

QUANTITATIVE TECHNIQUES

COURSE CODE: M26BS04DC

Discipline Core Course
Master of Business Administration
Self Learning Material

SREENARAYANAGURU
OPEN UNIVERSITY

SREENARAYANAGURU OPEN UNIVERSITY

The State University for Education, Training and Research in Blended Format, Kerala

SREENARAYANAGURU OPEN UNIVERSITY

SGOU

Vision

To increase access of potential learners of all categories to higher education, research and training, and ensure equity through delivery of high quality processes and outcomes fostering inclusive educational empowerment for social advancement.

Mission

To be benchmarked as a model for conservation and dissemination of knowledge and skill on blended and virtual mode in education, training and research for normal, continuing, and adult learners.

Pathway

Access and Quality define Equity.

Quantitative Techniques

Course Code: M26BS04DC

Semester - I

Discipline Core Course
Master of Business Administration
Self Learning Material

(With Model Question Paper Sets)

SGOU

SREENARAYANAGURU
OPEN UNIVERSITY

SREENARAYANAGURU OPEN UNIVERSITY

The State University for Education, Training and Research in Blended Format, Kerala

SREENARAYANAGURU
OPEN UNIVERSITY

QUANTITATIVE TECHNIQUES

Course Code: M26BS04DC
Semester- I
Discipline Core Course
Master of Business Administration

Academic Committee

Prof. (Dr.) S. Santhosh Kumar
Dr. R. Vasanthagopal
Dr. B. Vijayachandran Pillai
Dr. Hareesh N Ramanathan
Dr. Sreesha C.H.
Dr. Renjni D.
Dr. Sangeetha K. Pratap
Dr. Prashanth M.K.
Dr. Devi Soumyaja

Development of the Content

Dr. Surya Robert

Review and Edit

Radha Krishna Pillai G.

Proofreading

Dr. Nisa James

Scrutiny

Radha Krishna Pillai G.
Dr. Surya Robert

Cover Design

Jobin J.

Co-ordination

Dr. Bino Joy
HoS, School of LBS

Scan this QR Code for reading the SLM on a digital device.

Edition
April 2026

Copyright:
© Sreenarayanaguru Open University

ISBN 978-81-686027-9-3

9 788168 602793

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from Sreenarayanaguru Open University. Printed and published on behalf of Sreenarayanaguru Open University by Registrar, SGOU, Kollam.

www.sgou.ac.in

Visit and Subscribe our Social Media Platforms

MESSAGE FROM VICE CHANCELLOR

Dear Learner,

It gives me immense pride and heartfelt satisfaction to welcome you to Sreenarayanaguru Open University—a transformative space where aspirations are nurtured, potential is unlocked, and futures are shaped with purpose.

Established in September 2020 as a pioneering state initiative, the University stands as a symbol of inclusive, accessible, and high-quality higher education. Anchored in our guiding philosophy—"access and quality define equity"—we are committed not merely to expanding educational opportunities, but to ensuring that every learner receives an education that is meaningful, relevant, and transformative.

Our journey is deeply inspired by the timeless vision of Sreenarayana Guru, whose ideals of knowledge, equality, and social empowerment continue to illuminate our path. His philosophy reminds us that education is not just a means to an end, but a powerful instrument for personal growth, societal progress, and human dignity.

In a rapidly evolving and digitally driven world, we recognize that learning must be flexible, dynamic, and future-ready. Our thoughtfully designed blended learning ecosystem—integrating Self-Learning Materials, Classroom Counselling, and Virtual Learning Platforms—ensures that you experience the best of both independence and guided academic engagement. This model empowers you to learn at your own pace while staying meaningfully connected to academic support systems.

The MBA programme at Sreenarayanaguru Open University is not just a degree—it is a gateway to leadership, innovation, and professional excellence. The curriculum is meticulously curated to reflect the realities of the contemporary global business landscape, seamlessly blending foundational management principles with cutting-edge developments in business and technology. You will explore and engage with:

- Strategic management and entrepreneurial thinking
- Digital transformation, FinTech, and emerging technologies
- Logistics, supply chain, and global business ecosystems
- Consumer behaviour and evolving market intelligence
- Human resource management in the age of Artificial Intelligence
- Ethical leadership and sustainable decision-making

Designed with a multidisciplinary and future-oriented perspective, the programme offers diverse specializations, skill enhancement pathways, and exposure to innovative domains that align with industry expectations. It is structured to strengthen your employability, entrepreneurial mindset, and leadership capabilities—equipping you to excel in competitive and complex professional environments.

Our Self-Learning Materials are crafted with exceptional care, combining conceptual clarity with real-world relevance through case studies, contemporary examples, and practical insights. They are designed not only to inform but to inspire critical thinking, problem-solving, and adaptability—skills that are essential in today's disruptive and ever-changing business world.

At the heart of our mission lies an unwavering commitment to learner success. Our comprehensive learner support system is always within your reach, ensuring that your academic journey remains smooth, enriching, and fulfilling. We encourage you to actively engage, explore, question, and grow—because your success is the true measure of our purpose.

As you begin this important journey, I invite you to dream boldly, learn continuously, and lead responsibly. Let this programme be a stepping stone towards achieving your ambitions and making a meaningful impact in the world.

May your journey here be one of discovery, empowerment, and excellence.
With my best wishes for your success and lifelong learning,



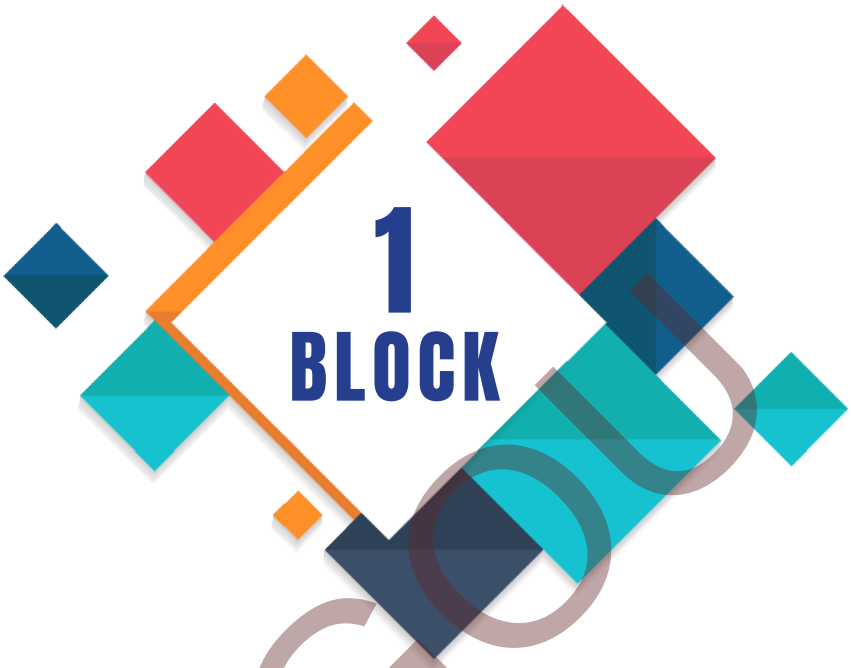
Prof. (Dr.) Jagathy Raj V. P.
Vice Chancellor

SGOU

Contents

BLOCK	**1**	**Overview of Quantitative Techniques**	**1**
Unit	1	Introduction to Quantitative Techniques	2
Unit	2	Probability Distribution	15
BLOCK	**2**	**Parametric and Non-Parametric Tests**	**55**
Unit	1	Fundamentals of Hypothesis Testing and Parametric Tests	56
Unit	2	Non-Parametric Tests and Their Applications	105
BLOCK	**3**	**Correlation and Regression Analysis**	**129**
Unit	1	Correlation	130
Unit	2	Regression Analysis	157
BLOCK	**4**	**Time Series Analysis and Index Numbers**	**178**
Unit	1	Time Series Analysis	179
Unit	2	Index Numbers	203
Model Question Paper Sets			234
Appendix Tables			241

SGOU



Overview of Quantitative Techniques

Unit 1

Introduction to Quantitative Techniques

Learning Outcomes

On completion of this unit, learners will be able to:

- ◊ introduce quantitative techniques and its relevance in solving complex business problems with analytical precision
- ◊ familiarise the core objectives of quantitative methods, including improving decision-making, enhancing resource allocation, and enabling predictive analysis
- ◊ make them aware of classification of Quantitative Techniques and its application in business and management

Background

Quantitative Techniques evolved from simple statistical tools into a comprehensive set of mathematical and analytical methods that transformed decision-making in business, economics, and social sciences. In the early stages, quantitative methods were primarily rooted in classical statistics and mathematics, focusing on measurement, probability, and optimisation. The 1920s and 1930s saw the rise of survey-based research and statistical sampling, which gave organisations a systematic way to analyse markets and performance. By the mid-20th century, especially during the Quantitative Revolution of the 1950s and 1960s, disciplines such as geography, economics, and management shifted from descriptive approaches to analytical and model-based techniques. This period marked the integration of linear programming, operations research, and econometrics, enabling managers and policymakers to apply mathematical models for resource allocation, forecasting, and strategic planning.

From the 1980s onwards, the evolution of quantitative techniques accelerated with the advent of computing and digital technologies, which allowed for complex simulations, regression modelling, and large-scale data analysis. The

SGOU

rise of big data, artificial intelligence, and machine learning further expanded the scope of quantitative methods, making them indispensable for evidence-based decision-making in areas such as finance, logistics, healthcare, and marketing. Today, quantitative techniques are not only about statistical analysis but also about predictive modelling, optimisation, and real-time analytics, reflecting their continuous adaptation to technological progress and global challenges. This unit explores the foundational concepts, goals, practical uses, and constraints of quantitative techniques, aiming to empower learners with analytical skills essential for effective decision-making in management.



Keywords

Quantitative Techniques, Decision Making, Forecasting, Limitations



Discussion

A manufacturing company producing consumer electronics was struggling to decide how many units of a new smartphone model to launch in the upcoming festive season. Instead of relying on intuition, management applied quantitative techniques in decision-making and forecasting. They used time-series forecasting to analyse past sales trends, regression models to study the impact of advertising spend and seasonal demand, and probability analysis to estimate risks of overproduction or shortages. Based on these insights, the managers decided to increase production moderately, aligning output with predicted demand while avoiding excess inventory. When the season arrived, sales closely matched the forecast, resulting in high customer satisfaction and reduced holding costs. This illustrates how forecasting and decision-making methods in quantitative techniques help managers make rational, data-driven choices that balance risk, cost, and opportunity.

Quantitative techniques have become essential in modern decision-making, helping businesses and organisations optimise productivity, reduce costs, and align operations with specific goals. These methods translate complex industrial and managerial issues into numerical models, offering structured and objective solutions. Their growing relevance has captured the attention of students, professionals, and policymakers alike. With the backing of scientific methods, these techniques have repeatedly proven effective in addressing various national and organisational challenges, ensuring accuracy and timeliness in outcomes. This approach supports every step of the decision-

SGOU

making process, from defining problems and setting constraints to building, solving, and validating models based on human behaviour and operational data.

1.1.1 Meaning

Quantitative techniques are analytical tools grounded in numerical data, offering managers a structured and practical approach to decision-making. These scientific methods help dissect complex business problems, enabling decision makers to analyse various scenarios and assess outcomes with clarity and precision. Organisations are better equipped to address issues systematically rather than relying on intuition or trial-and-error by employing such techniques.

Decision makers can develop, evaluate, and implement strategies that align with specific organisational goals through quantitative methods. These tools not only enhance problem-solving accuracy but also streamline the decision-making process. Quantitative techniques have become indispensable to modern management, offering a reliable foundation for making informed and strategic choices.

Example: A retail store relies on historical sales records to determine the quantity of stock required for the upcoming month. The manager gathers numerical information such as daily sales figures, seasonal demand patterns, and current inventory levels, and then applies statistical tools, including averages, trend analysis, and forecasting models, to predict future demand. By adopting these quantitative approaches, the manager can make precise and objective decisions, minimising the risk of excess stock or shortages, through data-driven analysis rather than relying on intuition.

1.1.2 Definition

Quantitative techniques are systematic, mathematical, and statistical methods used to analyse numerical data and support decision-making, especially in business and management contexts. These techniques help formulate problems, analyse alternatives, and select optimal solutions based on measurable data.

According to ***C.R. Kothari***, "Quantitative techniques may be defined as those techniques which provide the decision maker with a systematic and powerful means of analysis and help, based on quantitative data, in exploring policies for achieving predetermined goals".

"Quantitative techniques refer to those statistical and mathematical methods which are used in analyzing business and economic problems for decision-making." - ***D.R. Anderson***

"Quantitative techniques are a collection of mathematical and statistical tools used for the systematic analysis of data in order to help in decision-making under conditions of certainty and uncertainty." - ***S.P. Gupta***

"Quantitative techniques involve the use of mathematical models, statistical methods, and optimization techniques for decision-making in business and economics." - ***W.M. Harper***

SGOU

"Quantitative techniques are scientific methods that use quantitative data, mathematical models, and statistical tools to analyze and solve managerial problems." - ***N.D. Vohra***

1.1.3 Objectives of Quantitative Techniques

Quantitative techniques are powerful tools that help decision-making by applying mathematical and statistical methods. The following are some of the key objectives.

◊ **Facilitate Scientific Decision-Making**

Provide a logical, data-driven foundation for solving complex problems, especially in business, economics, and operations.

◊ **Optimise Resource Allocation**

Help determine the most efficient use of limited resources, time, money, staffing, or materials.

◊ **Minimise Costs and Maximise Profits**

Support cost-reduction strategies and profit-maximising decisions through models like linear programming and inventory control.

◊ **Improve Operational Efficiency**

Streamline processes by reducing waiting times, processing delays, and service bottlenecks (e.g., using queuing theory or simulation).

◊ **Support Forecasting and Planning**

Enable accurate predictions of future trends using time series analysis, regression, and probability models.

◊ **Aid in Risk and Uncertainty Management**

Provide tools like decision trees and game theory to evaluate alternatives under risk or uncertainty.

◊ **Enhance Quality Control**

Use statistical quality control techniques to monitor and improve product or service quality.

◊ **Enable Optimal Scheduling and Sequencing**

Assist in determining the best order of operations or job assignments to minimise total time or cost.

1.1.4 Classification of Quantitative Techniques

Quantitative techniques come in various forms and can be broadly grouped into two main categories:

SGOU

1. Mathematical Techniques
2. Statistical Techniques

A. Mathematical Techniques

Mathematical quantitative techniques involve systematically using numerical data and mathematical models to analyse patterns, predict outcomes, and optimise decisions. These techniques are widely applied in business, economics, and public policy to ensure data-driven planning, resource allocation, and strategic forecasting. These techniques typically include:

1. Permutations and Combinations

Permutations refer to how items can be arranged in a specific sequence. Such arrangements depend on the number of items and how many are selected at a time. These arrangements are calculated using established mathematical formulas.

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

A combination refers to selecting or grouping items where the arrangement or order does not matter. The total number of such groupings is determined using a specific mathematical formula.

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

2. Set Theory

Set theory provides a foundational mathematical framework for analysing and solving complex problems by organising elements into collections called sets. It allows researchers to systematically categorise, compare, and manipulate data, using operations like union, intersection, and difference to reveal underlying relationships. Widely applied in decision analysis, logic, data science, and economics, set theory helps structure analytical reasoning and model real-world scenarios with clarity and precision.

3. Matrix Algebra

Matrix algebra is a mathematical approach that organises numbers or variables into grids called matrices, made up of rows and columns. These structures allow for efficient operations such as addition, multiplication, inversion, and finding determinants, all of which help reveal patterns and solutions in complex problems. In fields like economics, computer science, and operations research, matrix algebra simplifies the modelling and analysis of large systems of equations and interconnected variables.

SGOU

4. Determinants

Determinants are scalar values derived from square matrices that help assess the solvability of systems of linear equations. If the determinant of a matrix is non-zero, the system has a unique solution; if it is zero, the system may be dependent or unsolvable. Beyond solving equations, determinants also play a key role in evaluating matrix invertibility and understanding properties like area, volume, and transformation effects in higher mathematics.

5. Differentiation

Differentiation is a fundamental concept in calculus that analyses how a function changes as its input (independent variable) changes. It helps determine the rate at which one quantity responds to shifts in another, often visualised as the slope of a curve at a given point. Commonly used in economics, engineering, and optimisation, differentiation enables precise modelling of dynamic systems, such as cost variations, speed, or marginal utility.

6. Integration

Integration is a fundamental operation in calculus that calculates the total accumulation of a quantity, often visualised as the area beneath a curve on a graph. It reverses the differentiation process by determining the original function from its rate of change. Typical applications in economics and engineering include finding distance from velocity, total cost from marginal cost, and evaluating resource consumption over time.

7. Differential Equations

Differential equations are mathematical formulas that involve derivatives, showing how one quantity changes for another over time or space. They are essential for modelling dynamic systems in physics, economics, biology, and engineering, capturing population growth, investment returns, or fluid motion.

B. Statistical Techniques

Statistical techniques are structured methods used to investigate, analyse, and draw conclusions about specific phenomena based on data. They cover the entire analytical workflow, including designing surveys or experiments, collecting and organising data, and applying regression analysis, hypothesis testing, or probability modelling. These techniques enable researchers, policymakers, and business professionals to make evidence-based decisions, uncover trends, and validate assumptions in real-world contexts. Key components include:

1. Data Collection

Data collection is essential to gathering relevant information to support statistical analysis, research, or decision-making. It encompasses techniques for acquiring primary data, such as surveys, interviews, and experiments, and secondary data from existing

SGOU

sources like reports, databases, or previous studies, ensuring a solid foundation for reliable insights.

2. Measures of Central Tendency, Dispersion, Skewness, and Kurtosis

- ◊ **Central Tendency:** Central tendency refers to statistical measures, such as the mean, median, and mode, that identify the centre or most typical value of a dataset. These metrics help summarise and compare datasets by providing a single representative value.
- ◊ **Dispersion:** Dispersion measures describe how much the data values vary or spread around the central value. Techniques like range, variance, and standard deviation help assess data consistency and identify patterns of variability.
- ◊ **Skewness:** Skewness quantifies the asymmetry of a distribution curve around its mean, indicating whether the data is stretched more to one side. A positively skewed distribution has a longer tail to the right, while a negatively skewed one stretches to the left.
- ◊ **Kurtosis:** Kurtosis evaluates the sharpness or flatness of a distribution's peak relative to a standard curve. High kurtosis signifies heavy tails and extreme outliers, whereas low kurtosis reflects flatter, more uniform distributions.

3. Correlation and Regression Analysis

- ◊ **Correlation:** Correlation measures how strongly two or more variables are related and whether that relationship is positive, negative, or neutral. It does not imply causation but helps identify associations, like sales increasing with advertising spend.
- ◊ **Regression:** Regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It enables prediction and estimation, such as forecasting future sales based on past trends in price, promotion, and consumer behaviour.

4. Index Numbers

Index numbers are statistical tools that summarise and compare changes in economic variables like prices, production, and income over time. Standardising data to a base period allows analysts to track trends and inflationary pressures across sectors. These indices serve as economic indicators, helping policymakers and businesses interpret market dynamics and plan strategically.

5. Time Series Analysis

Time series analysis examines data points collected or recorded at successive intervals to identify underlying patterns, trends, and seasonal effects. It helps understand how variables evolve and forecast future behaviour. Typical applications include financial market analysis, demand forecasting, and monitoring economic indicators.

SGOU

6. Interpolation and Extrapolation

Interpolation estimates missing values within the range of existing data by assuming a smooth and consistent trend. Extrapolation projects values beyond the current dataset based on the same pattern, extending predictions into the future or unknown intervals. These methods are helpful in fields like economics, environmental studies, and engineering where incomplete data must be interpreted meaningfully.

7. Statistical Quality Control (SQC)

SQC employs statistical methods to monitor, control, and improve manufacturing or service quality by analysing process variability. Techniques such as control charts, sampling plans, and process capability analysis help detect whether variations are random or due to assignable causes. It ensures compliance with standards, reduces defects, and enhances operational efficiency.

8. Ratio Analysis

Ratio analysis evaluates a firm's financial health and performance using metrics derived from its balance sheet, income statement, and cash flow data. Key ratios, like liquidity, profitability, and solvency, provide insight into operational efficiency and risk exposure. This analysis aids investors, managers, and lenders in making informed decisions.

