Remember

Forecasting Application in Real Life

Some of the forecasting applications in real life are listed as below:



Some of the listed forecasting techniques like Modeling Seasonality and Trend with seasonal index, Moving Average and Weighted Moving Average Forecasting Method, Exponential Smoothing Method, Trend Projection will be discussed.

Explanatory Models

Explanatory models, also called regression models, rely on the identification of related variables that can be used to predict values of the variable of interest. A mathematical relationship is developed between demands. For example, and some other factors that cause demand behavior.

5.5 Different Types of Forecasting Methods

Different types of forecasting methods are represented in the figure 5.3

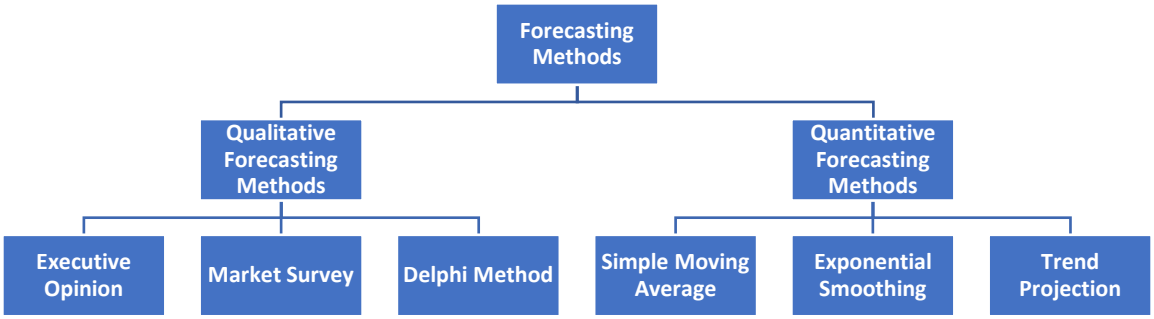


Fig. 5.3. Different Types of Forecasting Methods

Qualitative Forecasting Methods

These types of forecasting methods are based on judgments, opinions, intuition, emotions, or personal experiences and are subjective in nature.

Quantitative Forecasting Methods

These types of forecasting methods are based on mathematical (quantitative) models and are objective in nature.

Different Types of Qualitative Forecasting Methods

Executive Opinion Approach in which a group of managers meet and collectively develop a forecast	**Market Survey** Approach that uses interviews and surveys to judge preferences of customer and to assess demand	**Delphi Method** Approach in which consensus agreement is reached among a group of experts

Time series models look at past patterns of data and attempt to predict the future based upon the underlying patterns contained within those data

.

Different Types of Quantitative Forecasting Methods (Time Series Models)

Simple Moving Average The forecast for next period (period t+1) will be equal to the average of a specified number of the most recent observations, with each observation receiving the same emphasis	**Exponential Smoothing** weighted average procedure with weights declining exponentially as data become older	**Trend Projection** Technique that uses the least squares method to fit a straight line to the data

Decomposition of a Time Series

Patterns may be present in a time series. Different patterns of decomposition of a time series and examples are represented in the figure 5.4 and figure 5.5.