9. Probability Theory

Probability theory is a mathematical framework used to measure the likelihood of different outcomes under uncertainty. It supports predictive modelling and risk analysis through concepts like random variables, distributions, and expected values. The theory underpins decision-making in insurance, finance, and quality control domains.

10. Hypothesis Testing

Hypothesis testing is a statistical procedure used to validate assumptions or claims about a population by analysing sample data. It involves setting up a null and alternative hypothesis, selecting a significance level, and using test statistics to determine acceptance or rejection. This method ensures evidence-based conclusions and strengthens the credibility of research findings.

1.1.5 Application in Business and Management

The application of quantitative techniques in business and management is crucial for ensuring rational and scientific decision-making in today's competitive environment, particularly when dealing with conditions of certainty, risk, or uncertainty. Under certainty, managers rely on precise numerical data and deterministic models to guide their choices; in situations of risk, they employ probability and expected value analysis to assess possible outcomes; and when faced with uncertainty, they turn to decision rules such as maximin, minimax, and regret criteria. Forecasting and demand estimation also represent a key application area, where methods such as time series analysis, regression,

SGOU

and trend projection are utilised to predict future sales and market demand. In this way, quantitative techniques minimise reliance on intuition, enhance accuracy, and enable managers to make logical, forward-looking decisions.

a. Decision Making Under Certainty, Risk, and Uncertainty

Quantitative techniques serve as foundational tools in managerial decision-making, helping navigate scenarios with varying levels of predictability. Under certainty, managers use methods such as linear programming to allocate resources optimally when outcomes are known. In risky environments, tools like decision trees and expected monetary value analysis assist in evaluating different courses of action by factoring in the Probability and payoff of each option. When facing uncertainty, approaches like sensitivity analysis or scenario modelling help assess how changes in key variables influence potential outcomes. These techniques replace intuition with data-backed logic, enabling businesses to make more consistent and rational choices while balancing profitability, risk tolerance, and strategic goals.

b. Forecasting and Demand Prediction

Forecasting is crucial for aligning operations with future market behaviour, and quantitative tools play a vital role in this process. Techniques like time series analysis, regression models, moving averages, and exponential smoothing allow businesses to analyse historical patterns and project future demand. Accurate forecasts improve production planning, inventory management, workforce scheduling, and financial budgeting. For example, retail companies use sales data to estimate demand and prepare seasonal inventory, while manufacturers rely on economic indicators and trend analysis to plan production volumes. These techniques enhance overall agility and cost-efficiency by reducing uncertainty and improving responsiveness.

Example: Application of quantitative techniques can be seen in Zomato, India's leading online food delivery platform. Zomato employs methods such as probability, regression analysis, and time series forecasting to anticipate customer demand across different times of the day and various locations. Drawing on historical order data, it uses forecasting models to allocate rider availability, applies correlation and regression to examine how discounts influence order volumes, and adopts optimisation techniques to reduce delivery time and cost. Through the continuous analysis of numerical data with these quantitative tools, Zomato enhances customer satisfaction, minimises operational inefficiencies, and boosts profitability, demonstrating how quantitative techniques underpin real-time managerial decision-making.

1.1.6 Limitations of Quantitative Techniques

Quantitative Techniques rely heavily on accurate data, and flawed inputs can lead to misleading conclusions that compromise business decisions. They often oversimplify complex real-world problems by ignoring qualitative factors such as human behaviour, ethics, and social impact. Additionally, these techniques require specialised skills and

SGOU

may struggle to adapt in volatile, unpredictable environments where intuition and flexibility are crucial.

1. Dependence on Accurate and Relevant Data

Quantitative analysis hinges entirely on the quality and availability of data. If the input data is inaccurate, incomplete, outdated, or biased, the resulting conclusions can mislead decision-makers and create false confidence in flawed strategies. For example, forecasting models based on outdated consumer trends may cause overproduction or misallocation of resources. Additionally, collecting granular and reliable data can be costly and time-consuming, especially in decentralised or informal business environments. Without trustworthy data, even the most sophisticated techniques lose their effectiveness.

2. Oversimplification of Complex Realities

While mathematical models are powerful, they often reduce multifaceted business scenarios into simplified equations and assumptions. Human behaviour, ethical considerations, cultural dynamics, and qualitative insights are typically excluded from quantitative frameworks. For instance, a pricing model may optimise profit but ignore customer sentiment or brand perception, leading to reputational damage. This disconnect between numeric optimisation and social nuance can result in decisions that are logically sound but practically unsustainable or insensitive.

3. Limited Adaptability to Dynamic Environments

Many quantitative methods assume stable relationships and predictable conditions, yet real-world business environments are anything but static. Market disruptions, competitor innovations, regulatory shifts, and geopolitical events often render these models obsolete or inaccurate. A linear programming solution optimised for last quarter's conditions might fail to accommodate a sudden change in demand or supply chain constraints. Because models typically rely on historical data and fixed parameters, their applicability diminishes when unpredictability is high.

4. Requires Technical Expertise and Interpretation Skills

Advanced QT tools like simulation modelling, queuing theory, and Monte Carlo analysis demand a firm grasp of statistics, mathematics, and software tools. Without skilled analysts or decision-makers who can interpret results correctly, businesses risk misunderstanding the output or applying models inappropriately. Misinterpretation can lead to flawed strategic decisions, inefficiencies, or misuse of company resources. Moreover, training personnel or hiring experts can incur additional costs and slow decision timelines.

5. Can Neglect Human Judgment and Ethical Dimensions

A significant critique of purely quantitative methods is their potential to marginalise intuition, experience, and ethical reasoning. Business decisions based solely on

SGOU

numbers might overlook long-term values, stakeholder trust, or moral considerations. For instance, a cost-cutting model might recommend downsizing without factoring in employee morale or community impact. While QT brings precision, it must be balanced with qualitative judgment to ensure wise and humane decisions.

Summarised Overview

This unit offers a comprehensive foundation for understanding the analytical tools used in business and management decision-making. It covers these techniques' meaning and core objectives, emphasising their role in enhancing accuracy, efficiency, and rationality in managerial processes. Learners explore diverse applications across functional areas, such as finance, operations, and strategy, where quantitative methods support forecasting, optimisation, and risk analysis. It also introduces key methodologies, including mathematical modelling and statistical analysis. Finally, it critically examines the limitations of such techniques, including data dependency, unrealistic assumptions, and contextual constraints, helping students recognise both their potential and boundaries in practical use.

Mini Task:

Identify whether the following problems require mathematical or statistical techniques:

a. Allocating delivery routes

b. Estimating average customer waiting time

c. Studying the relationship between ad spend and sales

Self-Assessment Question

1. Define quantitative techniques.
2. What are the key objectives of applying quantitative techniques in management studies?
3. Describe the general methodology used in conducting a quantitative analysis.
4. Explain how quantitative techniques contribute to effective managerial decisions.

SGOU

5. Discuss limitations of quantitative techniques in the real-world business context.
6. What challenges do organisations face when implementing quantitative techniques?
7. Compare the role of quantitative techniques in business management
8. List three commonly used mathematical techniques in business analysis and decision-making.
9. Define statistical techniques and explain their role in quantitative research.



Assignments

1. Critically evaluate the role of quantitative techniques in strategic business decision-making.
2. Analyse how the methodology of quantitative analysis ensures objectivity and replicability in business research.
3. Examine the limitations of quantitative techniques when applied to dynamic, multi-variable market environments.
4. How would you choose between mathematical and statistical approaches to evaluate the profitability of a new product line?



Reference

1. Aggarwal, B. M. (2010). *Business statistics*. Ane Books Pvt. Ltd.
2. Chandrasekharan, N., & Umaparvathi, M. (2016). *Statistics for management*. PHI Learning Pvt. Ltd.
3. Gupta, M., & Khanna, R. (2011). *Quantitative techniques for decision making*. Prentice Hall of India.
4. Gupta, S. P. (2021). *Statistical methods*. S. Chand & Sons.
5. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International (P) Ltd.

SGOU

6. Levin, R. I., & Rubin, D. S. (1999). *Quantitative approaches for management*. Pearson Education.

7. Sharma, A. (2017). *Quantitative techniques for decision making*. Himalaya Publishing House.

8. Vohra, D. V. D. (2017). *Quantitative techniques for management*. New Delhi: McGraw Hill Education



Suggested Reading

1. Arora, P. N., Arora, S., & Arora, S. (2010). *Comprehensive statistical methods*. S. Chand & Company Pvt. Ltd.

2. Beri, G. C. (2017). *Business statistics*. Tata McGraw-Hill.

3. Chawla, D., & Sondhi, N. (2016). *Research methodology: Concepts and cases*. Vikas Publishing House.

4. Krishnaswami, O. R., Ranganathan, M., & Harikumar, P. N. (2016). *Research methodology*. Himalaya Publishing House.

5. Render, B. (2022). *Quantitative analysis for management*. Prentice Hall of India.

6. Singh, A. K. (2017). *Tests, measurements and research methods in behavioural sciences*. Bharti Bhawan.

SGOU

Unit 2

Probability Distribution

Learning Outcomes

On completion of this unit, learners will be able to:

- ◊ explain the foundational concepts and applications of the Bernoulli process and Binomial Distribution
- ◊ introduce the structure and properties of the Poisson Distribution
- ◊ familiarise the Normal Distribution and its curve characteristics
- ◊ apply normal approximation techniques to both Binomial and Poisson distributions
- ◊ compare the utility and limitations of Binomial, Poisson, and Normal distributions

Background

The formal study of Probability began in the 17th century with correspondence between French mathematicians Blaise Pascal and Pierre de Fermat, who sought to solve gambling-related problems. This laid the groundwork for the development of discrete probability models. The Binomial Distribution emerged from the Bernoulli trials formulated by Jakob Bernoulli in his seminal work *Ars Conjectandi* (1713), where he investigated repeated independent experiments with binary outcomes. Bernoulli's ideas introduced concepts of random variation and expected frequencies, making the binomial model foundational for statistical inference and decision-making.

The Poisson Distribution was introduced by French mathematician Siméon Denis Poisson in the early 19th century, particularly in his 1837 work *Recherches sur la probabilité des jugements*. Originally applied in legal statistics, the distribution modelled rare events occurring over time or space, such as the number of deaths or insurance claims. Its ability to approximate binomial probabilities

SGOU

when event probabilities were low and the number of trials was large made it immensely useful in industrial quality control, traffic flow, and queuing analysis. Over time, Poisson models became key tools in operations research and inventory management.

The Normal Distribution, also known as the Gaussian distribution, traces back to the work of Abraham de Moivre in the 18th century and Carl Friedrich Gauss in the 19th century. De Moivre discovered the normal approximation to the binomial distribution, while Gauss refined the bell-shaped curve in his study of astronomical errors. The formulation of the central limit theorem demonstrated that the normal distribution naturally arises from the aggregation of many independent random variables. This established its utility across sciences, from economics and psychology to biology and engineering, cementing the normal curve as the cornerstone of statistical analysis and probabilistic modelling. This unit will give insights into the theoretical foundations, properties, and practical applications of these core distributions, empowering learners to interpret data patterns, make informed decisions, and apply statistical tools across diverse real-world scenarios.

Keywords

Probability, Baye's Theorem, Probability Distribution, Binomial Distribution, Bernoulli's Process, Poisson Distribution, Normal Distribution

Discussion

Probability distribution is a fundamental concept in statistics that describes how the probabilities of different outcomes are distributed for a random variable. In business management, it serves as a powerful tool for decision-making under uncertainty. For instance, a normal distribution, characterised by its bell-shaped curve, is commonly used to model employee performance scores, customer satisfaction ratings, or demand forecasts, where most values cluster around the mean. A binomial distribution applies when analysing the probability of success or failure in repeated trials, such as the likelihood of a customer making a purchase during a promotional campaign. Meanwhile, a Poisson distribution is useful for modelling the number of customer complaints received per day or the frequency of machine breakdowns in a factory. These distributions help managers quantify risk, optimise resource allocation, and make informed predictions. By understanding the shape, spread, and central tendency of a distribution, business leaders can tailor strategies, such as inventory control, quality assurance, or marketing outreach, based on probabilistic insights rather than intuition alone.

SGOU

1.2.1 Introduction

The classical approach to probability is a theoretical method that defines probability as the ratio of the number of favourable outcomes to the total number of equally likely outcomes in a finite sample space. It assumes that all outcomes of an experiment are equally possible, making it straightforward to calculate probabilities in simple, well-defined situations.

The classical (or theoretical) approach to probability is based on the assumption that all outcomes of an experiment are equally likely. It is defined as:

$$P(E) = (\text{Number of favourable outcomes}) / (\text{Total number of possible outcomes})$$

Where:

- ◊ P(E) = Probability of event E
- ◊ Favourable outcomes = outcomes that satisfy the event
- ◊ Total outcomes = all possible outcomes in the sample space

Example: Tossing a Coin

- ◊ Sample space (S): {Head, Tail}
- ◊ Total outcomes: 2
- ◊ Probability of getting a Head: P(Head) = $1/2$
- ◊ Probability of getting a Tail: P(Tail) = $1/2$

Example: Rolling a Dice

- ◊ Sample space (S): {1, 2, 3, 4, 5, 6}
- ◊ Total outcomes: 6
- ◊ Probability of getting a 4: P (4) = $1/6$

Probability of getting an even number (2, 4, 6): :P (Even) = $3/6 = 1/2$

1.2.1.1 Probability

Probability is a mathematical discipline focused on quantifying the likelihood of events or the truth of propositions. It assigns values ranging from '0' to '1' to represent the chance of occurrence, where '0' denotes an impossible event and '1' signifies absolute certainty. Expanding on this concept, a probability distribution provides a structured framework for analysing and predicting outcomes across various scenarios. These distributions are highly valuable in practical applications, particularly in business contexts that require precise probability calculations. The fundamental aim of probability analysis is to support informed and rational decision-making. Through probability theory, we gain the tools to assess uncertainty and guide strategic choices.

SGOU

$$P(H_1) + P(H_2) + P(H_n) = 1$$

This implies that the total Probability of a definite event is allocated among a collection of mutually exclusive events that together form an exhaustive set. Broadly speaking, a probability distribution refers to a systematic listing, often in tabular form, of the probabilities associated with every possible outcome of a random experiment. It is also known as a theoretical frequency distribution, as it represents the expected frequencies of outcomes based on Probability rather than actual observation.

1.2.2 Definition of terms

a. Random Experiment

A random experiment refers to a process conducted under consistent conditions that can yield two or more possible outcomes, which vary unpredictably from one trial to another. While all potential results are known beforehand, the exact outcome of any single trial cannot be determined with certainty. Common examples include tossing a coin or rolling a die, each has defined outcomes, yet the result of each attempt remains uncertain. Key characteristics of a random experiment include having multiple possible outcomes, unpredictability of results, and the ability to repeat the experiment under identical conditions.

b. Sample Point

A sample point refers to an individual outcome of a random experiment. For instance, in a coin toss, obtaining a head is a sample point; in the case of rolling two dice, getting the result (1, 4) is also a sample point.

c. Sample Space

The sample space is the complete set of all possible outcomes of a random experiment. Examples include:

- ◊ Tossing a single coin: sample space = {Head, Tail}
- ◊ Tossing two coins: sample space = {HH, HT, TH, TT}

d. Event

An event consists of one or more outcomes from a random experiment and is considered a subset of the sample space. Events can be:

- ◊ **Simple**, involving a single outcome (e.g., getting two heads when two coins are tossed)
- ◊ **Compound**, involving multiple outcomes (e.g., getting heads in both tosses when a coin is tossed twice)

These concepts form the foundation for analysing probabilities in any random process.

SGOU

Events can be categorised in the following ways:

i. Certain, Impossible, and Uncertain Events

- ◊ A certain event is guaranteed to happen (e.g., drawing a white ball from a bag containing only white balls).
- ◊ An impossible event can never occur (e.g., drawing a white ball from a bag with only black balls).
- ◊ An uncertain event may or may not happen (e.g., drawing a white ball from a bag containing both white and black balls).

ii. Equally Likely Events

Events are equally likely when each has the same chance of occurring.

Examples:

- ◊ Tossing a coin: Head or Tail
- ◊ Rolling a die: Any number from 1 to 6

iii. Mutually Exclusive Events

Events are mutually exclusive if the occurrence of one prevents the occurrence of the other.

Example: In a coin toss, getting a head means you cannot get a tail in the same trial.

iv. Exhaustive Events

A set of events is exhaustive when it includes all possible outcomes of a random experiment.

Example: Rolling a die:- outcomes 1 through 6 cover all possibilities.

v. Complementary Events

Two events are complementary if one represents the occurrence and the other the non-occurrence of the same event.

- ◊ Denoted as A and A^c
- ◊ Probability: $P(A^c) = 1 - P(A)$

Example: In a coin toss, Head and Tail are complementary.

vi. Union of Events

The union of events A and B (denoted A∪B) includes all outcomes that belong to either A or B or both.

SGOU

Example:

- A = {1,2,3,4}, B = {3,4,5,6}
- A ∪ B = {1,2,3,4,5,6}

vii. Intersection of Events

The **intersection** of events A and B (denoted A ∩ B) includes only the outcomes common to both.

Example:

- A ∩ B = {3,4}

viii. Independent and Dependent Events

Independent events do not influence each other's occurrence.

Example: Tossing a coin twice: the result of one toss does not affect the other. Conditions:

- P(A/B) = P(A)
- P(B/A) = P(B)
- P(A and B) = P(A) P(B)

- Dependent events are influenced by each other.

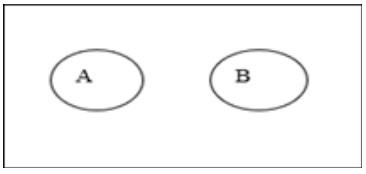
Example: Drawing two cards from a deck without replacement. the second draw depends on the first.

1.2.3 Theorems of Probability

a. Addition Theorem

The Addition Theorem in Probability helps determine the likelihood of either event A or event B occurring. This is represented as P(AUB), and its calculation depends on whether the events are mutually exclusive:

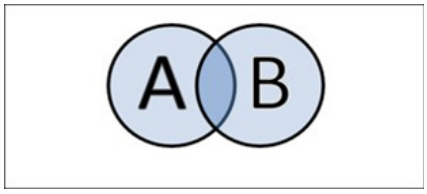
- If A and B are mutually exclusive (i.e., they cannot occur together), then: P(AUB) = P(A) + P(B)
- Since they are disjoint, their intersection A∩B is the empty set (∅), meaning no shared outcomes.
- If A and B are not mutually exclusive (i.e., they can occur simultaneously), then: P(AUB) = P(A) + P(B) - P(A∩B)



SGOU

◊ This formula adjusts for the overlap between A and B, subtracting the Probability of their intersection to avoid double-counting shared outcomes.

The proof is based on the principle that the total number of outcomes in A \cup B equals the sum of outcomes in A and B minus those common to both. Hence, n(A U B) = n(A) + n(B) - n(A∩B), which leads to the corresponding probability expression.



b. Multiplication Theorem

If A and B are two events, the Probability that both occur together, represented as P(A∩B), depends on whether they are independent or not:

◊ If A and B are not independent, then: P(A∩B) = P(A) P(B/A)

- This means the Probability of both events happening is the product of the Probability of A and the conditional Probability of B given that A has occurred.
- If A and B are independent, then: P(A∩B) = P(A) P(B)

◊ Since the occurrence of one does not affect the other, their joint Probability is simply the product of their individual probabilities.

Probability in the Context of Statistical Dependence

When two or more events are statistically dependent, it means that the occurrence of one event influences the likelihood of the others occurring.

Conditional Probability – the Probability of one event occurring given that another event has already occurred. Conditional Probability refers to the chance of an event A occurring, given that another event B has already occurred. It is denoted as: P(A/B)

Example:

Imagine rolling a fair six-sided die. The possible outcomes are:

[{1, 2, 3, 4, 5, 6}]

Let's say we're interested in the Probability that the number rolled is even, given that it is greater than 2.

◊ Total outcomes greater than 2: {3, 4, 5, 6} → 4 outcomes

◊ Among these, the even numbers are: {4, 6} → 2 outcomes

So, the conditional Probability is:

SGOU

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

This is a conditional probability because we're calculating the likelihood of rolling an even number under the condition that the number is already known to be greater than 2.

1.2.4 Baye's Theorem

Bayes' Theorem, developed by British mathematician Thomas Bayes, helps us revise probabilities when new evidence is introduced.

Example: Email Spam Detection

Let: A be the event that an email contains the word "free".