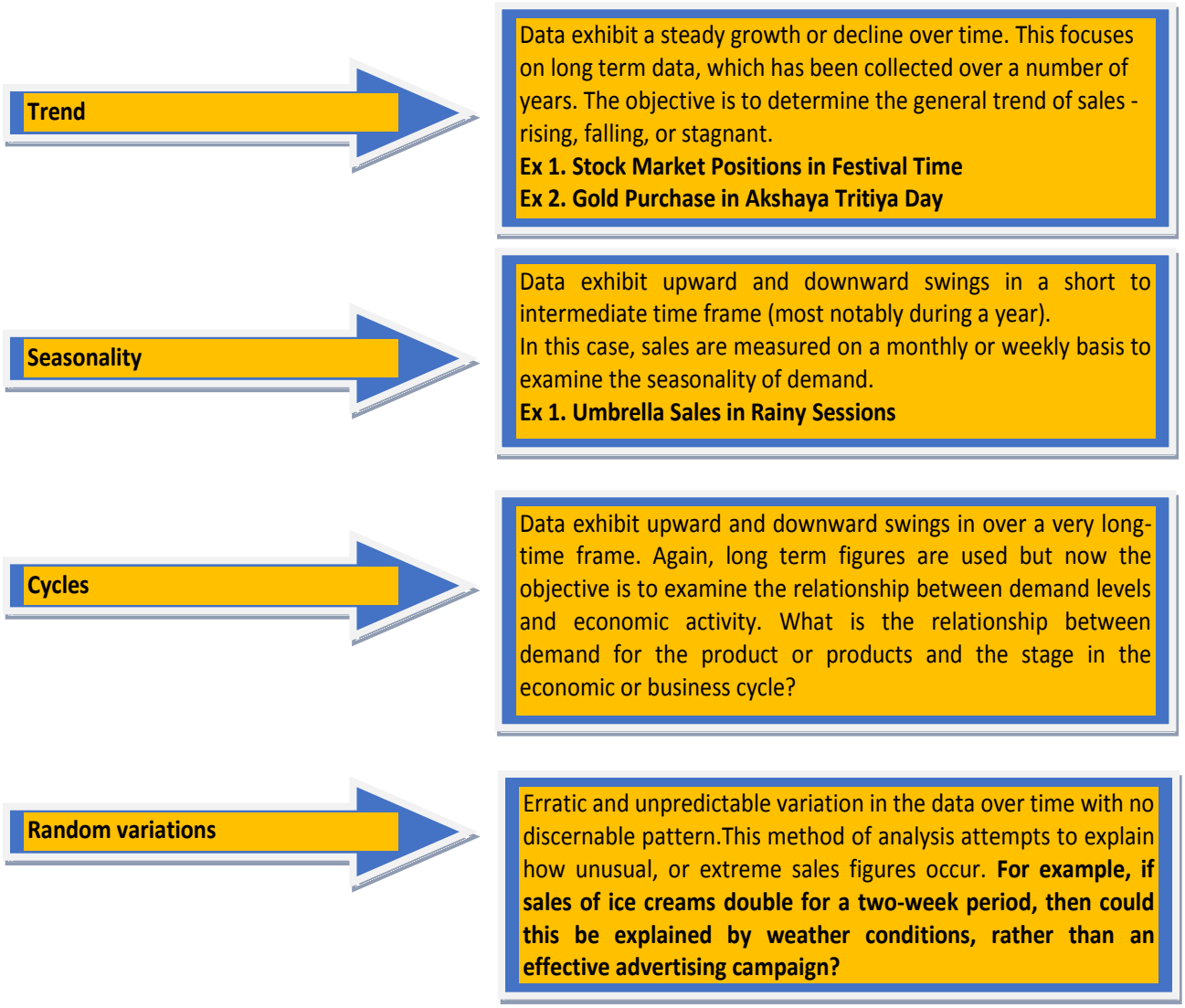


Fig. 5.4. Different Patterns for Decomposition of a Time Series

Seasonal Index:

[1]. A seasonal index of **1.00** for a particular month indicates that the expected value of that month is overall average.

[2]. A seasonal index of **1.50** indicates that the expected value for that month is **50%** greater than overall average.

[3]. A seasonal index of **0. 70** indicates that the expected value for that month is **30%** less than overall

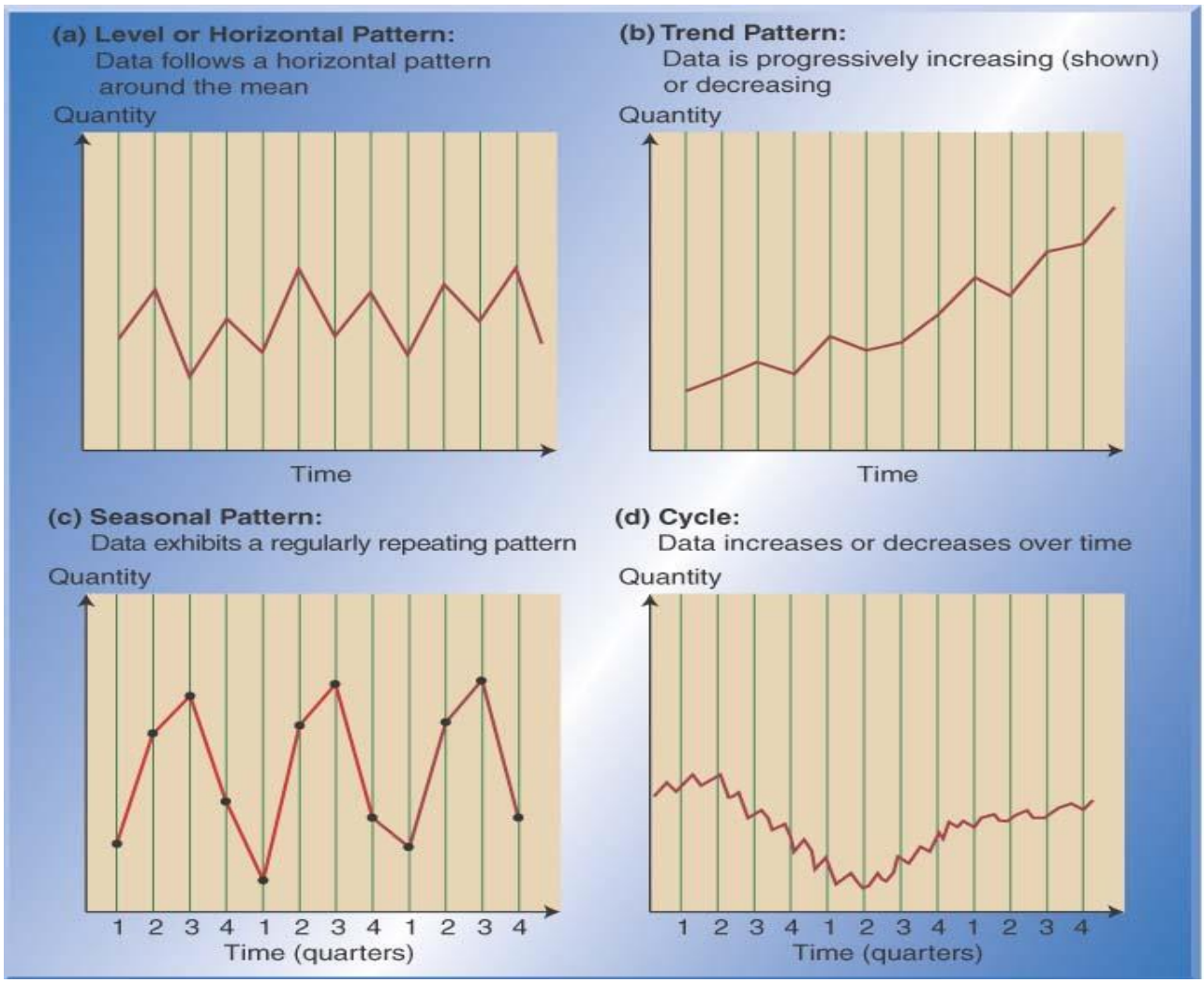


Fig. 5.5.Graphical Patterns for Decomposition of a Time Series

Modeling Seasonality and Trend

Seasonality is a pattern that repeats for each period. For example, annual agriculture product seasonal pattern has a cycle that is 12 periods long, if the periods are months, or 4 periods long if the periods are quarters. We need to get an estimate of the seasonal index for each month, or other periods, such as quarter, week, etc, depending on the data availability.

A seasonal index is how much the average for that particular period trends to be above (or below) the grand average. Therefore, to get an accurate estimate for the seasonal index, we compute the average of the first period of the cycle, and the second period, etc, and divide each by the overall average. The formula for computing seasonal factors is:

$$S_i = D_i/D$$

where:

S_i =Seasonal index for i^{th} period

D_i =Average value for i^{th} period

D = grand average

Problem 5.4:

The following table provides monthly agro product sales in terms of thousand at a agriculture form store. Find the seasonal index

Month → / Trend (quadrant) ↓	Jan	Feb	March	Total
1	3	2	1	6
2	2	4	1	7
3	4	3	3	10
4	2	1	2	5

Solution:

Month → / Trend (quadrant) ↓	Jan	Feb	March	Total
1	3	2	1	6
2	2	4	1	7
3	4	3	3	10
4	2	1	2	5
Mean	2.75	2.5	1.75	7
Seasonal Index	1.18	1.07	0.75	

Seasonal Index for Month:

D(Jan) = (3 + 2 + 4 + 2)/4 = 2.75

D = grand average = (2.75+2.5+1.75)/3 = 2.33

S(Jan) = D(Jan)/D = 2.75/2.33 = 1.18

Similarly, we can able to calculate

S(Feb) = D(Feb)/D = 2.5/2.33 = 1.07

S(March) = D(March)/D = 1.75/2.33 = 0.75

- **Seasonal index for January Month:** A seasonal index of **1.18** indicates that the expected value for that month is **18 %** greater than overall average.
- **Seasonal index for February Month:** A seasonal index of **1.07** indicates that the expected value for that month is **07%** greater than overall average.
- **Seasonal index for March Month:** A seasonal index of **0. 75**indicates that the expected value for that month is **25%** less than overall average.

Moving Average and Weighted Moving Average Forecasting Method

When using a moving average method described before, each of the observations used to compute the forecasted value is weighted equally. In certain cases, it might be beneficial to put more weight on the observations that are closer to the time period being forecast. When this is done, this is known as a weighted moving average technique. The weights in a weighted MA must sum to 1.

$$MA\ (3) = M_{t+1} = [D_t + D_{t-1} + D_{t-2}] / 3$$

$$\text{Weighted MA } (3) = F_{t+1} = wt_1\,(D_t) + wt_2\,(D_{t-1}) + wt_3\,(D_{t-2})$$

Problem 5.5:

The demand for Biomass Briquette Machine for a certain project is given each month as follows:

Month	1	2	3	4	5	6	7	8	9	10
Demand	120	110	90	115	125	117	121	126	132	128

The agriculture officer is asked to forecast the demand for the 11th month using three period moving average technique and weighted moving average technique.

Solution:

The agriculture officer has decided to use a weighting scheme of 0.5, 0.3, 0.2 and calculated the

Moving Average for the 11th month as follows.

MA (3) = $M_{t+1} = [D_t + D_{t-1} + D_{t-2}] / 3$

Weighted MA (3): F11 = (128) + (132) + (126) /3

=128.6

Weighted Moving Average for the 11th month as follows.