◊ C be the event that the email is actually spam.

From past data, we can estimate: P(free/spam) = P(A/C)

This tells us how likely a spam email is to contain the word "free".

However, in practice, we want to use this information to predict whether an email is spam based on the presence of the word "free": P(spam/free) = P(C/A)

Bayes' Theorem allows us to reverse the conditioning and calculate this Probability using known data about spam and word usage patterns. This is the foundation of many spam filters and predictive algorithms.

1.2.5 Frequency Distribution and Probability Distribution

Frequency Distribution

◊ Based on actual observations.

- Example: Suppose we record the test scores of 100 students and group them into intervals: 0–20, 21–40, 41–60, 61–80, 81–100

◊ We then count how many students fall into each range.

◊ The table has:

- Score intervals in one column
- Number of students in the other

◊ The total frequency equals the number of students (100).

Probability Distribution

◊ Based on theoretical expectations or historical patterns.

◊ Example: Assume we model the expected performance of students using a normal distribution.

SGOU

◊ The table includes:

- Possible score values
- Associated probabilities of each score occurring

◊ The sum of all probabilities equals 1.

◊ Like frequency distributions, probability distributions can be graphed and described using:

- Mean (average score)
- Variance (spread of scores)
- Skewness (symmetry or asymmetry of the distribution)

Let us consider an example of a probability distribution. Suppose we toss a fair coin twice; the possible outcomes are shown in the Table below:

Table 1.2.1 Possible Outcomes from a Two–Toss Experiment of a Fair Coin0

No. of possible outcomes	**1st toss**	**2nd toss**	**No of heads on two tosses**	**Probability of the possible outcomes**
1	Head	Head	2	0.5 x 0.5 = 0.25
2	Head	Tail	1	0.5 x 0.5 = 0.25
3	Tail	Head	1	0.5 x 0.5 = 0.25
4	Tail	Tail	0	0.5 x 0.5 = 0.25
				Total=1.00

We start by noting the outcome that has no heads, which corresponds to the fourth entry in Table 1.2.1. Then, we identify the outcomes that contain exactly one head these are the second and third entries. Lastly, we record the first outcome, which includes two heads. This classification is summarised in Table 1.2.2, where each outcome is grouped according to the number of heads it contains.

Table 1.2.2 Probability Distribution of the Possible No. of Heads from Two toss Experiment of a Fair coin

No. of Heads(H)	**Tosses**	**Probability of outcomes P (H)**
0	(TT)	$\frac{1}{4}$= 0.25
1	(HT)+(TH)	$\frac{2}{4}$= 0.50

SGOU

2	(HH)	$\frac{1}{4} = 0.25$

It is important to recognise that the tables presented do not reflect the actual results of tossing a fair coin twice. Rather, they illustrate the theoretical outcomes, that is, they represent the expected behaviour of a two-toss experiment involving an unbiased coin when considered over a long period.

1.2.6 Types of Probability Distribution

Probability distributions are generally categorised into two main types:

i. Discrete Probability Distribution

ii. Continuous Probability Distribution

◊ **Discrete Probability Distribution**

A discrete probability distribution is one in which the variable can assume only a finite or countable set of values. For instance, the Probability of someone being born in a particular month is discrete, as there are only 12 distinct outcomes, each corresponding to a month of the year.

◊ **Continuous Probability Distribution**

In contrast, a continuous probability distribution allows the variable to take on any value within a specified interval. The range of possible outcomes is uncountably infinite, making it suitable for modelling phenomena such as height, weight, or time.

Random Variable

A random variable is a numerical quantity whose value is determined by the outcome of a random experiment. The complete set of all possible outcomes from such an experiment is referred to as the sample space. Like probability distributions, random variables are classified as either:

◊ Discrete Random Variables

◊ Continuous Random Variables

Discrete Probability Distribution

A discrete probability distribution presents all the possible values that a discrete random variable can take, along with the associated probabilities of each value occurring. Within this category, two commonly studied distributions are:

i. Binomial Distribution

ii. Poisson Distribution

SGOU

1.2.7 Binomial Distribution

The binomial distribution is a fundamental probability model used to calculate the likelihood of a specific number of successful outcomes across a fixed number of trials, where each trial has the same probability of success. It is widely applied in various fields, especially in business scenarios. For instance, a quality control manager may use it to estimate the chance of finding defective items in a sample of 10 products, assuming a 10% defect rate. This can be quickly determined using binomial probability tables. The distribution is applicable to discrete data, not continuous, and is based on a type of experiment known as a Bernoulli process. Because of its simplicity and versatility, the binomial distribution serves as a foundation for more advanced statistical models.

1.2.8 Bernoulli Process

A Bernoulli process refers to a sequence of independent trials where each trial results in one of two outcomes: success (typically coded as 1) or failure (coded as 0). The Probability of success is denoted by p, while the Probability of failure is 1 − p. The distribution of outcomes from a single Bernoulli trial is called the Bernoulli distribution, named after the Swiss mathematician Jacob Bernoulli (1654–1705). This simple yet powerful model is the building block for the binomial distribution and is crucial in understanding more complex probabilistic frameworks.

Binomial Distribution and Its Conditions

In a binomial distribution, the Probability of obtaining 0, 1, 2, ..., up to **n** successes (such as heads or tails) in **n** independent tosses of a fair coin is represented by the successive terms in the expansion of $\mathbf{(q + p)^n}$,

- ◊ Each trial must result in one of two outcomes: success or failure.
- ◊ The Probability of success (p) must remain constant across all trials.
- ◊ All trials must be independent, meaning the result of one trial does not influence or get influenced by another

where:

- ◊ p is the Probability of success (e.g., getting heads),
- ◊ q is the Probability of failure (i.e., 1 − p).

This mathematical model is known as the binomial Law of Probability and is applicable under specific conditions:

Key Assumptions of Binomial Distribution

To apply the binomial distribution accurately, the following assumptions must be met:

SGOU

1. **Binary Outcomes:** Each trial yields only two possible results, such as Yes/No or Success/Failure.
2. **Constant Probability:** The likelihood of success or failure remains unchanged, no matter how many times the experiment is repeated.
3. **Statistical Independence:** The outcome of one trial does not affect the outcome of any other trial.
4. **Fixed Number of Trials:** The total number of trials is predetermined and finite (e.g., 1, 2, 3, 4, 5, etc.).

Binomial Probability Formula

$$P(x:n,p) = {}^{n}C_{x}\, p^{x}\, (q)^{n-x}$$

where,

- ◊ n = the number of experiments
- ◊ x = 0, 1, 2, 3, 4, …
- ◊ p = Probability of success in a single experiment
- ◊ q = Probability of failure in a single experiment (= 1 – p)

The binomial distribution formula is also written in the form of n-Bernoulli trials,

$$\text{where } {}^{n}C_{x} = \frac{n!}{x!(n-x)!}$$

$$\text{Hence, } P(x:n,p) = \frac{n!}{x!(n-x)!}\, p^{x}.(q)^{n-x}$$

n! can be simplified as follows:

$n! = n(n-1)! = (n-1)(n-2)! = (n-1)(n-2)(n-3)!$ and so on.

Binomial Distribution Formula

$$P(x) = \binom{n}{x} p^{x} q^{n-x} = \frac{n!}{(n-x)!x!} p^{x} q^{n-x}$$

where
n = the number of trials (or the number being sampled)
x = the number of successes desired
p = probability of getting a success in one trial
$q = 1 - p$ = the probability of getting a failure in one trial

SGOU

The symbol! means' factorial', which is computed as follows:

5! means 5x4x3x2x1=120

Mathematicians define 0! as 1

If n is large in number, say ${}^{50}C_3$, then we write (with the help of the above explanation)

$$
\begin{aligned}
{}^{50}C_3 &= \frac{50!}{3!(50-3)!} \\
&= \frac{(50)(49)(48)(47)!}{3!(47)!} \\
&= \frac{50 \times 49 \times 48}{3 \times 2 \times 1} = 19600
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
{}^{75}C_5 &= \frac{75!}{5!(75-5)!} \\
&= \frac{(75)(74)(73)(72)(71)(70)!}{5!(70)!} \\
&= \frac{75 \times 74 \times 73 \times 72 \times 71 \times 70}{5 \times 4 \times 3 \times 2 \times 1} \text{ and so, on}
\end{aligned}
$$

1.2.9 Properties of Binomial Distribution

i. Binomial distribution is a discrete probability distribution.

ii. The shape and location of the Binomial distribution change as 'p' changes for a given 'n'.

iii. The mode of the binomial distribution is at the value of x which has the maximum Probability. Binomial distribution has one or two modes.

iv. Mean of the binomial distribution increases as 'n' increases with 'p' remaining constant.

v. If 'n' is large and if neither 'p' nor 'q' is too close to zero binomial distribution may be approximately normal distribution.

vi. Binomial distribution has mean = np and S.D. = $\sqrt{npq}$

vii. If two independent random variables follow Binomial distribution, their sum also follows Binomial distribution.

SGOU

Illustration 1.2.1

A fair coin is tossed six times. What is the Probability of obtaining four or more heads?

Solution

When a fair coin is tossed, the probabilities of head and tail in case of an unbiased coin are equal, i.e.,

$$P = q = \frac{1}{2} \text{ or } 0.5$$

The probabilities of obtaining 4 heads are:

$$P(x:n,p) = [\, n!/x!\,(n-x)!\,] \cdot p^x.(q)^{n-x}$$

$$P(4) = {}^6C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{6-4}$$

$$P(4) = \frac{6!}{4!(6-4)!} (0.5)^4 (0.5)^2$$

$$P(4) = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{(4 \times 3 \times 2 \times 1)(2 \times 1)} (0.0625)\ (0.25)$$

$$= \frac{720}{(24)(2)} (0.0625)\ (0.25)$$

$$= 15 \times 0.0625 \times 0.25 = 0.234$$

The Probability of obtaining 5 heads is

$$P(5) = {}^6C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{6-5}$$

$$P(5) = \frac{6!}{5!\,(6-5)!} (0.5)^5 (0.5)^1$$

$$P(5) = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{(5 \times 4 \times 3 \times 2 \times 1)(1 \times 1)} (0.03125)\ (0.5)$$

$$= 6 \times 0.03125 \times 0.5 = 0.094$$

The Probability of obtaining 6 heads is:

$$P(6) = {}^6C_6 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^{6-6}$$

$$P(6) = \frac{6!}{6!(6-6)!} (0.5)^6 (0.5)^0$$

SGOU

$$P(6) = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{(6 \times 5 \times 4 \times 3 \times 2 \times 1)(1)} (0.015625)\ (1)$$

$$= 1 \times 0.015625 \times 1$$

$$= 0.016$$

The Probability of obtaining 4 or more heads is:

$$0.234 + 0.094 + 0.016 = 0.344$$

Illustration 1.2.2

The incidence of a certain disease is such that, on average, 20% of workers suffer from it. If 10 workers are selected at random, find the Probability that:

i. Exactly 2 workers suffer from the disease

ii. Not more than 2 workers suffer from the disease

iii. At least 2 workers suffer from the disease

Solution

Probability that a worker suffering from the disease $= \frac{20}{100} = \frac{1}{5}$ i.e., $p = \frac{1}{5}$ and

The Probability of a worker not suffering from the disease, i.e., $q = (1 - \frac{1}{5}) = \frac{4}{5}$

By the binomial probability law, the Probability that out of 10 workers, 'x' workers suffer from a disease is given by:

$$P(x:n,p) = {}^{n}C_{x}\ p^{x}\ (q)^{n-x}$$

$$= {}^{10}C\ (\frac{1}{5})^{r} (\frac{4}{5})^{10-x};\ x = 0,1,2,3,\ldots\ldots 10$$

The required Probability that exactly 2 workers will suffer from the disease is given by:

$$P(2) = {}^{10}C_{2}\ (\frac{1}{5})^{2}\ (\frac{4}{5})^{10-2}$$

$$P(2) = \frac{10!}{2!(10-2)!}\ (0.2)^{2}(0.8)^{8}$$

$$= \frac{10 \times 9 \times 8!}{(2 \times 1)\ (8)!} (0.04)\ (0.1677)$$

$$= 45\ (0.04)\ (0.1677)$$

$$= 0.302$$

SGOU

The required Probability that not more than 2 workers will suffer from the disease is given by:

$$P(0) + P(1) + P(2)$$

$$P(0) = {}^{10}C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{10-0} = 0.107$$

$$P(1) = {}^{10}C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^{10-1} = 0.269$$

$$P(2) = {}^{10}C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^{10-2} = 0.302$$

Probability of not more than 2 workers suffering from the disease

$$= 0.107 + 0.269 + 0.302 = 0.678$$

We have to find out $P(x \geq 2)$ i.e., $P(x \geq 2) = 1 - P(0) - P(1)$

$$= 1 - 0.107 - 0.269 = 0.624$$

Thus, the Probability of at least two workers suffering from the disease is 0.624

Frequency = Probability × N

Illustration 1.2.3

Dice are thrown 162 times. The occurrence of 2 or 3 is considered a success. In how many throws, do you expect (i) exactly 2 success (ii) at least one success?

Solution

$$P[\text{getting 2 or 3}] = \frac{2}{6} = \frac{1}{3}$$

$$p = \frac{1}{3} \text{ and } q = 1 - \frac{1}{3} = \frac{2}{3}$$

$$n = 4 \text{ and } N = 162$$

i. $P[\text{exactly 2 success}] = P(x = 2)$

$$= 4C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{4-2}$$

$$= 6 \times \frac{1}{9} \times \frac{4}{9}$$

$$= \frac{24}{81}$$

Number of throws in which exactly 2 success occur = Probability × N $= \frac{24}{81} \times 162 = 48$

SGOU

ii. P [at least one success] = 1 -P [no success]

$$= 1 - P[x = 0] = 1 - {}^4C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{4-0}$$

$$= \frac{65}{81}$$

Number of throws in which at least one success occurs

$$= \text{Probability} \times N = \frac{65}{81} \times 162 = 130$$

1.2.10 Mean and Standard Deviation of Binomial Distribution

The Binomial distribution has expected values of mean and a standard deviation (σ). We can express the mean of the binomial distribution as: Mean = np

Where n = number of trials; p = probability of success and, we can calculate the standard deviation by:

$$\sigma = \sqrt{npq}$$

Where n = Number of trials; p = probability of success and q = probability of failure =1- p

Illustration 1.2.4

If the Probability of defective bolts is 0.1, find the mean and standard deviation for the distribution of defective bolts in a total of 300.

Solution

P=0.1, n=300

Hence mean = np = 300 x 0.1 =30

Thus, we can expect 30 bolts to be defective

Standard Deviation = $\sqrt{npq}$

n =300, p = 0.1, q = 1 – p = 1 – 0.1 = 0.9

$\sigma = \sqrt{npq} = \sqrt{300 \times 0.1 \times 0.9} = 5.196$

1.2.11 Fitting a Binomial Distribution

When a binomial distribution is to be fitted to observed data, the following procedure is adopted:

◊ Determine the value of 'p' ,'q' and 'n' and substitute them in the function ${}^nC_x\ p^x\ (q)^{n-x}$, we get the probability function of the Binomial distribution.

SGOU

◊ Put x = 0,1,2,…. In the function ${}^nC_x\ p^x\ (q)^{n-x}$, we get n + 1 terms.

◊ Multiply each term of the expanded binomial by N (the total frequency), to obtain the expected frequency in each category.

Illustration 1.2.5

Eight coins were tossed together,256 times. Find the expected frequencies of Heads. Find mean and S.D.

Solution

p = P (getting head in a toss) = ½

q = 1 – p = ½ and n= 8

The Binomial Distribution is $P(x) = 8C_x\ (1/2)^x\ (1/2)^{8-x}$

Putting x = 0,1,2,3,…,8, we get all the terms of the Binomial Distribution

No. of heads (x)	P(x)	Expected Frequency P(x) = 256
0	$8C_0\ (1/2)^0\ (1/2)^8 = 1/256$	1
1	$8C_1\ (1/2)^1\ (1/2)^7 = 8/256$	8
2	$8C_2\ (1/2)^2\ (1/2)^6 = 28/256$	28
3	$8C_3\ (1/2)^3\ (1/2)^5 = 56/256$	56
4	$8C_4\ (1/2)^4\ (1/2)^4 = 70/256$	70
5	$8C_5\ (1/2)^5\ (1/2)^3 = 56/256$	56
6	$8C_6\ (1/2)^6\ (1/2)^2 = 28/256$	28
7	$8C_7\ (1/2)^7\ (1/2)^1 = 8/256$	8
8	$8C_8\ (1/2)^8\ (1/2)^0 = 1/256$	1

Mean = np = 8 x $1/2$ = 4

S.D = $\sqrt{npq}$ = $\sqrt{8 \text{ x } 1/2 \text{ x } 1/2}$ = $\sqrt{2}$ = 1.414

Illustration 1.2.6

The following data show the number of seeds germinating out of 5 ib damp filter for 80 sets of seeds. Fit a binomial distribution of this data and find the expected frequencies.

x	0	1	2	3	4	5
y	6	20	28	12	8	6

SGOU

Solution

Since 'p' is not given find mean of the given data.

x	f	fx
0	6	0
1	20	20
2	28	56
3	12	36
4	8	32
5	6	30
	80	174

$$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{174}{80} = 2.175$$

i.e . np = 2.175

n = 5

$$p = \frac{2.175}{5} = 0.435$$

and q = 1 – p = 0.565

Binomial distribution is $5C_x (0.435)^x (0.565)^{5-x}$ for x = 0,1,2,3,4,5

Expected frequencies are obtained from $N_x\ nC_x\ p^x\ q^{n-x}$, where N = 80, n = 5, p =0.4354, q = 0.565 and x = 0,1,2,3,4,5

x	Expected frequencies $N_x\ nC_x\ p^x\ q^{n-x}$
0	$80 \times 5C_0 (0.435)^0 (0.565)^5 = 5$
1	$80 \times 5C_1 (0.435)^1 (0.565)^4 = 18$
2	$80 \times 5C_2 (0.435)^2 (0.565)^3 = 27$
3	$80 \times 5C_3 (0.435)^3 (0.565)^2 = 21$
4	$80 \times 5C_4 (0.435)^4 (0.565)^1 = 6$
5	$80 \times 5C_5 (0.435)^5 (0.565)^0 = 7$

1.2.12 Utilities of the Binomial distribution

- ◊ Useful in decision-making situations in business.
- ◊ It serves as the basis for hypothesis testing and confidence interval estimation in situations involving binary outcomes, such as success/failure, yes/no, or presence/absence.

SGOU

- ◊ In finance, binomial models are employed to price options, where the underlying assets can either increase or decrease in value at each time step, akin to the binomial outcomes of success or failure.
- ◊ Used in sampling plans for inspecting the article drawn in the sample.

1.2.13 Poisson Distribution

The Poisson distribution is a type of discrete probability distribution used to estimate the likelihood of a specific number of events occurring within a fixed interval of time, space, distance, area, or volume. Named after the French mathematician Simeon Poisson, this distribution is particularly useful for modelling rare or infrequent events. Examples include the number of accidents in a day, defects in a manufacturing process, or employee absences over a given period.

It is ideal for scenarios where events occur independently and randomly, but at a known average rate, making it a powerful tool in fields like public health, quality control, and operations management.

Relationship with Binomial Distribution

While the binomial distribution depends on two parameters, n (number of trials) and p (Probability of success), there are situations where p is very small (e.g., less than 0.01) and n is quite large (e.g., greater than 50). In such cases, the product np remains relatively constant. Under these conditions, the Poisson distribution serves as a good approximation of the binomial distribution. Here ***m = np.***

The Poisson distribution process corresponds to a Bernoulli process with a very large number of trials (n) and a very low probability of success. Thus, it is also called the Law of improbable events. The parameter of the Poisson distribution is ***m***.

$$\text{Probable Density, } P(x) = \frac{e^{-m}m^{x}}{x!}$$

Where,

P (x) = Probability of success desired,

x = 0, 1,2,3, 4,……∞ (any positive integer)

e = a constant with value 2.7183 (the base of natural algorithms)

m = mean of the Poisson distribution, i.e., np or the average number of occurrences of an event.

When n is large, while the probabilities of occurrence of an event are close to 0, so that q = (1-p), the Binomial distribution is very closely approximated by the Poisson distribution with $m = np$.