Weighted MA (3) = $F_{t+1} = wt_1\,(D_t) + wt_2\,(D_{t-1}) + wt_3\,(D_{t-2})$

The agriculture officer has decided to use a weighting scheme of **0.5, 0.3, 0.2** (Assumed Value) and calculated the weighted moving average for the 11th month as follows.

Weighted MA (3): F11 = 0.5 (128) + 0.3 (132) + 0.2 (126)

= 64 + 39.6 + 25.2

= 128.2

Problem 5.6:

Suppose that the tractor ordered by the district administrative officer for the formers in the past 6 years are as shown in the following table

Year	1	2	3	4	5	6
No of Tractors	150	140	132	137	138	141

The management asked an officer to forecast the number of trackers required next month. The officer knows about the Moving Averages method and calculated the number of tractors required in the next month using the three period moving average as shown below.

Solution:

MA (3) = $M_{t+1} = [D_t + D_{t-1} + D_{t-2}] / 3$

Weighted MA (3): F11 = (141) + (138) + (137) /3

= 138.66

When using a moving average method, notice that each of the observations to compute the forecast value is weighted equally. In certain situations, it might be beneficial to put more weight on the observations that are closer to the time period being forecast. When this is done, it is referred to as the weighted moving average technique. The weights must sum to 1, alleviating the need to divide the total by N (number of observations). Let us apply a weighting scheme of (0.5, 0.3, 0.1) to the above three month moving average example. Note that 0.5 is applied to the most recent data point.

Weighted MA (3): F11 = 0.5 (141) + 0.3 (138) + 0.2 (137)

= 70.5 + 41.4 + 27.4

= 139.7

Find the forecasted value for the fourth month, fifth month, sixth month and seventh month for demand by using moving average and weighted average method

Year	1	2	3	4	5	6
Demand	310	365	395	415	450	465

To Do Activity

Exponential Smoothing Method

The new forecast for next period (period t) will be calculated as follows:

$$F_{t+1} = \alpha D_t + (1 - \alpha) F_t$$

D_t = Actual Value

F_t= Forecasted Value

α = Smoothing factor, which range from 0 to 1

t = Current time period

Problem 5.7:

Suppose we have set of data on the average monthly network traffic on a software module during a year. The data contains the month M and the average level of traffic T as shown below.

Month	1	2	3	4	5	6	7	8	9	10	11	12	13
Traffic	1050	1120	980	1110	1200	900	1040	990	1200	1190	1170	1080	1147

Now let us look at the original problem. Assume $\alpha = 0.6$ and initial forecast of 1000,

Solution:

Month	1	2	3	4	5	6	7	8	9	10	11	12	13
Traffic	1050	1120	980	1110	1200	900	1040	990	1200	1190	1170	1080	1147
Forecasted	1000	1030	1084	1022	1075	1150	1000	1024	1004	1022	1167		

The exponential smoothing value for the 2^{nd} month

$$F_{t+1} = \alpha D_t + (1 - \alpha) F_t$$

= (0.6) (1050) + (1 - 0.6) (1000) = 1030, and the other values were calculated in a similar fashion.

Trend Projection

This method is a version of the linear regression technique. It attempts to draw a straight line through the historical data points in a fashion that comes as close to the points as possible. (Technically, the approach attempts to reduce the vertical deviations of the points from the trend line and does this by minimizing the squared values of the deviations of the points from the line). Ultimately, the statistical formulas compute a slope for the trend line (b) and the point where the line crosses the y-axis (a). The results in the straight-line equation

$$Y = a + b X$$

Where X represents the values on the horizontal axis (time), and Y represents the values on the vertical axis (demand).

Problem 5.8:

Find out the 7^{th} year, 8^{th} year, 9^{th} year and 10^{th} year forecasting value:

Year	1	2	3	4	5	6
Demand	310	365	395	415	450	465

Using regression analysis, we found the relational formula

$$Y = 295 + 30\,X$$

Solution:

This equation can be used to forecast for any year into the future.

Year 7: Forecast = 295 + 30(7) = 505

Year 8: Forecast = 295 + 30(8) = 535

Year 9: Forecast = 295 + 30(9) = 565

Year 10: Forecast = 295 + 30(10) = 595

Problem 5.9:

A linear regression model for import and consumption of pesticides 1981-2002.It reveals the pesticide import during 1981 to 2002. The table shows that a small quantity of 665 metric tons were imported in 1980, on the other hand a huge bulk of 68,804 metric tons were imported in 2002. During, 23 years, a linear trend was observed using regression model. The regression model is given as under:

$$\text{Total import} = -5762.2 + 2632.1\,(\text{year})$$

Find out the 3^{rd} year, 5^{th} year of forecasting value:

Solution:

This equation can be used to forecast for any year into the future.

Year 3: Forecast = - 5762.2 + 2632.1 (3) = 2134.1

Year 5: Forecast = - 5762.2 + 2632.1 (5) = 7398.3

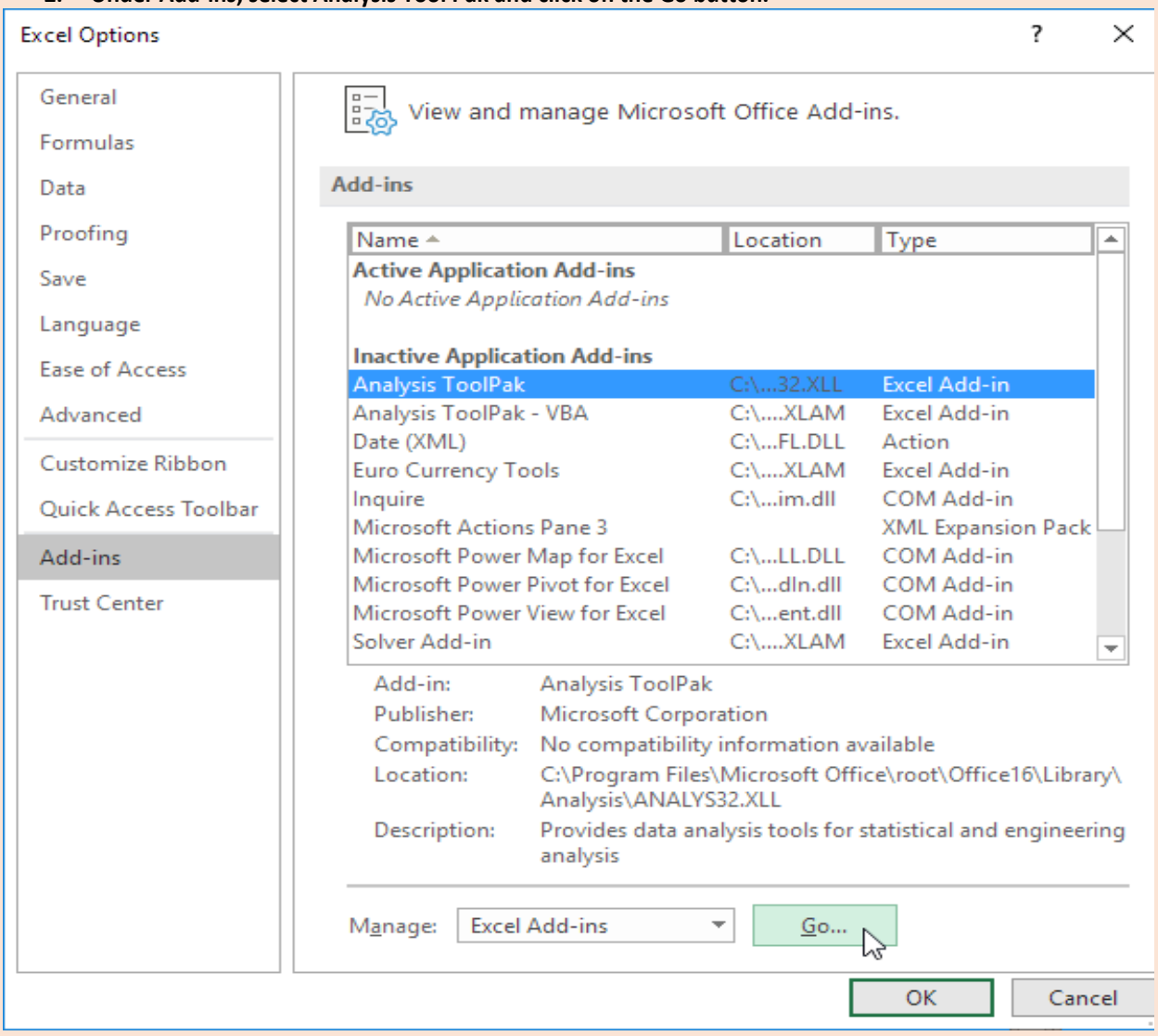
To Do Activity:

Excel – Analytical Tool Practice

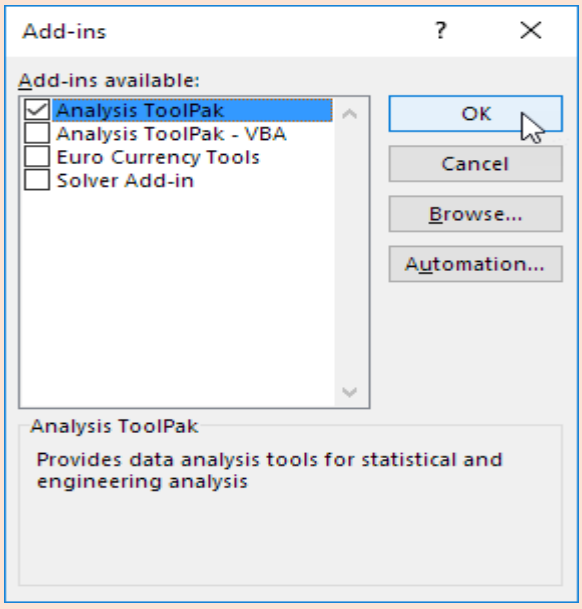
The Analysis Tool Pak is an Excel add-in program that provides data analysis tools for financial, statistical and engineering data analysis.