1.2.14 Properties of the Poisson Distribution

a. It is a discrete probability distribution, the limiting form of the binomial distribution.

b. The range of the random variable is $0 \leq x < \infty$.

c. It consists of a single parameter "m" only. So, the entire distribution can be obtained by knowing this value only.

d. Mean and variance of Poisson distribution are equal to "m".

e. It is a positively skewed distribution.

Measures of Central Tendency

In the Poisson distribution, the mean (m) and the variance (s^2) represent the same value, i.e., Mean = variance = m

$$SD = \sqrt{m}$$

$$S.D\ (\sigma) = \sqrt{variance} = \sqrt{np}$$

Illustration 1.2.7

2% of the electronic toys produced in a specific manufacturing process are defective. What is the Probability that a shipment of 200 toys will contain exactly 5% defectives? Also, the mean and standard deviation were found.

Solution

n = 200

Probability of a defective toy (p) = $\frac{2}{100}$

= 0.02

Since n is large and p is small, the Poisson distribution is applicable. Apply the formula as follows:

$$P(x) = \frac{e^{-m} m^x}{x!}$$

The Probability of 5 defective pieces in 200 toys is given by:

$$P\ (5) = \frac{m^5\ e^{-m}}{5!}$$

Where,

m = np = 200 x 0.02 = 4

SGOU

$$P(5) = \frac{18.7392}{120}$$

$$= 0.156$$

Mean = np = 200 x 0.02 =4

S.D (σ) = $\sqrt{4} = 2$

Illustration 1.2.8

On average, a certain intersection results in 3 traffic accidents per month. What is the Probability that in any given month that through this intersection, no accidents will occur?

Solution

$$P(x) = \frac{e^{-m} m^x}{x!}$$

$$= \frac{e^{-3} 3^0}{0!}$$

$$= 0.0496$$

1.2.15 Fitting of a Poisson Distribution

To fit a Poisson distribution to a given observed data (frequency distribution), the procedure is as follows:

a. We must obtain the value of its mean, i.e., m = np

b. The probabilities of various values of the random variables (x) are to be computed:

$$P(x) = \frac{e^{-m} m^x}{x!}$$

c. Each Probability so obtained is then multiplied by N (the total frequency) to get expected frequencies. i.e. N × $P(x)$

Illustration 1.2.9

Fit a Poisson distribution to the following data and calculate the theoretical frequencies.

x	0	1	2	3	4
f	123	59	14	3	1

SGOU

Solution

x	f	fx
0	123	0
1	59	59
2	14	28
3	3	9
4	1	4
	200	100

$$\text{on } \bar{x} = \frac{\sum fx}{N}$$

$$= \frac{100}{200} = 0.5$$

Thus m = 0.5

Therefore, the Poisson distribution,

$$P(x) = \frac{e^{-m} 0.5^x}{x!}$$

Theoretical Frequencies are

x	P(x)	Theoretical frequency = N × P(x)
0	$\frac{e^{-0.5}0.5^0}{0!} = 0.6065$	200 × 0.6065 = 121
1	$\frac{e^{-0.5}0.5^1}{1!} = 0.30325$	200 × 0.303 = 61
2	$\frac{e^{-0.5}0.5^2}{2!} = 0.0766$	200 × 0.0766 = 15
3	$\frac{e^{-0.5}0.5^3}{3!} = 0.0127$	200 × 0.0127 = 3
4	$\frac{e^{-0.5}0.5^4}{4!} = 0.0016$	200 × 0.0016 = 0

SGOU

1.2.16 Utility of Poisson Distribution

The Poisson distribution is a powerful statistical tool used to model the Probability of a given number of events occurring within a fixed interval of time, space, or volume. Its utility spans various domains due to its simplicity and effectiveness in handling rare or random events.

Key Applications

- ◊ **Quality Control:** Estimating the number of defects in a batch of products.
- ◊ **Healthcare:** Modelling the number of patient arrivals at a clinic or the incidence of rare diseases.
- ◊ **Traffic Engineering:** Predicting the number of cars passing through a checkpoint in a given time.
- ◊ **Telecommunications:** Analysing call arrivals at a call centre.
- ◊ **Natural Sciences:** Measuring radioactive decay events or mutations in DNA sequences.

Advantages

Table 1.2.3 Advantages of Poisson distribution

Advantage	Description
Versatility	Applicable across fields like biology, finance, engineering, and logistics.
Simplicity	Easy to understand and implement, even with large datasets.
Predictive Power	Useful for forecasting rare events with high precision.
Efficiency	Handles large-scale data with minimal computational complexity.
Foundation for Other Models	Serves as a basis for exponential and compound distributions.

1.2.17 Normal Distribution

Binomial and Poisson distributions are widely used for modelling discrete random variables, while the normal distribution is the most important theoretical model for continuous variables. Known as the Gaussian distribution, it is symmetric around the mean, indicating that values closer to the mean occur more frequently than those further away. The normal distribution plays a central role in statistical inference, helps quantify uncertainty in real-world scenarios, and approximates other distributions.

SGOU

Initially introduced by De Moivre in 1733 as a limit of the binomial distribution, the normal distribution is commonly observed in business and natural phenomena. Examples include measurements like height, weight, and product dimensions. Its widespread applicability aids in calculating probabilities and making informed decisions. A normal distribution is fully described by just two parameters: the mean and the standard deviation (σ), which are visually identifiable in its characteristic bell-shaped curve.

a. Definition

A continuous random variable X is said to follow a normal distribution if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

where,

- μ is the mean of X,
- σ is the standard deviation of X.

The variable X varies between -∞ and + ∞.

b. Normal Distribution as a Limiting Case of the Binomial Distribution

The Binomial distribution is an important theoretical distribution for discrete random variables. Under certain conditions, the Binomial distribution tends to the Normal distribution. This approximation holds when:

1. The number of trials (n) is very large.
2. The probabilities of success (p) and failure (q=1-p) are nearly equal.

In such cases, the Binomial distribution can be approximated by the Normal distribution.

c. Probability Density Function and Parameters of the Normal Distribution

If X is a continuous random variable following a normal distribution with mean \μ and standard deviation σ, its probability density function (PDF) is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

where,

- ◊ π = 3.14159
- ◊ e = 2.71828

SGOU

◊ Mean (μ) – determines the location of the curve along the X-axis.

◊ Standard deviation (σ) – determines the spread or shape of the curve.

Different values of **μ** and **σ** produce different Normal curves.

The variable X can take any real value in the range -∞ and + ∞.

1.2.18 Normal distribution curve

The normal distribution curve, also known as the Gaussian distribution or bell curve, is a continuous probability distribution that is symmetrical around its mean. It represents how data values are distributed, with most observations clustering near the centre (mean), and fewer occurring as you move further away from it. The curve is highest at the mean and tapers off equally on both sides, reflecting that values closer to the mean are more frequent than extreme values.

This distribution is entirely defined by two parameters:

◊ **Mean (μ)**: the central value

◊ **Standard deviation (σ)**: which determines the spread or width of the curve

The total area under the curve equals 1, representing the complete probability space.

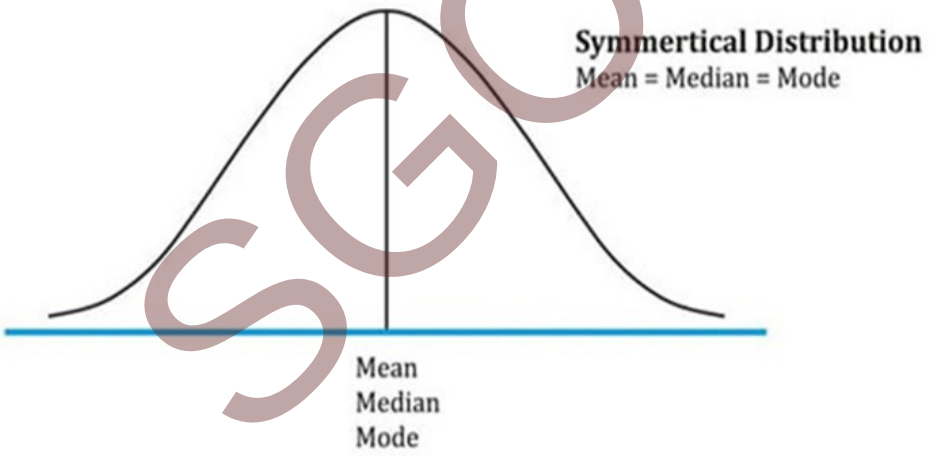


Fig. 1.2.1 Normal probability distribution

1.2.19 Properties of Normal Distribution

i. The normal curve is continuous.

ii. It is bell-shaped.

iii. It is symmetrical about the mean.

iv. Mean, Median and Mode are equal for a normal curve.

v. The height of the normal curve is at its maximum at the mean.

SGOU

vi. There is only one maximum point, which occurs at the mean.

vii. The ordinate at the mean divides the whole area into two equal parts. The area to the right of the ordinate and that to the left are both 0.5.

viii. The curve is asymptotic to the base on either side: i.e., it approaches nearer and nearer to the base but never touches the base.

ix. The coefficient of skewness is 0 (i.e., $\beta = 0$).

x. The normal curve is unimodal, i.e., it has only one mode.

xi. The point of intersection occurs at $\mu \pm \sigma$. At the point of inflexion, the curve changes from concavity to convexity.

xii. Q1 and Q3 are equidistant from the median.

xiii. Mean deviation for the normal curve is $\frac{4}{5}\sigma$, and QD is $\frac{2}{3}\sigma$.

xiv. The normal curve is mesokurtic. That is measure of kurtosis is β_2=3.

xv. No portion of the curve lies below the X-axis.

xvi. Theoretically, the range of the normal curve is $-\infty$ *to* $+\infty$. But practically the range is -3σ to $+3\sigma$.

xvii. A linear function of several normal variates is also a normal variate. This is known as the additive property of the normal distribution.

xviii. Area under the normal curve is distributed as follows:

Mean $\pm \sigma$ covers 68.27 % area

Mean $\pm 2\sigma$ covers 95.45 % area

Mean $\pm 3\sigma$ covers 99.73 % area

This is the area property of the normal distribution

1.2.20 Utility of Normal Distribution

The study of the normal distribution is of central importance in statistical analysis because of the following reasons:

i. Most discrete probability distributions (Binomial distributions, Poisson distributions) tend to approximate the normal distribution as 'n' becomes large.

ii. Almost all sampling distributions, such as Student's t- distribution, F-distributions, z-distributions, and chi-square distributions, conform to normal distribution for large values of n.

iii. The various tests of significance like t-test, F –test etc. are based on the assumption that the parent population from which the sample have been drawn follows Normal distribution.

SGOU

iv. It is extensively used in large sampling theory to find estimates of parameters from statistics, confidence limits etc.

v. Normal distribution has the remarkable property stated in the central limit theorem. As per the theorem, when the sample size is increased, the sample means will tend to be normally distributed. The central limit theorem gives the normal distribution its central place in the sampling theory, since this can solve many important problems.

vi. In theoretical statistics as well as applied works, many problems can be solved only under the assumptions of a normal population.

vii. The normal distribution has numerous mathematical properties, making it popular and comparatively easy to manipulate. The normal curve is reasonably close to many distributions.

viii. It finds many applications in statistical quality control and industrial experiments. Many distributions in social and economic data are approximately normal e.g. Birth, death, etc. are normally distributed. In psychological and educational data many distributions are of normal type.

Area under the Normal Curve

The area under the normal curve gives us the proportion of the cases falling between two numbers or the Probability of getting a value between two numbers. Irrespective of the value of the mean (μ) and standard deviation (σ), for a normal distribution, the total area under the curve is 1.00.

The area under the normal curve is approximately distributed by its standard deviation as follows:

$\mu \pm 1\sigma$ covers 68.27 % area, i.e., 34.135% area will lie on either side of μ

$\mu \pm 2\sigma$ covers 95.45 % area, i.e., 47.725% will lie on either side of μ

$\mu \pm 3\sigma$ covers 99.73 % area, i.e., 49.865 % will lie on either side of μ

1.2.21 Normal Approximation to Binomial Distribution

The continuous normal distribution can sometimes be used to approximate the discrete binomial distribution. It is often used in statistical inference. The binomial distribution is perfectly symmetric if p = 0.5 and has some skewness when p ≠ 0.5. The normal approximation works best when p is close to 0.5 and when n is large. For a binomial random variable X.

$\mu = np$ and $\sigma^2 = np\ (1\text{-}p)$

$z = \dfrac{x - \mu}{\sigma}$ has (approximately) the standard normal distribution

$z \sim N\ (0, 1)$

If p and q are nearly equal (i.e., p is nearly 1/2), then the normal approximation is surprisingly good even for small values of n. However, when p and q are not equal, i.e., when p and q are small, even then the Binomial distribution tends to the normal distribution, but in this case, the convergence is slow. By this, we mean that if p and q are not equal, then for the Binomial distribution to tend to the Normal distribution, we need a relatively larger value of n as compared to the value of n required in case when p and q are nearly equal. Thus, the normal approximation to the Binomial distribution is better for increasing values of n and is exact in the limiting cases as $n \to \infty$.

1.2.22 Normal approximation to Poisson distribution

The Poisson distribution helps you understand the likelihood of getting a specific number of rare events in a fixed amount of time or space. Sometimes, when you are dealing with a lot of events that are quite rare, it becomes challenging to use the Poisson distribution directly. So, statisticians use a trick to make the calculations easier. If the average number of events (mean) is relatively large (like more than 10), they can pretend that the Poisson distribution looks like a normal distribution (bell curve).

Conditions for Approximation:

- ◊ **Large Mean:** If, on average, you get a lot of events, the normal approximation tends to work well.
- ◊ **Continuity Correction:** When making approximations for specific values, a small adjustment is made to the boundaries to make the calculations smoother.

The normal distribution is mathematically convenient and has well-known properties, making calculations easier. It is like using a good approximation when math gets too complicated. The approximation is better suited for large average event counts and rare and independent events. If the events are not rare or the mean is small, it is usually better to stick with the original Poisson distribution.

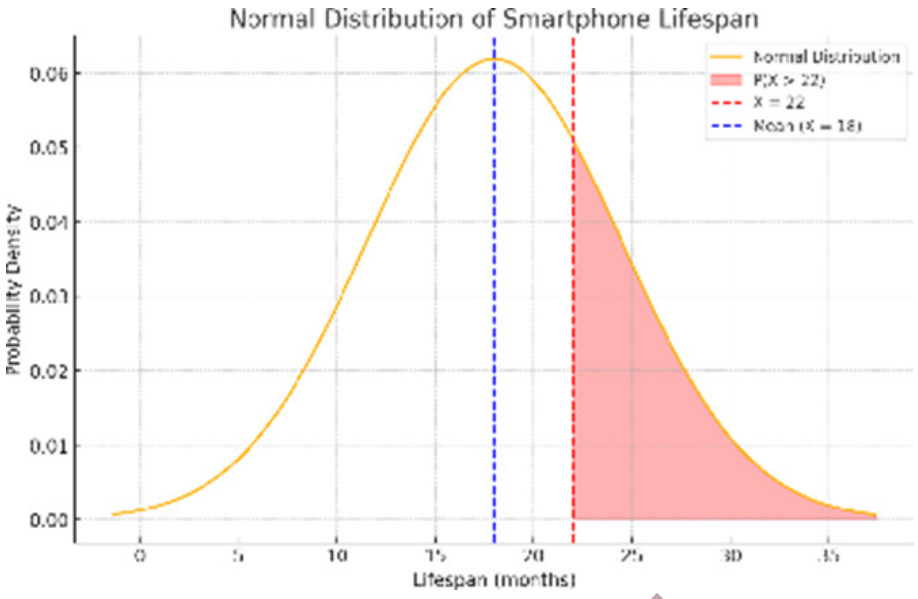
Illustration 1.2.10

An economist asserts that the average lifespan of a particular smartphone model is 18 months, with a standard deviation of 6.45 months. What is the Probability that a smartphone purchased by a customer will still be functional after 22 months?

Solution

For finding the required Probability, we are interested in the area of the problem of the standard curve, as shaded and shown below:

SGOU



$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

σ = Standard Deviation

$$Z = \frac{22 - 18}{6.45}$$

$$= 0.62$$

The value from the Standard Normal Distribution Table showing the area under the normal curve for Z = 0.62 is 0.2324. This means that the area of the curve between μ = 18 and x = 22 is 0.2324. Hence, the area of the shaded portion of the curve is (0.5-0.2324) = 0.2676, since the area of the entire right-hand portion of the curve always happens to be 0.5. Thus, the Probability that the smartphone will still be functional after 22 months is 0.2676.

Illustration 1.2.11

Suppose the waist measurement (w) of 800 girls is normally distributed with a mean 66 cms and a standard deviation of 5 cms. Find the number of girls with waists-

i. Between 65cms and 70 cms;

ii. Greater than or equal to 72 cms.

Solution

Mean , μ = 66

σ = 5cms

SGOU

W (In cms)	$z = \frac{x - \mu}{\sigma}$ μ = Mean σ = Standard Deviation
65	$\frac{65-66}{5} = 0.2$
70	$\frac{70-66}{5} = 0.8$
72	$\frac{72-66}{5} = 1.2$

i. The Probability that a girl has a waist between 65cms and 70 cms is given by:

$P(65 \leq W \leq 70)$

$= P(-0.2 \leq Z \leq 0.8)$

$= P(-0.2 \leq Z \leq 0) + P(0 \leq Z \leq 0.8)$

$= P(0 \leq Z \leq 0.2) + P(0 \leq Z \leq 0.8)$

$= 0.0793 + 0.2881$

$= 0.3674$

Hence the group of 800 girls, the expected number of girls with waists between 65 cms and 70 cms is 800 x 0.3674 = 293.92 =294

ii. The Probability that a girl has a waist greater than or equal to 72 cms is given by

$P(W \geq 72) = P(Z \geq 1.2)$

$= 0.5 - P(0 \leq Z \leq 1.2)$

$= 0.5 - 0.3849$

$= 0.1151$

Hence, in a group of 800 girls, the expected number of girls with a waist greater than or equal to 72 cms is 800 x 0.1151 = 92.08 = 92

Illustration 1.2.12

Assume the mean height of NCC volunteers to be 68.22 inches with a standard deviation of 3.286 inches. How many NCC volunteers in a school of 1000 would you expect to be

SGOU

i. Over 6 feet tall

ii. Below 5.5 feet

Assume heights to be normally distributed.

Solution

Let the variable X denote the height (in inches) of the NCC volunteers. Then we are given:

Mean = μ = 68.22 and standard deviation = σ = 3.286

An NCC volunteer will be over 6 feet tall if X is greater than

12 x 6 = 72 (Because x is height in inches)

When X =72,

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

σ = Standard Deviation

$$= \frac{72\text{-}68.22}{3.286} = \frac{3.78}{3.286} = 1.15$$

The Probability that an NCC volunteer is over 6 feet = 72" tall is given by

$$P(X > 72) = (Z > 1.15)$$

$$= 0.5 - P(0 \leq Z \leq 1.15)$$

$$= 0.5 - 0.3749 = 0.1251$$

Here, in a school of 1000 NCC volunteers, the number of volunteers over 6 feet tall is:

$$= 1000 \times 0.1251$$

$$= 125.1$$

$$= 125$$

The probability that a NCC volunteer is below 5.5 = 66 is given by: P (X< 66) = P $(Z < \frac{66 - 68.22}{3.286})$

$$= P(Z < \frac{-2.22}{3.286})$$

$$= P(Z < -0.6756) = P(Z > 0.6756)$$

$$= 0.5\text{-}P(0<Z<0.6756)$$

$$= 0.5\text{-}0.2501$$

$$= 0.2499$$

SGOU

Hence, the number of soldiers over 5.5 feet in a regiment of 1000 soldiers is

1000 x 0.2499 = 249.9 = 250

Illustration 1.2.13

The hourly wages of 1000 workmen are normally distributed around a mean of ₹70, with a standard deviation of ₹5. Estimate the number of workers whose hourly wages will be:

i. Between ₹69 and ₹72

ii. More than ₹75

iii. Less than ₹63

iv. Also estimate the lowest hourly wages of the 100 highest paid workers

Solution

Let the random variable X denote the hourly wages in rupees. Then X is a normal variable with $\mu = 70$ and $\sigma = 5$. The standard normal variable corresponding to X is

X	$z = \frac{x - \mu}{\sigma}$ μ = Mean σ = Standard Deviation
63	-1.4
69	-0.2
72	0.4
75	1

i. P (69 < X <72)

= P (-0.2 < Z < 0.4)

= P (-2 < 2 < Z < 0) + P (0 < Z < 0.4)

= P (0 < Z < 0.2) + P (0 < Z <0.4)

= 0.0793 + 0.1554

= 0.2347

Hence, the required number of workers is 1000 x 0.2347 =234.7 = 235

ii. We want P (X >75)

P (X > 75) = P (Z > 1)

= 0.5 – P (0 < Z < 1)

= 0.5 -0.3413

SGOU

= 0.1587

Thus, the number of workers with hourly wages more than 75 is 1000 x 0.1587 = 158.7 = 159

iii. P (X < 63)

= P (Z < -1.4)

= P (Z > 1.4)

= 0.5 -P (0 < Z < 1.4)

= 0.5 – 0.4192

= 0.0808

Hence, the number of workers with hourly wages less than ₹ 63 is 1000 × 0.808 = 80.8 =81

Proportion of the 100 highest paid workers is $\frac{100}{1000}=\frac{1}{10}=0.10$

We want to determine X = x_1, Such that (P X > x_1) = 0.10

When X = x_1,

$$Z=\frac{x-70}{5}=z_{1\ (Say)}$$

Then P (Z > z_1) = 0.10 → P (0 < Z < z_1) = 0.5 -0.1 = 0.40

From the normal probability table, we get

$$Z=\frac{x-70}{5}=1.28 \rightarrow x_1 = 70 + 5 \times 1.28 = 70 + 6.48 = 76.40$$

Hence, the lowest hourly wages of the 100 highest paid workers are ₹ 76.40

Summarised Overview

The probability distributions in statistics, beginning with the Binomial Distribution, which models outcomes of a Bernoulli process, a sequence of independent trials with success or failure, are defined by fixed trials, constant success probability, and known mean and standard deviation. It is widely used in scenarios like quality control and fitting real data to binomial patterns. The Poisson Distribution is introduced next, modelling rare events over time or space where occurrences are independent and the mean equals the variance, which is especially useful for traffic or call arrivals. Finally, the Normal Distribution, represented by its bell-shaped curve, describes continuous data with symmetrical

SGOU

properties around the mean, and finds practical use in natural and social sciences. The syllabus also covers how the binomial and Poisson distributions can approximate the normal distribution under certain conditions, enhancing analytical flexibility when working with large datasets or complex probabilities.



Self-Assessment Question

1. How is Probability defined?
2. What is meant by a random experiment?
3. Can you explain conditional Probability?
4. What are mutually exclusive events?
5. What is a complementary event in Probability?
6. How do we define independent events?
7. What are dependent events in Probability?
8. What does the term sample space refer to?
9. What is meant by a specific event?
10. What is an impossible event in Probability?
11. What is a discrete probability distribution?
12. What steps are involved in fitting a binomial distribution to observed data?
13. How is the binomial distribution defined?
14. How do you calculate a binomial distribution's expected value and standard deviation?
15. What is a random variable in probability theory?
16. What are the key assumptions behind the binomial distribution?
17. What are the practical uses of the binomial distribution?
18. What are the main properties of the binomial distribution?
19. How is the Poisson distribution defined?
20. What type of probability distribution is the Poisson model?

SGOU

21. What are the applications of the Poisson distribution?
22. What are the defining features of the Poisson distribution?
23. How does the Poisson distribution differ from the binomial distribution?
24. What is the process for fitting a Poisson distribution to data?
25. What conditions must be met to apply the Poisson distribution?
26. Why is the Poisson distribution referred to as the Law of rare events?
27. What is the normal distribution?
28. What are the key traits of the normal distribution?
29. In what way does the normal distribution serve as an approximation to the binomial distribution?
30. Under which conditions is the normal approximation to the binomial distribution most accurate?
31. How does the area under the normal curve relate to Probability and proportion?
32. When can the normal distribution be used to approximate the Poisson distribution?
33. What are the benefits of using the normal distribution in statistical analysis and decision-making?



Assignments

1. A card is drawn at random from an ordinary pack of 52 cards. Find the Probability that the card drawn is either a spade or a diamond. **(Answer- 1/2)**
2. A bag contains 5 white, 2 black, 3 yellow and 3 red balls. What is the Probability of getting a white or a red ball at random in a single draw of one? **(Answer- 8/13)**
3. There are 6 candies in a bowl, 3 are red, 2 are yellow, and 1 is blue. What is the probability of picking a yellow candy? **(Answer- 1/3)**

SGOU

4. A survey of the residents of a particular city showed that 20% preferred a white telephone over any other colour available. What is the Probability that more than one half of the next 20 telephones installed in this city will be white. **(Answer- 0.0006)**

5. The Probability that a patient recovers from a delicate heart operation is 0.9. What is the Probability that exactly 5 of the next 7 patients having this operation survive? **(Answer- 0.124)**

6. A multiple-choice quiz has 15 questions, each with 4 possible answers of which only one is correct. What is the Probability that sheer guesswork is:

    a. Exactly 2 correct answers **(Answer- 0.15)**

    b. 5-10 correct answers **(Answer- 0.31)**

    c. No correct answers **(Answer- 0.0.134)**

7. Eight unbiased coins were tossed simultaneously. Find the Probability of getting

    a. Exactly 4 heads **(Ans. 0.273)**

    b. No head at all **(Ans 0.0039)**

    c. 6 or more heads **(Ans 0.144)**

    d. Number of heads varying from 3 to 5 **(Ans 0.710)**

8. The Probability that a person dies from a certain respiratory infection is 0.002. Find the Probability that fewer than 5 of the next 2000 persons so infected will die. **(Answer- 0.6288)**

9. Suppose on average, one person in every 1000 is an alcoholic. Find the Probability that a random sample of 8000 people will yield 6 to 8 alcoholics. **(Answer- 0.4013)**

10. State the relationship between the Binomial distribution and the Poisson distribution.

11. If 3% of electric bulbs manufactured by a company are defective, find the Probability that in a sample of 100 bulbs, exactly 5 bulbs are defective. **(Answer: 0.1006)**

12. Give a normal distribution with Mean-300 and Standard deviation =50, what is the Probability that x assumes a value greater than 362? **(Answer: 0.1075)**

SGOU

13. Given a normal distribution with μ= 40 and σ= 6, find the value of x that has
    - ◊ 5% of area above the curve **(Answer: 49.87)**
    - ◊ 38% of area below the curve **(Answer: 38.17)**
14. Given a normal distribution with μ= 50 and σ= 10, find the Probability that x assumes a value:
    - ◊ 50-65 **(Answer: 0.4332)**
    - ◊ 50-55 **(Answer: 0.1915)**
15. In a normal distribution, 31% of the items are under 45 and 8 % are over 64. Find the mean and standard deviation of the distribution. **(Answer: Mean 50, SD 10)**
16. A normal distribution has 77 as its mean. Find its standard deviation if 20% of the area under the curve lies to the right of 90. **(Answer: SD 15.48)**
17. Marks obtained by several students are assumed to be normally distributed with a mean of 50 and a variance 36. If 4 students are taken at random, what is the Probability that exactly two of them will have marks over 62? **(Answer: 0.00298)**
18. The probability that a batsman scores a century in a cricket match is 1. What is the probability that in 4 matches he will score centuries in at least 3 innings. **(Ans:- 0.97^4)**
19. A box containing 100 transistors, 20 of which are defective, 10 are selected for inspection. Indicate what is the probability that (i) all 10 are defective (ii) 10 are good (iii) at least one is defective and (iv) at the most three are defective? **(Ans:- 1/9)**
20. It is found that the number of accidents occurring in a factory follows Poisson distribution with a mean of 2 accidents per week. Find the probability that (i) no accident occurs in a week (ii) number of accidents in a week exceeds 2 (Given that $e^{-2} = 0.135$) **(Ans:- 0.135,0.325)**
21. A manufacturer of pins knows that on an average 5% of his products is defective. He sells pins in boxes of 100 and guarantee that no more than 4 pins will be defective. What is the probability that a box will meet guaranteed quality ($e^{-5} = 0.0067$) **(Ans:- 0.438)**
22. During two hours between 8 and 10 AM on an average 1.5 number of phone calls per minute are reported in the switch board of a com-pany. Find the probability that during one particular minute, in the said interval, there will be (i) no phone call at all (ii) exactly 3 calls and (iii) at least 4 calls. (Given $e^{-1.5} = 2231$) **(Ans:- (i) 0.2231, (ii) 0.1255,(iii) 0.07)**

SGOU

23. The mean height of the soldiers is 68.22 inches with a variance of 10.8 (in). How many soldiers in a regiment of 1000 would you expect to be over 6 feet tall? [Hint: 1 feet 12 inches] **(Ans:- 125)**

24. The Bombay Municipal Corporation installed 2000 bulbs in the streets of Bombay. If these bulbs have an average life of 1000 burning hours with a standard deviation of 200 hours, what number of bulbs might be expected to fail in the first 700 burning hours? **(Ans:- 134)**



Reference

1. Aggarwal, B. M. (2010). *Business statistics*. Ane Books Pvt. Ltd.
2. Chandrasekharan, N., & Umaparvathi, M. (2016). *Statistics for management*. PHI Learning Pvt. Ltd.
3. Gupta, M., & Khanna, R. (2011). *Quantitative techniques for decision making*. Prentice Hall of India.
4. Gupta, S. P. (2021). *Statistical methods*. S. Chand & Sons.
5. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International (P) Ltd.
6. Levin, R. I., & Rubin, D. S. (1999). *Quantitative approaches for management*. Pearson Education.
7. Sharma, A. (2017). *Quantitative techniques for decision making*. Himalaya Publishing House.
8. Vohra, D. V. D. (2017). *Quantitative techniques for management*. New Delhi: McGraw Hill Education



Suggested Reading

1. Arora, P. N., Arora, S., & Arora, S. (2010). *Comprehensive statistical methods*. S. Chand & Company Pvt. Ltd.
2. Beri, G. C. (2017). *Business statistics*. Tata McGraw-Hill.
3. Chawla, D., & Sondhi, N. (2016). *Research methodology: Concepts and cases*. Vikas Publishing House.

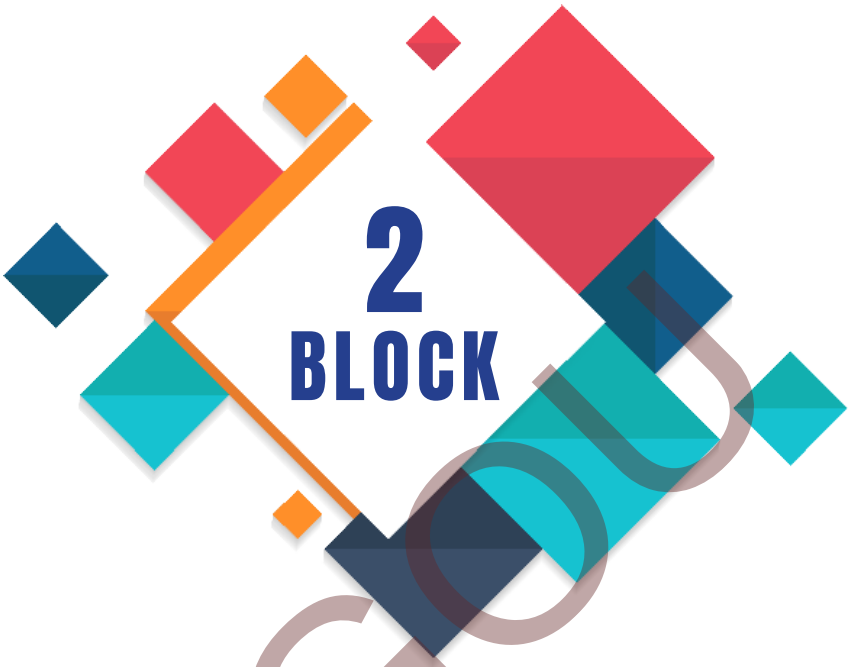
SGOU

4. Krishnaswami, O. R., Ranganathan, M., & Harikumar, P. N. (2016). *Research methodology*. Himalaya Publishing House.

5. Render, B. (2022). *Quantitative analysis for management*. Prentice Hall of India.

6. Singh, A. K. (2017). *Tests, measurements and research methods in behavioural sciences*. Bharti Bhawan.

SGOU



Parametric and Non-Parametric Tests

SGOU

Unit 1

Fundamentals of Hypothesis Testing and Parametric Tests

Learning Outcomes

On completion of this unit, learners will be able to:

- ◊ introduce the difference between sample statistics and population parameters, and explain the role of standard error in statistical inference
- ◊ make them aware of the hypothesis testing process thoroughly, including formulating hypotheses, determining significance levels, and recognising common errors such as Type I and II
- ◊ explain the proficiency in selecting and applying appropriate parametric tests, such as Z-tests, t-tests, F-tests, and tests for proportions, based on data characteristics and distributional assumptions
- ◊ narrate the properties of the normal distribution in the context of inferential statistics, including critical values and confidence intervals
- ◊ familiarise Analysis of Variance (ANOVA) to evaluate differences in group means, and understand its extension of t-tests in multi-group comparisons

SGOU

Background

The foundations of hypothesis testing and parametric methods lie in the early development of modern statistics during the late 19th and early 20th centuries. The concept of statistical inference, concluding populations based on sample data, was formalised by pioneers such as Ronald A. Fisher, Karl Pearson, and Jerzy Neyman. Fisher introduced the notion of null hypothesis and p-values in the 1920s, laying the groundwork for the standard approach to testing statistical Significance. Around the same time, Pearson contributed essential tools such as the chi-square test, while Neyman and Egon Pearson advanced the framework by distinguishing between Type I and Type II errors and developing the concept of power in hypothesis testing.

Parametric tests evolved alongside these theoretical advancements, grounded in assumptions about the underlying population distribution, most notably, normality. The Z-test and t-test emerged as essential tools for comparing means when population variances were known or unknown, respectively, with William Sealy Gosset ("Student") developing the t-distribution to handle small samples. The F-test and Analysis of Variance (ANOVA), credited to Fisher, provided a robust mechanism for comparing multiple groups means, particularly in experimental designs. As computational capabilities grew throughout the 20th century, these techniques became cornerstones of statistical practice in fields as varied as economics, psychology, biomedicine, and business research, informing decisions through structured, evidence-based reasoning. This unit will give an idea about the principles and practices of statistical inference, focusing on hypothesis testing and parametric techniques used to analyse and interpret data.



Keywords

Statistical Inference, Significance Testing, Parametric Analysis, Error Types, Z-test, t-test, F-test. ANOVA



Discussion

Hypothesis testing is a fundamental statistical tool used in business management to make informed decisions based on data. It involves formulating two competing hypotheses the null Hypothesis (H_0), which represents a default or status quo assumption, and the alternative Hypothesis (H_1), which reflects a new claim or effect. Managers use hypothesis testing to assess whether observed differences or relationships in data are statistically significant or likely due to chance. For example, a retail chain might hypothesise that a new store layout increases average customer spending. By collecting sample data before and after the redesign, they can test whether the observed increase is statistically significant, thereby guiding future layout decisions.

Parametric tests are a category of hypothesis tests that assume the underlying data follows a specific distribution, typically the normal distribution, and that certain parameters (like mean and variance) are known or estimated. These tests are powerful when their assumptions are met and include methods such as the t-test, ANOVA, and regression analysis. In business management, a common parametric test is the independent samples t-test, used to compare the mean performance of two groups. For instance, a company might use a t-test to evaluate whether the average sales generated by employees who received a new training programme differ significantly from those

SGOU

who did not. If the test reveals a statistically significant difference, it supports the effectiveness of the training intervention.

2.1.1. Basic Concepts

a. Statistic and Parameter

A parameter is a function of the population values. It is a statistical measure derived from the population. A parameter is usually denoted using Greek letters.

- ◊ Population mean is a parameter which is represented using the symbol 'μ'.
- ◊ Population Standard Deviation 'σ'.

A function of the sample values is a sample statistic. It is a statistical measure derived from the sample. A sample statistic is usually denoted using English letters.

- ◊ Sample mean is a statistic which is represented using the letter ' $\bar{X}$'.
- ◊ Sample Standard Deviation 's'.

b. Statistical Inference

Statistical inference refers to the process of selecting and using a sample statistic to conclude the population parameter. Statistical inference deals with two types of problems.

They are:-

- ◊ Testing of Hypothesis
- ◊ Estimation of Parameters

Test of hypothesis

A statistical test of hypothesis is a systematic procedure in which a statistical hypothesis is formulated and then accepted or rejected on the basis of a random sample drawn from the population. The methods used to accept or reject a hypothesis are known as statistical tests of hypothesis. Commonly used statistical tests include the Z-test, t-test, chi-square (χ^2) test, and F-test.

c. Statistical Hypothesis

In any test of a hypothesis, we begin with an assumption about the population from which the sample is drawn. The assumption may be about the form of the population or about the parameter of the population. This assumption is called a hypothesis.

Example: We may be interested in testing whether a coin is unbiased or whether the average height of people in a group.

A statistical hypothesis can be defined as the tentative conclusion logically drawn concerning any parameter or the form of distribution of the population.

SGOU

Simple and Composite Hypotheses

If a hypothesis is concerning the population completely Such as functional form and the parameter, it is called simple hypothesis.

If a hypothesis is not simple, then it is called composite hypothesis.

Eg: The hypothesis population is normal with mean 25 and SD 10 is a simple hypothesis, while the hypothesis "population follows normal distribution with mean 25" is a composite hypothesis.

Parametric and Non-Parametric Hypotheses

A hypothesis which specifies only the parameters of the probability density function is called a parametric hypothesis.

A hypothesis specifies only the form of the density function in the population, it is called a non-parametric hypothesis.

Eg: The hypothesis of population that "mean of hypothesis is 25 "is parametric while the hypothesis that population is Normal" is non-parametric.

Null and Alternative hypotheses

In a test of hypothesis, two different hypotheses are laid down.

The two hypotheses are so constructed that if one hypothesis is accepted, the other is rejected. These two hypotheses are known as Null hypoth e sis and alternative hypothesis. The null hypothesis is denoted by 'Ho' and the alternative is denoted by H1.

Eg: When we want to test whether the population mean is 65, the null hypothesis is that population mean is 65, and alternative hypothesis is that the population mean is not 65,

From the above discussion the null hypothesis can be defined as the statistical hypothesis which is stated for the purpose of possible acceptances. A null hypothesis is original hypothesis. Any hypothesis Other than null hypothesis is called an alternative hypothesis. So, when the null hypothesis is rejected, we accept the other hypothesis is known as alternative hypothesis.

Level of Significance

Confidence with which the null hypothesis is accepted on rejected depends on what is called Significance level. The probability which we may reject a null hypothesis when it is true, is called the level of significance.

So, when the level of significance is 0.5, it means that in the long run the stratification is rejecting true null hypothesis 5 times out of every 100 times. The level of significance is denoted by ∞.

SGOU

Degree of Freedom

It is defined as no: of independent observation which is obtained by Subtracting no: of constraints from the total no of observations.

Type I and Type II Errors

Type I error is committed by rejecting null hypothesis even when the null hypothesis is true. Similarly type II error is committed by accepting a null hypothesis when it is false.

Type I error -> Rejecting Ho when it is true.

Type II error -> Accepting Ho when Ho is false.

For eg: Suppose the test conducted on the basis of a sample drawn from the population mean ≠ 125 even though the population. is 125. This is type I error. Similarly on the basis of the sample drawn, if the test conducted leads to the conclusion that the test conducted population mean is 125 even though population mean is not 125, it is type II errors.

Standard Erro□

The standard deviation of a sampling distribution is called Standard Errors (S.E) of the statistic whose sampling distribution is it. Therefore, Standard error of sample of sample mean is $\sigma/\sqrt{n}$

Use of Standard Error

Standard Error plays a very important role in the large sample theory and forms the basis of the testing of hypothesis.

1. It is used for testing a given hypothesis.
2. SE gives an idea about, the reliability of a sample. The reciprocal of SE is a measure of reliability of the sample.
3. SE can be used to determine the confidence limits for population Values like mean, population and SD.

Critical Region

The range of variation is divided into two regions acceptance region and rejection region. It is the critical value which separates the acceptance region. If the computed value of the test statistic falls in the rejection region; we reject the null hypothesis. The rejection region is known as the Critical Region corresponding to a predetermined level of significance (Say ∞) and the acceptance region corresponding to region '1-∞'.

SGOU

Two Failed and One Tailed Test

A two tailed test of hypothesis will reject the null hypothesis if the Computed value of the test statistic is significantly greater than or lower than the Critical value. Thus, in two tailed test the critical region is represented by both tails.