To load the Analysis Tool Pak add-in, execute the following steps.

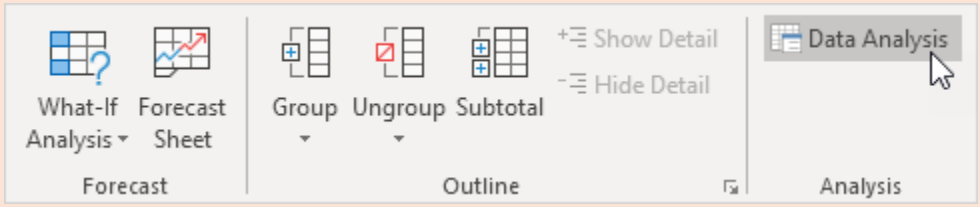
1. **On the File tab, click Options.**
2. **Under Add-ins, select Analysis Tool Pak and click on the Go button.**



3. Check Analysis Tool Pak and click on OK.

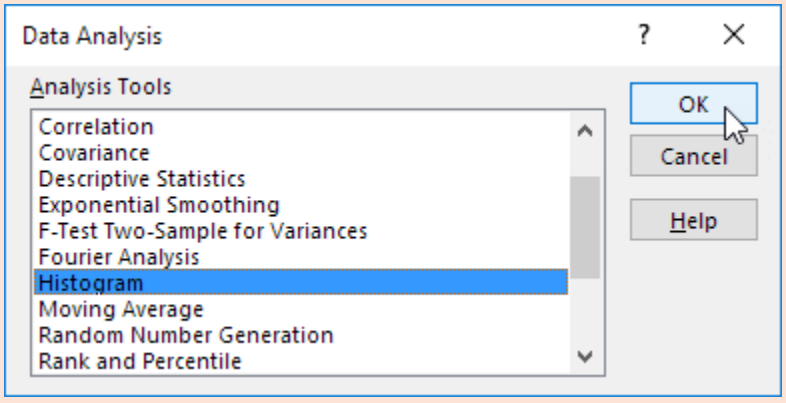


4. On the Data tab, in the Analysis group, you can now click on Data Analysis.



The following dialog box below appears.

5. For example, select Histogram and click OK to create a Histogram in Excel.



Use Analysis Tools, Practice Histogram, Correlation Analysis, Regression Analysis, Moving Average Forecasting with real examples

Summary of the Chapter

1. **Correlation is a statistical measure which determines co-relationship or association of two variables.**
2. **Regression describes how an independent variable is numerically related to the dependent variable.**
3. **Qualitative Forecasting Methods are based on judgments, opinions, intuition, emotions, or personal experiences and are subjective in nature.**
4. **Quantitative Forecasting Methods are based on mathematical (quantitative) models and are objective in nature.**
5. **Formula for Moving Average: MA (3) = $M_{t+1} = [D_t + D_{t-1} + D_{t-2}] / 3$**
6. **Formula for Weighted Moving Average:MA (3) = $F_{t+1} = wt_1 (D_t) + wt_2 (D_{t-1}) + wt_3 (D_{t-2})$**
7. **Formula for Exponential Smoothing Method: $Ft+1 = \alpha D_t + (1 - \alpha) F_t$**
8. **Formula for Trend Analysis and Regression Analysis: Y = a + b X**

Model Questions

Part A

[1]. List out the difference between Correlation and Regression Analysis
[2]. Define: Forecasting
[3]. List out the applications of forecasting
[4]. Write down the formula for Moving Average Forecasting Method
[5]. Write down the formula for Weighted Moving Average Forecasting Method
[6]. Write down the formula for Exponential Smoothing Forecasting Method
[7]. List out the difference between Moving Average Forecasting Method and Weighted Moving Average Forecasting Method
[8]. Define: Seasonal Index
[9]. Write down the formula for Seasonal index
[10]. Write down the formula for Trend Analysis and Regression Analysis
[11]. Define Linear Regression Equation

Part B

[1]. The last five months demand for the fertilizer materials for is shown below. Find out the 6^{th} month demand for the fertilizer materials using Moving Average Forecasting Method and Weighted Moving Average Forecasting Method?

Month	1	2	3	4	5
Demand	13	17	19	23	24

[2]. From the records of previous orders, management has accumulated the following data for the past 10 months:

Month	Jan	Feb	March	April	May
Orders	120	90	100	75	110
Month	June	July	August	September	October
Orders	50	75	130	110	90

Compute the monthly demand forecast for November through Exponential Smoothing Forecasting. Use weights of 0.5, 0.3, and 0.2, with the heavier weights on the more recent months.

[3]. Elaborate different types of forecasting techniques with examples

[4]. The last five months demand for the fertilizer materials for is shown below. Find out the 6^{th} month demand for the fertilizer materials using regression model?