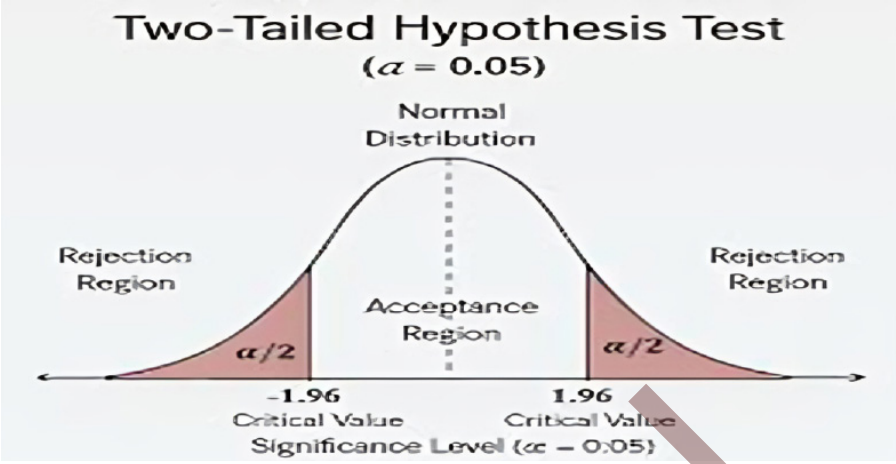


Fig. 2.1.1 Two Tailed Test

In one tailed test, the rejection region will be located in only one tail which may be either left or right, depending on alternative hypothesis. Thus, in one tailed test Critical region is represented by only one tail.

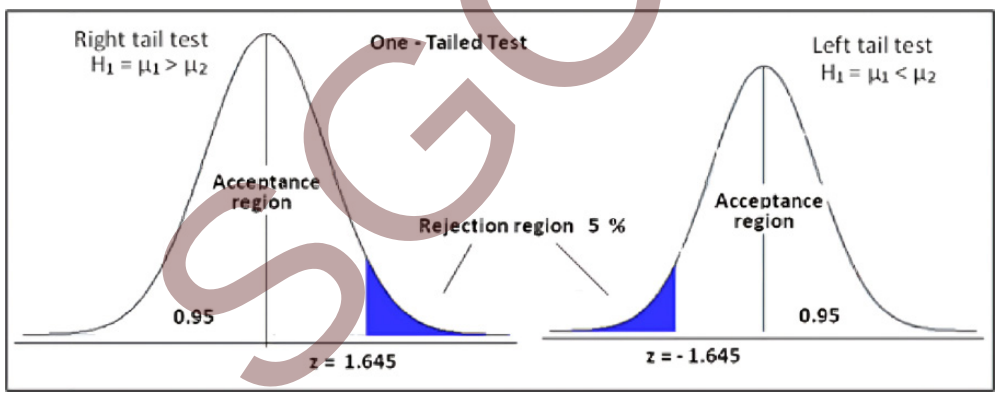


Fig. 2.1.2 One - Tailed Test

Power of a test

Probability for rejecting the null hypothesis when the alternative hypothesis is true is called a power of test

Power of a test = 1 - P (type II error)

Procedure for Testing a Hypothesis

The following steps are generally followed in hypothesis testing:

1. **Formulate the hypotheses:** Set up a null hypothesis (H_0) and an alternative hypothesis (H_1) appropriate to the problem under study.

SGOU

2. **Select an appropriate test:** Choose a suitable test criterion such as the Z-test, t-test, chi-square (χ^2) test, or F-test, depending on the nature of the data and the objective of the test.

3. **Specify the level of significance:** Decide the level of significance (α), which is commonly set at 5% or 1%. In the absence of specific instructions, a 5% level of significance is generally adopted.

4. **Compute the test statistic:** Calculate the value of the test statistic using the appropriate formula.

5. **Determine the critical (table) value:** Obtain the critical value of the test statistic corresponding to the chosen level of significance and the relevant degrees of freedom.

6. **Make a decision:** Compare the calculated test statistic with the critical value. If the computed value falls within the acceptance region, the null hypothesis is accepted. If it falls within the rejection region, the null hypothesis is rejected.

2.1.2 Parametric Tests

Parametric tests are statistical tests that assume the data follows a specific distribution, most commonly the normal distribution. These tests are based on assumptions about population parameters such as the mean and variance, and they are considered highly powerful and efficient when these assumptions are satisfied. However, if the underlying assumptions are violated, the results of parametric tests may not be reliable. Statistical tests that rely on the assumption of a normally distributed population or population parameters are known as parametric tests, and the most important among them include the Z-test, t-test, and F-test.

a. Z-test

Z-test is applied when the test statistic follows a normal distribution. It was developed by Prof. R.A. Fisher. The following are the important uses of the z-test: -

i. To test the population mean when the sample is large or when the population standard deviation is known.

ii. To test the equality of two sample means when the samples are large or when the population standard deviations are known.

iii. To test the population proportion.

iv. To test the equality of two population proportions.

v. To test the equality of correlation coefficients.

The z-test is used in testing of hypothesis on the basis of some assumptions. The important assumptions in the z-test are: -

i. Sampling distribution of the test statistic is normal.

SGOU

ii. For finding standard error, sample statistics are used in place where population parameters are to be used.

b. t-test

t- distribution was originated by W.S. Gosset in the early 1900. t-test is applied when the test statistic follows a t-distribution.

Uses of the t-test are: -

i. To test the given population, the mean when the sample is small, and the population S.D. is unknown.

ii. To test the equality of two sample means when the samples are small, and the population S.D. is unknown.

iii. To test the difference in values of two dependent samples.

iv. To test the significance of correlation coefficients.

The following are the important assumptions in t- test: -

i. The population from which the sample drawn is normal.

ii. The sample observations are independent.

iii. The population S.D. (σ)is unknown.

iv. Have a similar amount of variance within each group being compared

2.1.3 Testing of Given Population Mean

This testing of hypothesis is used to test whether the given population mean is true or not. In other words, it is used to test whether there is significant difference between sample mean and population mean.

Procedure

1. Set up H_0 and H_1

H_0: There is no significant difference between sample mean and population mean (i.e., $\mu = \mu_0$)

H_1: There is significant difference between sample mean and population mean (i.e., $\mu \neq \mu_0$)

2. Decide the test statistic

The test statistic applicable here is z-test or t-test. If population S.D (i.e., σ) is known, apply z-test. . If population S.D. (i.e., σ) is unknown but sample is large, apply z-test. If population S.D. (i.e., σ) is unknown but sample is small, apply t-test.

SGOU

3. Apply the appropriate formula for computing the value of the test statistic

Compute the test statistics = $$\frac{\bar{X} - \mu 0}{Standard\ error}$$

$\bar{X}$ is the sample mean and μ_0 is the population mean.

Standard Error = $\frac{\sigma}{\sqrt{n}}$ (If population S.D. is known)

$\frac{s}{\sqrt{n}}$ Standard Error = (If population S.D. is unknown, but sample is large)

Standard Error = $\frac{s}{\sqrt{n-1}}$ (If population S.D. is unknown and sample is small)

Where,

σ = population SD.

s = sample S.D.

n = sample size

4. Specify the level of Significance

If nothing is mentioned about the level of Significance, take 5%.

5. Fix the degree of freedom

For z-test, d.f. = infinity; For t-test, d.f = n-1

6. Locate the table value (critical value)

Table value of the test statistic at specified significance level and fixed degree of freedom.

7. Compare the calculated value of test statistic with the table value and decide whether to accept or reject the null hypothesis.

If calculated value of the test statistic is numerically less than the table value, the null hypothesis is accepted. If calculated value of the test statistic is numerically more than the table value, the null hypothesis is rejected.

Illustration 2.1.1

The mean life of random sample of 100 tyres is 15269 km. The manufacturer claims that the average life of tyres manufactured by the company is 15200 km with S.D. of 1248 km. Test the validity of company's claim.

SGOU

Solution

H_0: There is no significant difference between sample mean and population mean (i.e., $\mu = 15200$)

H_1: There is significant difference between sample mean and population mean (i.e., $\mu \neq 15200$)

Since population S.D is known, the test statistic applicable here is z-test

$$z = \frac{\bar{x} - \mu 0}{Standard\ error}$$

$$\bar{X} - \mu = 15269 - 15200$$

$$= 69$$

$$S\,E = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{1248}{\sqrt{100}} = 124.8$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$z = \frac{69}{124.8}$$

$$= 0.553$$

Level of significance = 5%

Degree of freedom = infinity (population S D is known)

Table value (Critical Value) at 5 % level of significance and infinity degree of freedom is 1.96. Since the calculated value of z is numerically less than the critical Value, H_0 is accepted. That is, there is no significant Difference between the sample mean and the population mean μ=15200. So, we may conclude that the claim of company is valid.

Illustration 2.1.2

A sample of size 400 was drawn, with a mean of 99. Test whether this sample could have come from the normal population with mean = 100 and S.D. = 8 at 5% significance level.

Solution

H_0: There is no significant difference between sample mean and population mean (i.e.; $\mu = 100$)

H_1: There is significant difference between sample mean and population mean (i.e., $\mu \neq 100$)

SGOU

Since population S.D. is known, the test statistic applicable here is Z-test

$$z = \frac{\bar{x} - \mu 0}{Standard\ error}$$

$\bar{X} - \mu = 100 - 99 = 1$

$$S\,E = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{8}{\sqrt{400}} = 0.4$$

$$z = \frac{1}{0.4} = 2.5$$

Level of significance = 5%

Degree of freedom= infinity (population S.D. is known)

Table value (Critical Value) at 5 % level of significance and infinity degree of freedom is 1.96. Since calculated value of z is more than the critical Value, H_0 is rejected. H_1 is accepted. That is, there is significant difference between sample mean and population mean. So, we may conclude that $\mu \neq 100$

Illustration 2.1.3

A random sample of 200 talcum powder bottles gave an average weight of 49.5 grams with a S.D. of 2.1 grams. Do we accept the hypothesis that the weight per bottle is 50 grams at 1% significance level?

Solution

H_0: There is no significant difference between sample mean and population mean (i.e., $\mu = 50$)

H_1: There is significant difference between sample mean and population mean (i.e., $\mu \neq 50$)

Since sample is large, the test statistic applicable here is z-test.

$$z = \frac{\bar{x} - \mu 0}{Standard\ error}$$

$\bar{X} - \mu = 49.5 - 50 = -0.5$

$$S\,E = \frac{s}{\sqrt{n}}$$

$$= \frac{2.1}{\sqrt{200}}$$

SGOU

$= 0.148$

$$z = \frac{-0.5}{0.148} \text{ (Calculated value)} = -3.38$$

Level of significance = 1%

Degree of freedom= infinity (population is large)

Table value (Critical Value) at 1% level of significance and infinity degree of freedom is 2.58. Since calculated value of z is more than the critical Value, H_0 is rejected. H_1 is accepted. That is, there is a significant difference between the sample mean and the population mean. So, we may conclude that $\mu \neq 50$ grams.

Illustration 2.1.4

The average life of 26 electric bulbs were found to be 1200 hours with a standard deviation of 150 hours. Test whether these bulbs could be considered as a random sample from a normal population with mean 1300 hours.

Solution

Ho: $\mu = 1300$

Since sample is small and population S.D is not known, apply - t-test.

$$t = \frac{\bar{X} - \mu}{Standard\ error}$$

where $\bar{x} = 1200$, $\mu = 1300$ and

$$\text{S.E.} = \frac{s}{\sqrt{n-1}} = \frac{150}{\sqrt{26-1}} = \frac{150}{\sqrt{25}} = 30$$

$$t = \frac{1200 - 1300}{30} = -3.3$$

For 't - test', degree of freedom = n-1 ie 26-1=25

Level of significance = .05

Table value of t at 0.05 level of significance = 2.06

Calculated value is greater than the table value numerically.

Therefore, we reject the null hypothesis.

The bulbs could not be a random sample from a normal population with a mean of 1300 hours.

One-tailed tests

The questions discussed so far relate to two-tailed tests, since in all those cases the alternative hypothesis is expressed as $\mu \neq \mu_0$. However, when the alternative hypothesis is directional, such as $\mu > \mu_0$ or $\mu < \mu_0$, the test is known as a one-tailed test.

SGOU

In other words, two-tailed tests are applied when testing for equality or any difference in either direction, whereas one-tailed tests are used when testing for conditions such as *more than, less than, superior, inferior, promotes* etc in a specific direction.

Illustration 2.1.5

A stenographer claims that she can take dictations at the rate of more than 120 words per minute. Of the 12 tests given to her she could perform an average of 135 words with a standard deviation of 40. Is her claim valid.

Solution

This is one-tailed test as the claim is more than 120.

Null hypothesis (H_0) is $\mu=120$

Alternative hypothesis (H_1) is $\mu > 120$ (ie more than 120)

$$t = \frac{\bar{X} - \mu}{Standard\ error}$$

where $\bar{x} = 135$, $\mu = 120$

$$S.E. = \frac{s}{\sqrt{n-1}} = \frac{40}{\sqrt{12-1}} = \frac{40}{\sqrt{11}} = 12.06$$

Degree of freedom = n-1 = 11

Level of significance = .01 (given)

Table value of t=2.718 (for one tailed test)

Calculated value of t is less than table value numerically.

Therefore, we accept H_0

ie, $\mu = 120$ (μ is not greater than 120)

Her claim that she can take dictation at the rate of more than 120 is not valid.

2.1.4 Testing Significance of the Difference between Two Population Means (or Testing Equality of two population means)

Procedure

H_0: There is no significant difference between two means (i.e., $\mu_1 = \mu_2$)

H_1: There is a significant difference between two means (i.e., $\mu_1 \neq \mu_2$)

1. Decide the test statistic

The test statistic applicable here is z-test or t-test.

SGOU

If population S.D. (i.e., σ) is known, when samples are large apply z-test, otherwise t-est.

If population S.D. (i.e., σ) is unknown but sample is large, apply z-test

If population S.D. (i.e., σ) is unknown but sample is small, apply t-test

2. Apply the appropriate formula for computing the value of the test statistic

$\frac{z}{t}$ = Difference/Standard Error

Difference = Difference between two sample means

Test statistic = $\frac{\bar{x}_1 - \bar{x}_2}{SE}$ where $\bar{x}_1$ and $\bar{x}_2$ are sample means.

Standard Error = $\sigma\sqrt{\frac{1}{n1} + \frac{1}{n2}}$ (If two samples are drawn from the same SD)

Standard Error = $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1}}$ (If population S.D. are known)

Standard Error = $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}}$ (If population S.D. are unknown, but samples are large)

Standard error = $\sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{n_1 + n_2 - 2} \times \frac{1}{n_1} + \frac{1}{n_2}}$

(If population S.D. are unknown, and samples are small)

Where,

σ_1 = population S.D. of sample 1

s_1 = sample S.D. of sample 1

n_1 = sample size of sample 1

σ_2 =population S.D. of sample 2

s_2 = sample S.D. of sample 2

n_2 = sample size of sample 2

3. Specify the level of Significance

If nothing is mentioned about the level of significance, take 5%.

SGOU

4. Fix the degree of freedom

For z-test, d.f = infinity; For t-test, d.f = $n + n_2 - 2$

5. Locate the table value

Critical value of the test statistic at specified level of significance and fixed degree of freedom.

6. Compare the calculated value of test statistic with the table value

Compare the calculated value of test statistic with the table value and decide whether to accept or reject the null hypothesis. If calculated value of the test statistic is numerically less than the table value, the null hypothesis is accepted. If calculated value of the test statistic is numerically more than the table value, the null hypothesis is rejected.

Illustration 2.1.6

The mean yield of wheat from District I was 210Kg per acre from a sample of 100 plots. In another District II, the mean yield was 200 Kg per acre from a sample of 150 plots. Assuming that the S.D. of yield of the entire State was 11 Kg, test whether there is any significant difference between the mean yields of the crop in the two districts.

Solution

District I	District II
$n_1 = 100$	$n_2 = 150$
$\bar{x}_1 = 210$	$\bar{x}_2 = 200$
$\sigma_1 = 11$	$\sigma_2 = 11$

H_0: There is no significant difference between two sample means (i.e., $\mu_1 = \mu_2$)

H_1: There is significant difference between two sample means (i.e., $\mu_1 \neq \mu_2$)

Since population S.D. are given, the test statistic applicable here is z-test.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{SE}$$

$\bar{x}_1 - \bar{x}_2 = 210 - 200 = 10$

$$SE = \sigma\sqrt{\frac{1}{n1} + \frac{1}{n2}} = 11\sqrt{\frac{1}{100} + \frac{1}{150}}$$

$$= 11\sqrt{.01 + .007}$$

$$= 11 \times 0.1304 = 1.43$$

$$Z = \frac{10}{1.42} = 7.04$$

Table value of z at 5% level of significance and infinity degrees of freedom = 1.96

Since the calculated Value of Z is more than the table value, H_0 is rejected. We accept H_1. So we may conclude that there is a significant difference in the mean yields of crops in two districts.

Illustration 2.1.7

Electric bulbs manufactured by X Ltd. and Y Ltd. gave the following results:

Particulars	X Ltd	Y Ltd
Number of bulbs used	100	100
Mean Life in Hours	1300	1248
Standard Deviation	82	93

State whether there is any significant difference in the life of bulbs of the two makes.

Solution

H_0: There is no significant difference between two sample means (i.e; $\mu_1=\mu_2$)

H_1: There is significant difference between two sample means (i.e., $\mu_1 \neq \mu_2$)

Since population S.D. are unknown but samples are large, the test statistic applicable here is z-test.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{SE}$$

$\bar{x}_1 - \bar{x}_2 = 1300 - 1248 = 52$

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}} \text{ (If population S.D.s are unknown)}$$

$$= \sqrt{\frac{82^2}{100} + \frac{93^2}{100}}$$

$$= \sqrt{67.24 + 86.49}$$

$$= \sqrt{153.73}$$

$$= 12.4$$

$$z = \frac{52}{12.4} = 4.19$$

Level of significance = 5%, Degree of freedom = infinity

Table value of Z at 5% level of significance and infinity degrees of freedom = 1.96

SGOU

Since the calculated Value of Z is more than the table value, H_0 is rejected. We accept H_1. (i.e., $\mu_1 \neq \mu 2$). So we may conclude that there is a significant difference in the mean life of bulbs of the two makes.

Illustration 2.1.8

Two batches of same product are tested for their mean life. Assuming that lives of the two products follow a normal distribution, test the hypothesis that the mean life is same for both the batches, given the following information:

Batch	Sample Size	Mean life in hours	S.D
A	10	750	12
B	8	820	14

Solution

H_0: There is no significant difference between two sample means (i.e: $\mu_1=\mu_2$)

H_1 : There is significant difference between two sample means (i.e., $\mu_1 \neq \mu_2$)

Since population S.D. are unknown and samples are small, the test statistic applicable here is t-test.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$$

$\bar{x}_1 - \bar{x}_2 = 820 - 750 = 70$

$$\text{Standard error} = \sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{n_1 + n_2 - 2} \times \frac{1}{n_1} + \frac{1}{n_2}}$$

$$= \sqrt{\frac{(10 \times 12^2 + 8 \times 14^2)}{10+8-2} \times \frac{1}{10} + \frac{1}{8}}$$

$$= \sqrt{\frac{3008}{16} \times \frac{1}{10} + \frac{1}{8}}$$

$$= \sqrt{188 \times \frac{1}{10} + \frac{1}{8}}$$

$$= \sqrt{42.3} = 6.5$$

$$t = \frac{70}{6.5} = 10.77 \text{ (Calculated value) Level of significance = 5\%}$$

Degree of freedom = 10 + 8 - 2 = 16

Table value of t at 5% level of significance and 16 degrees of freedom =2.12

Since the calculated value of t is more than the table value, H_0 is rejected. We accept H_1. (i.e; $\mu 1 \neq \mu 2$).

SGOU

So, we may conclude that the lives of products produced in two batches are not same.

Illustration 2.1.9

In a test given to 2 groups of students, the marks obtained were as follows:

Group I:	18	20	36	50	49	36	34	49	41
Group II:	29	26	28	35	30	44	46		

Test whether the group means are equal.