Month	1	2	3	4	5
Demand	13	17	19	23	24

[5]. The table shows the quarterly sales for gel making machines. The sales in Lakhs. Determine the seasonal index

Season / Year	winter	spring	summer	fall
2010	2	4	3.3	2.1
2011	2.2	1.8	1.6	1.7
2012	3	2	1.1	1.2

References

Chapter 1

[1]. Chess prodigy Joshua Waitzkin (2007). The Art of Learning: A journey in the Pursuit of Excellence, Wiley

[2]. GalitShmueli, Nitin R. Patel, and Peter C. (2010). Data Mining for Business Intelligence, Wiley

[3]. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran (2015). Quantitative Methods for Business, CENGAGE Learning Custom Publishing

[4]. I. Levin Richard, H. SiddiquiMasood, S. Rubin David, Rastogi Sanjay (2017). Statistics for Management Paperback, Pearson Education

[5]. R.N. Prasad, SeemaAcharya (2011). Fundamentals of Business Analytics, Wiley

[6]. T N Srivastava, ShailajaRego (2015). Statistics for Management, Tata McGraw Hill Publishing Company Limited Press.

Chapter 2

[1]. Seymour Lipschutz , John J. Schiller (2017). Schaum's Outline: Introduction to Probability and Statistics, McGraw Hill Education

[2]. HenkTijms (2007). Understanding Probability, second edition, Cambridge University Press

[3]. Ramesh Chandra (2014). Permutation and Combinations Paperback, Notion Pres

Chapter 3

[1]. Sharma J K (2014). Business Statistics, Vikas publishing house: Uttar Pradesh.

[2]. Raghavachari (2004). M-Mathematics for management, Tata McGraw Hill: New Delhi.

[3]. Sancheti , D.C., and Kapoor, V.K (1982). -Business Mathematics, Sultan Chand &Sons: New Delhi.

[4]. Noah Iliinsky, Julie Steele (2010). Designing Data Visualizations, O'Reilly Media

[5]. Stevens, S. S. (1959). Measurement. In C. W. Churchman, ed., Measurement: Definitions and Theories, pp. 18-36. New York: Wiley. Reprinted in G. M. Maranell, ed., (1974) Scaling: A Sourcebook for Behavioral Scientists. Chicago, Aldine: 22-41.

Chapter 4

[1]. John Carratello, Patty Carratello (2010). Data Handling: Charts, Graphs and Diagrams, Teacher Created Resources.

[2]. Bill Jelen (2010). Charts and Graphs: Microsoft Excel 2010 (MrExcel Library), Que Publishing

[3]. Kristen Sosulski (2018). Data Visualization Made Simple: Insights into Becoming Visual Hardcover, Routledge

[4]. Claus O. Wilke (2018). Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures, O'Reilly Media

Chapter 5

[1]. Khanna. O.P (2018). Industrial Engineering –DhanpatRai Publications.

[2]. Stephen N, Chapman (2006). Fundamentals of Productions Planning and Control, Pearson Edition.

[3]. MarthandTselang (2018). Industrial Engineering and Production Management, S. Chand & Co.

Further Reading Materials

Chapter 1

[1]. National Council of Educational Research and Training Study Material : http://ncert.nic.in/ncerts/l/keep209.pdf

[2]. Coursera: Business Statistics and Analysis
https://www.coursera.org/specializations/business-statistics-analysis

[3]. NPTEL: Business Statistics
https://nptel.ac.in/courses/110107114/

Chapter 2

[1]. MIT 6.041SC Probabilistic Systems Analysis and Applied Probability course, which is available in MIT Open Cours Ware for free :
https://www.youtube.com/playlist?list=PLUl4u3cNGP60A3XMwZ5sep719_nh95qOe

Chapter 3

[1]. The Tableau Platform
https://www.tableau.com/products

[2]. 12 great books about data visualization
https://www.tableau.com/learn/articles/books-about-data-visualization

[3]. How to Make a Chart or Graph in Excel [With Video Tutorial]
https://blog.hubspot.com/marketing/how-to-build-excel-graph

Chapter 4

[1]. 10 Different kinds of Graphs for your Data
https://www.udemy.com/blog/different-kinds-of-graphs/

Chapter 5

[1]. NPTEL Course Material:
Linear Regression Analysis and Forecasting Course
Reference: https://nptel.ac.in/courses/111104098/#

[2]. MIT, USA Course Material:
Demand Forecasting, Planning, and Management
Reference: https://ocw.mit.edu/courses/engineering-systems-division/esd-260j-logistics-systems-fall-2006/lecture-notes/lect6.pdf

Editors' Profile

Dr W G Prasanna Kumar

Dr. W. G. Prasanna Kumar, PhD in Education with basic degree in Social Work and Master's Degrees in Sociology, Public Administration and Political Science has professional education in Environmental Economics, Public Relations, Communication and Training and Development. Presently Chairman, Mahatma Gandhi National Council of Rural Education (MGNCRE) under the Ministry of Human Resource Development, in Government of India strives to promote resilient rural India through Higher Education interventions. The national initiative of reviving Mahatma Gandhi's ideas of NaiTalim, spearheaded by Dr. W G Prasanna Kumar, has met unprecedented success at both national and state levels. The primary objective of this initiative is to promote Gandhiji's ideas on Experiential Learning, NaiTalim, Work Education and Community Engagement, and mainstreaming them in School Education and Teacher Education Curriculum & Pedagogy. As Professor and Head Centre for Climate Education and Disaster Management in Dr MCR HRD Institute, conducted several capacity building and action research programmes in climate education, disaster management and crowd management. He has handled many regional, national and international environmental education programmes and events including UN CoP11 to Convention on Biological Diversity and Media Information Management on Environmental Issues.