Solution

Here we have to find the Means and S.D.s of the two samples**.**

Computation of Mean and S.D. of two Groups					
Group I			**Group II**		
X	**X - $\bar{X}$**	**(x- $\bar{X}$)2**	**X**	**X - $\bar{X}$**	**(x- $\bar{X}$)²**
18	-19	361	29	-5	25
20	-17	289	26	-8	64
36	-1	1	28	-6	36
50	13	169	35	1	1
49	12	144	30	-4	16
36	-1	1	44	10	100
34	-3	9	46	12	144
49	12	144			
41	4	16			
∑X =333		**1134**	**∑X = 238**		**386**

Mean of Group I ($\bar{X}$) = $\frac{333}{9}$ = 37

Mean of Group II($\bar{X}$) = $\frac{238}{7}$ =34

S.D of Group I = $\sqrt{\frac{\sum(x-\bar{x}_1)^2}{n}}$

= $\sqrt{\frac{1134}{9}}$

= $\sqrt{126}$

S_1 =11.22

SGOU

$$\text{S.D of Group II} = \sqrt{\frac{\sum(x-\bar{x}_2)^2}{n}}$$

$$= \sqrt{\frac{386}{7}}$$

$$= \sqrt{55.14}$$

$$S_2 = 7.42$$

H_0:There is no significant difference between two sample means (i.e., $\mu_1 = \mu_2$)

H_1: There is significant difference between two sample means (i.e., $\mu_1 \neq \mu_2$)

Since population S.D.s are unknown and samples are small, the test statistic applicable here is t-test.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$$

$$\bar{x}_1 - \bar{x}_2 = 37 - 34 = 3$$

$$SE = \sqrt{\frac{(n_1 s_1^2 + n_2 s_2^2)}{n_1 + n_2 - 2} \times \frac{1}{n_1} + \frac{1}{n_2}}$$

$$= \sqrt{\frac{9 \times 126 + 7 \times 55.14)}{2} \times \frac{1}{9} + \frac{1}{7}}$$

$$= \sqrt{\frac{1520}{14} \times 0.254}$$

$$= \sqrt{27.58}$$

$$= 5.25$$

$$t = \frac{3}{5.25} = 0.571 \text{ (Calculated Value) Level of significance = 5\%}$$

Degree of freedom= 9 + 7 - 2 = 14

Table value of t at 5% level of Significance and 14 degrees of freedom = 2.145

Since the calculated value of t is less than the table value, H_0 is accepted. (i.e., $\mu_1 = \mu_2$). So, we may conclude that the difference in the group means are not significant. They are equal.

SGOU

2.1.5 Testing of Significance of the Difference in case of Dependent Samples (Paired observations)

Here, the observations in one sample are somehow related to the observations in the other. Therefore, they are called paired observations. The test statistic applicable here is t-test.

Procedure

a. Set up H_0 and H_1

H_0: There is no significant difference between samples

H_1: There is significant difference between samples

b. Decide test statistic

Since the paired data are comparatively less, the test statistic applicable here is always t-test.

c. Apply the appropriate formula for computing the value of the test statistic

$$t = \bar{d}/SE$$

Where,

$\bar{d}$ is Arithmetic mean of the difference between the values

SE is $\frac{s}{\sqrt{n-1}}$ [s = standard deviation of the difference]

d. Specify the level of Significance

Take 5%, if nothing is mentioned in the question.

e. Fix the degree of freedom

d.f = n – 1 , where n= Number of pairs of observations.

f. Locate the critical value of the test statistic (t-test) at specified level of significance and fixed degree of freedom

g. Compare the calculated value of test statistic with the table value

Compare the calculated value of test statistic with the table value and decide whether to accept or reject the null hypothesis. If calculated value of the test statistic is numerically less than the table value, the null hypothesis is accepted. If calculated value of the test statistic is numerically more than the table value, the null hypothesis is rejected.

Illustration 2.1.10

The marks scored by 10 students, before and after providing special coaching, are given in the following table:

SGOU

Before:	67	24	57	55	63	54	56	68	33	43
After:	70	38	58	58	56	67	68	72	42	38

Test whether there is any significant difference in their performance.

Solution

H_0: There is no significant difference between samples

H_1: There is significant Difference between samples

Test statistic applicable here is t-test.

Computation of the mean and standard deviation of the difference between the values			
Score (Before)	**Score (After)**	**Difference (d)**	**d^2**
67	70	3	9
24	38	14	196
57	58	1	1
55	58	3	9
63	56	-7	49
54	67	13	169
56	68	12	144
68	72	4	16
33	42	9	81
43	38	-5	25
		$\sum d = 47$	**$\sum d^2 = 699$**

Arithmetic mean of d values $= \bar{d} = \frac{47}{10} = 4.7$

$$\text{SD of values} = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$$

$$= \sqrt{\frac{699}{10} - \left(\frac{47}{10}\right)^2}$$

$$= \sqrt{69.9 - 22.09}$$

$$= \sqrt{47.81}$$

$$= 6.91$$

$$SE = \frac{6.91}{\sqrt{10-1}} = 2.3$$

$$t = \frac{4.7}{2.3} = 2.04$$

SGOU

Level of significance = 5% Degree of freedom = 10-1 = 9

Table value (critical value) of t at 5% level of Significance and 9 degree of freedom is 2.262. Since the calculated value of t is less than the critical value, the null hypothesis is accepted. So, we may conclude that there is no significant difference in the performance of the students.

2.1.6 Testing of Given Population Proportion

This type of testing of hypothesis is used to test whether there is any significant difference between the sample proportion and the given population proportion.

Procedure

a. Set up H_0 and H_1

H_0: There is no significant difference between sample proportion and population proportion

(i.e., H_0: P = p)

H_1: There is significant difference between sample proportion and population proportion

(i.e., H_0: P ≠ p)

b. Decide the test statistic

The test statistic applicable here is Z-test

c. Apply appropriate formulae for computing the Value of Z (i.e., calculated value)

$$Z = \frac{D}{SE}$$

$$\text{i.e., } Z = \frac{p - P}{SE}$$

Where p = sample proportion, P =Population proportion

$$S\,E = \sqrt{\frac{PQ}{n}} \text{ Where, } Q = 1 - P$$

d. Decide the level of Significance

Take 5%, if nothing is mentioned in the question.

SGOU

e. Fix the degree of freedom

For z-test,d.f. = ∞

f. Locate the table value of Z

Locate the value at specified level of Significance and fixed degree of freedom.

g. Compare the calculated Value of Z with the table value and decide whether to accept or reject the null hypothesis

If calculated Value of Z is numerically less than the table value, the null hypothesis is accepted. If calculated Value of Z is numerically more than the table value, the null hypothesis is rejected.

Illustration 2.1.11

It is found that out of 500 units of a product produced by a machine, 30 are defectives. Test whether the machine produces 2% defective items on an average.

Solution

H_0: There is no significant difference between sample proportion and population proportion(i.e., H_0: P =0.02)

H_1: There is significant difference between sample proportion and population proportion (i.e., H_0:P≠ 0.02)

$$Z = \frac{p - P}{SE}$$

P = 0.02,

$$p = \frac{30}{500} = 0.06,$$

Q = 1 – P

=1 - 0.02

= 0.98,

n = 500

$$SE = \sqrt{\frac{PQ}{n}}$$

$$= \sqrt{\frac{0.02 \times 0.98}{500}}$$

$$= \sqrt{\frac{0.0196}{500}}$$

SGOU

$$= \sqrt{0.0000392}$$

$$= 0.0063$$

$$Z = \frac{0.06 - 0.02}{0.0063} = \frac{0.04}{0.0063} = 6.349$$

Level of significance = 5% degree of freedom = infinity

Table value of Z at 5% level of significance and infinity degree of freedom is 1.96.

Since the calculated Value of Z is more than the table value, null hypothesis is rejected. We accept alternate hypothesis. $P \neq 0.02$. So, it is not possible to think that the machine produces 2% defective items.

2.1.7 Testing Equality of Two Population Proportions

Procedure

a. Set up H_0 and H_1

H_0: There is no significant difference between the two population proportions (i.e., $P_1 = P_2$)

H_1: There is a significant difference between the two population proportions (i.e., $P_1 \neq P_2$)

b. Decide the test statistic

The test statistic applicable here is Z-test.

c. Apply the appropriate formula for computing the value of the test statistic

$$Z = \text{Difference/Standard Error i e; } Z = \frac{p_1 - p_2}{SE}$$

Where p_1 and p_2 are the proportions of two samples

$$SE = \sqrt{p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Where,

$$p_0 = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$q_0 = 1 - p_0$$

$n_1 = number\ of\ events\ in\ sample\ number\ one$

SGOU

$n_2 = number\ of\ events\ in\ sample\ number\ two$

d. Specify the level of Significance

If nothing is mentioned about the level of significance, take 5%.

e. Fix the degree of freedom (d.f. = infinity)

f. Locate the table value (critical value)

Locate the table value (critical value) of the test statistic at specified level of significance and fixed degree of freedom.

g. Compare the calculated value of test statistic with the table value

Compare the calculated value of test statistic with the table value and decide whether to accept or reject the null hypothesis. If calculated value of the test statistic is numerically less than the table value, the null hypothesis is accepted. If calculated value of the test statistic is numerically more than the table value, the null hypothesis is rejected.

Illustration 2.1.12

In a sample of 800 people selected from district X, 400 were regular drinkers of coffee. In another sample of 1000 people drawn from district Y, 450 were regular drinkers of coffee. Test whether there is significant difference between the two districts, regarding the coffee drinking habit of people.

Solution

H_0: There is no significant difference between two districts regarding the coffee drinking habits of people (i.e., $P_1 = P_2$)

H_1: There is significant difference between two districts regarding the coffee drinking habits of people (i.e., $P_1 \neq P_2$)

The test statistic applicable here is Z-test.

$$Z = \text{Difference/Standard Error i e; } Z = \frac{P_1 - P_2}{SE}$$

$$P_1 = \frac{400}{800} = 0.5$$

$$P_2 = \frac{450}{1000} = 0.45$$

$$SE = \sqrt{p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad n_1 = 800,\ n_2 = 1000$$

$$p_0 = \frac{(800 \times 0.5 + 1000 \times 0.45)}{(800 + 1000)}$$

$$= \frac{850}{1800} = 0.472$$

$$q_0 = 1 - 0.472 = 0.528$$

$$\therefore SE = \sqrt{0.472 \times 0.528 \left(\frac{1}{800} + \frac{1}{1000}\right)}$$

$$= \sqrt{0.249\ (0.00125 + 0.001)}$$

$$= \sqrt{0.249 \times 0.00225}$$

$$= \sqrt{0.00056}$$

$$= 0.0237$$

$$\therefore Z = \frac{(0.5 - 0.45)}{0.0237}$$

$$= \frac{0.05}{0.0237}$$

$$= 2.11$$

Level of Significance = 5%.

Fix the degree of freedom = infinity

Table value (critical Value) of Z at 5% level of significance and infinity degrees of freedom is 1.96

The null hypothesis is rejected since the calculated Value of Z is more than the table value. The alternate hypothesis is accepted. $p_1 \neq p_2$. So, we may conclude that there is a significant difference between the two districts regarding people's coffee drinking habits.

2.1.8 F - Test (Variance Ratio Test)

Variance Ratio Test

The test based on the test statistics which follows F distribution is called F-test. To test whether there is a significant difference between the variance of two populations based on a small sample, we can use the F-test. Let n_1 and n_2 be the size of the samples and s_1^2 and s_2^2 be the variance of the samples respectively. The F-statistic (or F-ratio) formula is generally given as the ratio of the larger variance to the smaller one.

$$F = \frac{n_1 s_1^2/(n_1 - 1)}{n_2 s_2^2/(n_2 - 1)}$$

SGOU

or simply the ratio of two sample variances:

$$F = \frac{s_1^2}{s_2^2} (\text{where } s_1^2 > s_2^2)$$

From the value of the F-statistic, we can conclude that the population variances are equal or that the two samples are selected from populations with the same variance.

F-test is used to test the equality of variance of two populations and to test the equality of means of three or more variances (ANOVA).

F-test is based on the following assumptions:

1. Normality: Values in each group are normally distributed.
2. Homogeneity: Variances within each group should be equal (homoscedasticity).
3. Independence of Errors: Errors should be independent of each value.

The degrees of freedom for the F-test are $(n_1 - 1), (n_2 - 1)$

Illustration 2.1.13

The SD of two sample sizes 10 and 14 of two normal populations are 3.5 and 3 respectively. Examine whether the SD of the population are equal at 5% level of significance.

Solution

H_0 : The S.D of two populations are equal, $\sigma_1 = \sigma_2$ (or) $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1 \neq \sigma_2$

$$n_1 = 10, s_1 = 3.5$$

$$n_2 = 14, s_2 = 3$$

$$\frac{n_1 S_1^2}{n_1 - 1} = \frac{10 \times (3.5)^2}{10 - 1} = \frac{122.5}{9} = 13.61$$

$$\frac{n_2 S_2^2}{n_2 - 1} = \frac{14 \times (3)^2}{14 - 1} = \frac{126}{13} = 9.69$$

$$F = \frac{13.61}{9.69} = 1.40$$

SGOU

The degrees of freedom are $df = [(n_1 - 1), (n_2 - 1)] = (9,13)$

The table value of F at 5% level of significance with $df(9,13)$ is 2.72.

Table value (2.72) > Calculated value (1.40), hence we accept H_0

Conclusion: At a 5% level of significance, the S.D. of the two populations are equal.

2.1.9 Analysis of Variance

The testing of hypotheses so far discussed consists of different sample groups which do not exceed two. If there are three or more sample groups, testing equality cannot be done in any of the already discussed methods. The testing of the significance of the difference among three or more samples is generally done using the analysis technique. In case of analysis of variance, as part of testing procedure, we have to prepare a separate statement called Analysis of Variance Table or ANOVA Table. Therefore, this type of hypothesis testing is also called variance analysis. The test statistic used for Analysis of Variance is F-test. F-test is a parametric test.

Types of Analysis of Variance

There are two types of variance analysis. They are:

i. One-way classification of data (One-way analysis of variance)

ii. Two-way classification of data (Two-way analysis of variance)

2.1.9.1 One-way classification of data (One way analysis of variance)

In one way, observations are classified into groups based on a single criterion. Suppose we want to study about the yield of a particular crop. You know there a number of factors which influence the productivity of crops. If we undertake this study to know the effect of seed quality on crop yield, it is called one- way analysis of variance. Here yield of crops based on different seeds must be given in columns. In other words, in case of one-way analysis of variance, the samples must always be in columns.

Types of variances in One-way ANOVA

a. Variance between samples (Columns): This is the net result of the variation of different sample means from the grand mean. The grand mean is the mean of all the observations under all sample groups.

b. Variance within the sample: This is the net result of variations of different items of the sample from the respective sample means.

c. Variance about the sample: This is the sum of the variance between samples and the variance within the sample.

SGOU

Table 2.1.1 Proforma of One-way ANOVA Table

One-way ANOVA Table				
Source of variation	Sum of Squares	Degree of Freedom	Mean Sum of Squares	F-Ratio
Between samples	SSC	C – 1	MSC = $\frac{SSC}{C-1}$	F = $\frac{MSC}{MSE}$ Or $\frac{MSE}{MSC}$ F =
Within Sample	SSE	N – C	MSE = $\frac{SSE}{N-C}$	
Total	SST=	N –1		

SSC = Sum of Squares between Columns (Samples)

SSE = Sum of Square within Column (Sample)

SST = Sum of Square Total

MSC = Mean Sum of Squares between Columns (Samples)

MSE = Mean Sum of Squares within Column (Sample)

C = Number of Columns (Samples)

Procedure for carrying out One-way Analysis of variance

a. Set up H_0 and H_1

H_0: There is no significant difference between samples

H_1: There is significant difference between samples

b. Decide the test statistic

Test statistic applicable here is F-test

c. Apply the appropriate formula for computing the value of f-test

$$\frac{MSC}{MSE} \text{ or } \frac{MSE}{MSC}$$

i. Find SST

$$\text{SST} = \text{Sum of square of all items} - \frac{T^2}{N}$$

SGOU

Where,

T = Total of all observations,

N = Total Number of observations

$\frac{T^2}{N}$ is generally called correction factor

ii. Find SSC

$$SSC = \left[\frac{(\sum x_1)^2}{n_1}\right] + \left[\frac{(\sum x_2)^2}{n_2}\right] + \left[\frac{(\sum x_3)^2}{n_3}\right] + \ldots\ldots\ldots - \frac{T^2}{N}$$

Where,

$\sum x_1$= sum of items in the first column

$\sum x_2$= sum of items in the second column

n_1= number of items in the first column

n_2 = number of items in the second column

iii. Draw one-way ANOVA Table and enter the values of SST and SSC

iv. Find the Value of SSE

SSE = SST – SSC

v. Find the degree of freedom in the third column as indicated in the proforma

vi. Find MSC

$$MSC = \frac{SSC}{C-1}$$

vii. Find MSE

$$MSE = \frac{SSE}{N-C}$$

viii. Find F–Ratio

$\frac{MSC}{MSE}$ or $\frac{MSE}{MSC}$

d. Specify the level of significance

Take 5%if nothing is mentioned.

SGOU

e. Fix the degrees of freedom

Here we have to fix a pair of d.f. If 'F' is obtained by using F= $\frac{MSC}{MSE}$, then pair of d.f. is (d.f. of MSC, d f of MSE) If 'F' is obtained by using F= $\frac{MSE}{MSC}$, then pair of d.f. is (d.f. of MSE, d f of MSC)

f. Obtain table value of F at specified level of significance and fixed degree of freedom

g. Compare the Calculated Value of F with the Table value

Compare the Calculated Value of F with the Table value and decide whether to accept or reject the null hypothesis. If calculated value is less than the table value, H_0 is accepted. If calculated value is more than the table value, H_0 is rejected.

Illustration 2.1.14

Four varieties of a crop were grown on 3 plots, and the following yield was obtained. You are required to test whether there is a significant difference in the productivity of seeds:

Plot	Variety of Seeds			
	P	Q	R	S
I	10	7	8	5
II	9	7	5	4
III	8	6	4	4

Solution

H_0: There is no significant difference in the productivity of seeds.

H_1: There is significant difference in the productivity of seeds.

Test Statistic applicable here is F-test

$$\frac{MSC}{MSE} \text{ or } \frac{MSE}{MSC}$$

$$T = 10+7+8+5+9+7+5+4+8+6+4+4 = 77$$

$$SST = 10^2 + 7^2 + 8^2 + 5^2 + 9^2 + 7^2 + 5^2 + 4^2 + 8^2 + 6^2 + 4^2 + 4^2 - \frac{T^2}{N}$$

$$= 100+49+64+25+81+49+25+16+64+36+16+16 - \frac{77^2}{12}$$

$$= 541 - (5929/12) = 541 - 494 = 47$$

$$SSC = \left[\frac{(\sum x_1)^2}{n_1}\right] + \left[\frac{(\sum x_2)^2}{n_2}\right] + \left[\frac{(\sum x_3)^2}{n_3}\right] + \ldots\ldots\ldots - \frac{T^2}{N}$$

SGOU

$$= \left[\frac{(10+9+8)^2}{3}\right] + \left[\frac{(7+7+6)^2}{3}\right] + \left[\frac{(8+5+4)^2}{3}\right] + \left[\frac{(5+4+4)^2}{3}\right] - \frac{T^2}{N}$$

$$= \frac{729}{3} + \frac{400}{3} + \frac{289}{3} + \frac{169}{3} - \frac{77^2}{12}$$

$$= \frac{1587}{3} - 494$$

$$= 529 - 494$$

$$= 35$$

SSE = SST- SSC

= 47-35 = 12

Level of significance = 5%

One-way ANOVA Table				
Source of variation	Sum of Squares	Degree of Freedom	Mean Sum of Squares	F-Ratio
Between samples	SSC = 35	C – 1= 4 -1 = 3	MSC = $\frac{SSC}{C-1}$ =11.67	F = $\frac{MSC}{MSE}$
Within Sample	SSE = 12	N–C =12- 4 = 8	MSE = $\frac{SSE}{N-C}$ =1.5	$\frac{11.67}{1.5}$ = 7.78
Total	SST = 47	N –1=12-1 =11		

Degree of freedom = (3, 8)

Table value of F at 5% level of Significance and (3, 8) degrees of freedom is 4.07. Since calculated Value of F is more than the table value, H_0 is rejected. We accept alternate hypothesis. So, we may conclude that there is significant difference in the productivity of three varieties of seeds.