He was Director in National Green Corps in the State Government for over 11 years and Senior Social Scientist in State Pollution Control Board for 6 years. Conducted various curriculum and non- curriculum related training programmes in environmental education. He was a Resource Person for AP Judicial Academy, AP Police Academy, AP Forest Academy, EPTRI, Commissonerate of Higher Education and Intermediate Education, State Council for Educational Research and Training and National Council for Educational Research and Training New Delhi, CCRT, Bharathiya Vidyapeet University Pune, CPR Environmental Education Centre Chennai and Centre for Environment Education Ahmedabad. Dr W G Prasanna Kumar was trained in Community Consultation for Developmental Projects in EPA Victoria Australia in 1997 trained as State Chief Information Officer by IIM Ahmedabad and MCRHRDI Government of Andhra Pradesh in 2004 and trained in Environmental Education and Waste Management Technique by JICA, Japan in 2011.

He was awarded Best State Nodal Officer of National Green Corps Award from Centre for Science and Environment, New Delhi, 2008, Jal Mithra Award from Earthwatch Institute of India and Water Aid New Delhi, 2014 and Certificate of Commendation for the services in UN Conference of Parties to Convention for Biodiversity conducted at Hyderabad from 1-20 October 2012 by the Government of Andhra Pradesh 2012.

Dr K N Rekha

Dr K N Rekha, is a PhD Graduate from IIT Madras. She has 14 years of experience in training and education Industry. She works at Mahatma Gandhi National Council of Rural Education (MGNCRE), Hyderabad as Senior Faculty. She is involved in curriculum development on Rural Management and Waste Management. Prior to this, she worked as a researcher at Indian School of Business, Hyderabad, a short stint at Centre for Organisation Development (COD), Hyderabad. She has co-authored a book on "Introduction to Mentoring", written book chapters, peer reviewed research papers, book reviews, Case studies, and caselets in the area of HR/OB. She also presented papers in various national and international conferences. Her research areas include Mentoring, Leadership, Change Management, and Coaching. She was also invited as a guest speaker at prominent institutions like IIT Hyderabad.

Subject Authors' Profile

Mr. Kaja Bantha Navas Raja Mohamed

Mr. Kaja Bantha Navas Raja Mohamed, works as an Assistant Professor in the School of Mechanical Engineering at Sathyabama Institute of Science and Technology, India. He received his Engineering Master's in Industrial Engineering from Thiagarajar College of Engineering in 2007. His research interests are Business Analytics, Data Analytics, Computational Intelligent, Industrial Engineering, Design of Experiments, Optimization, Mathematical Modelling and Machine Learning Techniques. He has wide knowledge and experience in modelling and optimization using various tools and software packages like SPSS, SAS, Minitab, Design Expert, MATLAB etc. He has published more than 40 papers in refereed international journals and conference proceedings. He got seven Industrial Design Patents grants and eight patents published. He is a CMI Level 5 Certified Profession for Change Management - Technical Leadership Development Program funded by AICTE UK India Education and Research Initiative (UKIERI) and one of the Mentor of Change for Atal Tinkering Labs as part of Atal Innovation Missions, NITI Aayog, Govt of India. Recently he received Global Engineer Leadership Scholarship for Summer Research School, National TsingHua University, Taiwan.

Ms. S. Sharmila Parveen

Ms. S. Sharmila Parveen, works as an Assistant Professor in Department of Business Administration at Nazareth College of Arts and Science, Chennai. She has completed her MBA in College of Engineering, Guindy. Her specializations are Finance and Systems. She has qualified both NET and SLET examination. She has done various case studies in Agriculture and Service Sectors fields.

सत्यमेव जयते

Mahatma Gandhi National Council of Rural Education (MGNCRE)

Department of Higher Education

Ministry of Human Resource Development, Government of India

Where there is Rural Wellbeing
there is Universal Prosperity

सत्यमेव जयते

Mahatma Gandhi National Council of Rural Education
Department of Higher Education
Ministry of Education, Government of India

Where there is Rural Wellbeing
there is Universal Prosperity

040 - 2321 2120

admin@mgncre.in
www.mgncre.in

#5-10-174, Shakkar Bhavan, Fateh Maidan Lane
Band Colony, Basheer Bagh,
Hyderabad-500004