Illustration 2.1.15

The following table shows the yield of 3 varieties. Perform analysis of variance and test whether there is significant difference between varieties:

SGOU

Varieties	Plots				
	A	B	C	D	E
I	30	27	42		
II	51	47	37	48	
III	44	35	41	36	42

Solution

Here, we are asked to test whether there is significant difference between varieties. But varieties are given in rows, not in columns. In one way ANOVA, the samples must be in columns. Therefore, we have to rearrange the given data so as to bring the samples in columns as shown below:

Plots	**Varieties**		
	I	**II**	**III**
A	30	51	44
B C D	27	47	35
E	42	37	41
		48	36
		42	

H_0: There is no significant difference in the productivity of varieties.

H_1: There is significant difference in the productivity of varieties.

Test Statistic applicable here is F-test

$$F = \frac{MSC}{MSE} \text{ or } \frac{MSE}{MSC}$$

$$SST = 30^2 + 51^2 + 44^2 + 27^2 + 47^2 + 35^2 + 42^2 + 37^2 + 41^2 + 48^2 + 36^2 + 42^2 - \frac{T^2}{N}$$

$$= 900 + 2601 + 1936 + 729 + 2209 + 1225 + 1764 + 1369 + 1681 + 2304 + 1296 + 1764 - \frac{480^2}{12}$$

$$= 19778 - 19200 = 578$$

$$SSC = \left[\frac{(\sum x_1)^2}{n_1}\right] + \left[\frac{(\sum x_2)^2}{n_2}\right] + \left[\frac{(\sum x_3)^2}{n_3}\right] + \ldots\ldots\ldots - \frac{T^2}{N}$$

$$\left[\frac{(30+27+42)^2}{3}\right] + [(51+47+37+48+42)2/5] + \left[\frac{44+35+41+36^2}{4}\right] - \frac{480^2}{12}$$

$$= 9801/3 + 50625/6 + 24336/4 - 19200$$

SGOU

= 3267 + 10125 + 6084 – 19200

= 19476 –19200

= 276

SSE = SST- SSC

= 578-276

= 302

Level of significance = 5% degree of freedom = (2, 9)

One-way ANOVA Table				
Source of variation	Sum of Squares	Degree of freedom	Mean Sum of Squares	F-Ratio
Between samples	SSC = 276	3-1= 2	MSC = $\frac{276}{2}$ =138	
Within sample	SSE = 302	12-3 = 9	MSE = $\frac{302}{9}$ = 33.56	F= $\frac{138}{33.56}$ = 4.112
Total	SST = 578	12-1=1		

Table value of F at 5% level of Significance and (2, 9) degrees of freedom is 4.26. Since calculated Value of F is less than the table value, H_0 is accepted. So we may conclude that there is no significant difference in the productivity of three varieties.

2.1.9.2 Two-way classification of data (Two-way analysis of variance)

In two-way classification, observations are classified into different groups on the basis of two criteria. Consider the example mentioned in one- way classification. Suppose we study the effect of both the quality of seeds and the type of fertilisers on the productivity of crop. In that case, the data are to be classified based on two criteria: type of seed and type of fertiliser. This is called two- way analysis of variance. In the case of a two-way analysis of variance, we do not need to rearrange the given data. Since two criteria are considered, there will be two sets of hypotheses here.

Types of variances in Two-way ANOVA

a. **Variance between samples (Columns):** This is the net result of the variation of different sample means (in respect of columns) from the grand mean. The grand mean is the mean of all the observations under all sample groups.

b. **Variance between rows:** This is the net result of the variation of different sample means (in respect of rows) from grand mean.

SGOU

c. **Variance within the sample (Residual):** This is the net result of variations of different sample items from the respective sample means.

d. **Variance about the sample:** This is the sum of the variance between columns, variance between rows and the variance within the sample(residual).

Table 2.1.2 Proforma of Two-way ANOVA Table

SSC = Sum of Squares between Columns

Two-way ANOVA Table				
Source of variation	**Sum of Squares**	**Degree of Freedom**	**Mean Sum of Squares**	**F-Ratio**
Between Columns	SSC	c – 1	MSC = $\frac{SSC}{C-1}$	$F_C = \frac{MSC}{MSE}$ or $= \frac{MSE}{MSC}$
Between Rows	SSR	r – 1	MSR= $\frac{SSR}{R-1}$	$F_R = \frac{MSR}{MSE}$ or $= \frac{MSE}{MSR}$
Within Sample	SSE	(c – 1) x (r – 1)	MSE = $\frac{SSE}{(C-1)x\,(r-1)}$	
Total	SST	N –1		

SSR = Sum of Squares between Rows

SSE = Sum of Square within Samples

SST = Sum of Square Total

MSC = Mean Sum of Squares between Columns

MSR = Mean Sum of Squares between Rows

MSE = Mean Sum of Squares within Samples

c= Number of Columns

r = Number of Rows

SGOU

Procedure for carrying out Two-way Analysis of Variance

a. Set up H_0 and H_1

H_0: There is no significant difference between samples (in respect of columns)

H_1: There is a significant difference between samples (in respect of columns)

H_0: There is no significant difference between samples (in respect of rows)

H_1: There is a significant difference between samples (in respect of rows)

b. Decide the test statistic

Test statistic applicable here is F-test

c. Apply the appropriate formula for computing the values of F ratios

$$F_C = \text{. i.e., } \frac{MSC}{MSE} or \frac{MSE}{MSC}$$

$$F_R = \text{i.e., } \frac{MSR}{MSE} or \frac{MSE}{MSR}$$

i. Find SST

$$SST = \text{Sum of square of all items} - \frac{T^2}{N}$$

Where,

T = Total of all observations,

N = Total Number of observations

$\frac{T^2}{N}$ is generally called correction factor

ii. Find SSC

$$SSC = \left[\frac{(\sum x_1)^2}{n_1}\right] + \left[\frac{(\sum x_2)^2}{n_2}\right] + \left[\frac{(\sum x_3)^2}{n_3}\right] + \ldots\ldots\ldots - \frac{T^2}{N}$$

Where,

$\sum x_1$=sum of items in the first column

$\sum x_2$= sum of items in the second column

n_1 = number of items in the first column

n_2 = number of items in the second column

SGOU

iii. Find SSR

$$SSC = \left[\frac{(\sum x_1)^2}{n_1}\right] + \left[\frac{(\sum x_2)^2}{n_2}\right] + \left[\frac{(\sum x_3)^2}{n_3}\right] + \ldots\ldots\ldots - \frac{T^2}{N}$$

Where,

$\sum x_1$ = sum of items in the first row

$\sum x_2$ = sum of items in the second row

n_1 = number of items in the first row

n_2 = number of items in the second row

d. Draw ANOVA Table and enter the values of SST, SSC and SSR

e. Find the Value of SSE

SSE = SST – (SSC + SSR)

f. Find the degree of freedom in the third column as indicated in the proforma

g. Find MSC

$$MSC = \frac{SSC}{C\text{-}1}$$

h. Find MSR

$$MSR = \frac{SSR}{r\text{-}1}$$

i. Find MSE

$$MSE = \frac{SSE}{[(c-1)\,x\,(r-1)]}$$

j. Find F–Ratios (i.e., F_C and F_R)

$$F_C = [\text{i.e., } F = \frac{MSC}{MSE} \text{ or } \frac{MSE}{MSC}$$

$$F_R = [\text{i.e., } \frac{MSR}{MSE} \text{ or } \frac{MSE}{MSR}]$$

k. Specify the level of Significance

Take 5% if nothing is mentioned.

l. Fix the degrees of freedom

Fix a pair of d.f. in respect of F_C and F_R.

SGOU

m. Obtain table value of F_C and F_R at specified level of Significance and fixed degree of freedom

n. Compare the calculated value of Fc with the table value

Compare the calculated value of Fc with the table value, and decide whether to accept or reject the null hypothesis. If calculated value is less than the table value, H_0 is accepted. If the calculated value is more than the table value, H_0 is rejected.

o. Compare the calculated value of F_R with the table value

Compare the calculated value of F_R with the table value, and decide whether to accept or reject the null hypothesis. If the calculated value is less than the table value, H_0 is accepted. If calculated value is more than the table value, H_0 is rejected.

Illustration 2.1.16

Following table shows the yield of crops using 3 varieties of seeds:

Plots	Varieties of Seeds		
	P	Q	R
I	6	7	8
II	4	6	5
III	8	6	10
IV	6	9	9

Test whether there is significant difference in the productivity of varieties of seeds. Also test the significance of the difference between plots.

Solution

H_0: There is no significant difference in the productivity of varieties of seeds.

H_1: There is significant difference in the productivity of varieties of seeds.

H_0: There is no significant difference in the productivity from plots.

H_1: There is significant difference in the productivity from plots.

Test statistic applicable here is F-test F_C [i.e., F= $\frac{MSC}{MSE}$ or $\frac{MSE}{MSC}$

FR = ; [i.e., $\frac{MSR}{MSE}$ or $\frac{MSE}{MSR}$

SST = Sum of square of all items - $\frac{T^2}{N}$

$$= 6^2 + 4^2 + 8^2 + 6^2 + 7^2 + 6^2 + 6^2 + 9^2 + 8^2 + 5^2 + 10^2 + 9^2 - \frac{84^2}{12}$$

$$=36+16+64+36+49+36+36+81+64+25+100+81-\frac{7056}{12}$$

$$= 624 - 588$$

$$= 36$$

$$SSC = \left[\frac{(\sum x_1)^2}{n_1}\right] + \left[\frac{(\sum x_2)^2}{n_2}\right] + \left[\frac{(\sum x_3)^2}{n_3}\right] + \ldots\ldots\ldots \frac{T^2}{N}$$

$$= \left[\frac{(6+4+8+6)^2}{4}\right] + \left[\frac{(7+6+6+9)^2}{4}\right] + \left[\frac{(8+5+10+9)^2}{4}\right] + - \frac{84^2}{12}$$

$$= \frac{576}{4} + \frac{784}{4} + \frac{1024}{4} - 588$$

$$= 144 + 196 + 256 - 588 = 8$$

$$SSR = \left[\frac{(\sum x_1)^2}{n_1}\right] + \left[\frac{(\sum x_2)^2}{n_2}\right] + \left[\frac{(\sum x_3)^2}{n_3}\right] + \ldots\ldots\ldots \frac{T^2}{N}$$

$$= \left[\frac{(6+7+8)^2}{3}\right] + \left[\frac{(4+6+5)^2}{3}\right] + \left[\frac{(8+6+10)^2}{3}\right] + \left[\frac{(6+9+9)^2}{3}\right] \ldots - \frac{84^2}{12}$$

$$= \frac{441}{3} + \frac{225}{3} + \frac{576}{3} + \frac{576}{3} - 588$$

$$=147+75+192+192-588$$

$$= 18$$

$$SSE = SST - (SSC+SSR)$$

$$= 36 - (8+18) = 10$$

Between Columns

Two-way ANOVA Table				
Source of variation	Sum of Squares	Degree of Freedom	Mean Sum of Squares	F-Ratio
Between Columns	SSC=338.8	4 – 1 = 3	MSC = $\frac{338.8}{3}$ = 112.93	$F_C = \frac{MSC}{MSE}$ $\frac{112.93}{6.142}$ = 18.387

SGOU

Between Rows	SSR=161.5	5 – 1 = 4	MSR = $\frac{161.5}{4}$ = 40.375	
Within Sample	SSE=73.7	(4 – 1) x (5 –1) = 12	MSE = $\frac{73.7}{12}$ =6.142	$F_R = \frac{MSR}{MSE} = \frac{40.375}{6.142} = 6.574$
Total	**SST= 574**	**20 –1** = 19		

Calculated value of F_C = 2.395 Level of Significance = 5% Degrees of freedom = (2, 6) Table value of F_C at 5% level of Significance and (2, 6) degrees of freedom = 5.14 Since calculated value is less than the table value, null hypothesis is accepted. So we may conclude that there is no significant difference in the productivity of three varieties of seeds.

Between Rows

Calculated value of F_R = 3.593 Level of Significance = 5% Degrees of freedom = (3, 6) Table value of F_R at 5% level of Significance and (3, 6) degrees of freedom = 4.76

The null hypothesis is accepted since the calculated value is less than the table value. So we may conclude that there is no significant difference in the productivity from plots.

Coding Method

In analysis of variance, while preparing ANOVA table (one-way and two-way), we first have to find the values of SST, SSC, SSR, etc. However, if the individual observations of the given data are of large values, the computation of SST, SSC, SSR, etc. becomes tedious. So, as to avoid this complication, we may apply coding method. The coding method refers to the addition, subtraction, multiplication, and division of individual data observations using a constant. The addition, subtraction, multiplication or division of all the individual items by a constant will not affect the Value of F.

Illustration 2.1.17

The following table shows the number of units of a product produced by 5 workers using 4 different types of machines:

You are required to test:

Whether there is significant difference in the mean productivityof machines.

Whether there is significant difference in the mean productivity of workers.

SGOU

Workers	Machines			
	P	Q	R	S
I	44	38	47	36
II	46	40	52	43
III	34	36	44	32
IV	43	38	46	33
V	38	42	49	39

Solution

Let us apply coding method by subtracting 45 from each observation of the given data.Then we get;

Workers	Machines			
	P	Q	R	S
I	-1	-7	2	-9
II	1	-5	7	-2
III	-11	-9	-1	-13
IV	-2	-7	1	-12
V	-7	-3	4	-6

H_0: There is no significant difference in the productivity of machines.

H_1: There is significant difference in the productivity of machines.

H_0: There is no significant difference in the productivity of workers.

H_1: There is significant difference in the productivity of workers.

Test statistic applicable here is F-test

$$FC = [\text{i.e., } F = \frac{MSE}{MSC} \text{ or } \frac{MSC}{MSE}$$

$$F_R = [\text{i.e., } \frac{MSR}{MSE} \text{ or } \frac{MSE}{MSR}]$$

$$T = -1+-7+2+-9+1+-5+7+-2+-11+-9+-1+-13+-2+-7+1+$$
$$-12+-7+-3+4+-6 = 80$$

$$SST = \text{Sum of square of all items} - \frac{T^2}{N}$$

$$= -1^2 + 1^2 + -11^2 + -2^2 + -7^2 + -7^2 + -5^2 + -9^2 + -7^2 + -3^2$$

$$+ 2^2 + 7^2 + -1^2 + 1^2 + 4^2 + 9^2 + -2^2 + -13^2 + -12^2 + -6^2 - \frac{80^2}{20}$$

$$= 1+1+121+4+49+49+25+81+49+9+4+49+1+1+16+81+4+169+144+36 - \frac{6400}{20}$$

$$= 894 - 320 = 574$$

$$SSC = \left[\frac{(\sum x_1)^2}{n_1}\right] + \left[\frac{(\sum x_2)^2}{n_2}\right] + \left[\frac{(\sum x_3)^2}{n_3}\right] + \ldots\ldots\ldots \frac{T^2}{N}$$

$$= \left[\frac{(-1+1-11-2-7)^2}{5}\right] + \left[\frac{(-7-5-9-7-3)^2}{5}\right] + \left[\frac{(2+7-1+1+4)^2}{5}\right]$$

$$+ \left[\frac{(-9-2-13-12-6)^2}{5}\right] - \frac{80^2}{20}$$

$$= \frac{400}{5} + \frac{961}{5} + \frac{169}{5} + \frac{1764}{5} - 320$$

$$= 80+192.2+33.8+352.8 = 658.8$$

$$= 658.8 - 320 = 338.8$$

$$SSR = \left[\frac{(\sum x_1)^2}{n_1}\right] + \left[\frac{(\sum x_2)^2}{n_2}\right] + \left[\frac{(\sum x_3)^2}{n_3}\right] + \ldots\ldots\ldots \frac{T^2}{N}$$

$$= \left[\frac{(-1-7+2-9)^2}{4}\right] + \left[\frac{(1-5+7-2)^2}{4}\right] + \left[\frac{(-11-9-1-13)^2}{4}\right] +$$

$$\left[\frac{(-2-7+1-12)^2}{4}\right] + \left[\frac{(-7-3+4-6)^2}{4}\right] - \frac{80^2}{20}$$

$$= \frac{225}{4} + \frac{1}{4} + \frac{1156}{4} + \frac{400}{4} + \frac{144}{4} - 320$$

$$= 481.5 - 320$$

$$= 161.5$$

$$SSE = SST - (SSC+SSR)$$

$$= 574 - (338.8+161.5)$$

$$= 73.7$$

SGOU

Two-way ANOVA Table				
Source of variation	Sum of Squares	Degree of Freedom	Mean Sum of Squares	F- Ratio
Between Columns	SSC = 338.8	4 – 1= 3	MSC = $\frac{338.8}{3}$ =112.93	$F_C = \frac{MSC}{MSE}$ = $\frac{112.93}{6.142}$ =18.387
Between Rows	SSR = 161.5	5 – 1 = 4	MSR = $\frac{161.5}{4}$ = 40.375	
Within Sample	SSE = 73.7	(4 –1) * (5-1) = 12	$\frac{73.7}{12}$ MSE = = 6.142	$F_R = \frac{MSR}{MSE}$ = $\frac{40.375}{6.142}$ =6.574
Total	SST = 574	20 –1 =19		

Between Columns

Calculated Value of F_C = 18.387

Level of Significance = 5%

Degrees of freedom = (3, 12)

Table value of F_C at 5% level of significance and (3, 12)

Degrees of freedom = 3.49

Since calculated value is more than the table value, null hypothesis is rejected. Alternate hypothesis is accepted. So, we may conclude that there is significant difference in the mean productivity of machines.

Between Rows

Calculated Value of F_R = 6.574

Level of Significance = 5%

Degrees of freedom = (4, 12)

Table value of F_R at 5% level of significance and (4, 12)

Degrees of freedom = 3.26

SGOU

Since calculated value is more than the table value, null hypothesis is rejected. Alternative hypothesis is accepted. So, we may conclude that there is a significant difference in the mean productivity of workers.



Summarised Overview

The Fundamentals of Hypothesis Testing and Parametric Tests introduces learners to the foundational principles of statistical inference, beginning with the distinction between sample statistics (e.g., sample mean, sample proportion) and population parameters (e.g., population mean, proportion), which are central to concluding a population based on sample data. It covers the concept of standard error, which quantifies the variability of a sample statistic and underpins confidence intervals and hypothesis testing. The hypothesis testing process is explored in depth, including formulating null and alternative hypotheses, selecting appropriate test statistics, and decision-making based on calculated p-values or critical values. Learners are introduced to the level of significance (α), typically set at 0.05 or 0.01, which determines the threshold for rejecting the null hypothesis. The syllabus also explains Type I Error (false rejection of a true null hypothesis) and Type II error (failure to reject a false null hypothesis), highlighting the trade-offs between sensitivity and specificity in statistical decision-making.

The parametric tests, which assume underlying population distributions, typically assume normality. It begins with the normal distribution, a cornerstone of parametric inference, and progresses to specific tests: the Z-test for large samples and known population variance, the t-test for small samples or unknown variance (including one-sample, independent two-sample, and paired sample variants), and the F-test for comparing variances across groups. The test for proportion is included to evaluate hypotheses about population proportions. The Analysis of Variance (ANOVA) introduces a powerful technique for comparing means across multiple groups simultaneously, using the F-distribution to assess whether observed differences are statistically significant. The emphasis is placed on selecting appropriate tests based on data characteristics and research objectives, ensuring learners gain conceptual clarity and practical competence in applying parametric methods.

SGOU

Self-Assessment Question

1. Differentiate between sample statistic and population parameter.
2. What do you mean by statistical inference?
3. Explain the meaning and types of hypotheses.
4. What is sampling distribution?
5. Elucidate the process involved in testing hypotheses.
6. How does Type I error differ from Type II error?
7. What is a standard error?
8. State the symbols that are used to represent Type I and Type II errors.
9. What do you mean by parametric tests?
10. What is a Type I Error?
11. What is Type II error?
12. What do you mean by power of a test?
13. What is meant by critical region and acceptance region?
14. What is one tailed test?
15. What is two-tailed test?
16. What do you mean by analysis of variance?
17. Explain the two types of analysis of variance.
18. What are the different types of variances in case of one-way analysis of variance?
19. What are the different types of variances in case of two-way analysis of variance?
20. Draw the proforma of one-way ANOVA table.
21. Draw the proforma of two-way ANOVA table.
22. Explain the hypothesis testing procedure in case of one-way ANOVA.
23. Explain the hypothesis testing procedure in case of two-way ANOVA.
24. What do you mean by coding method in analysis of variance?

SGOU

Assignments

1. Identify and list the different formulas for calculating the test statistic.
2. Explain the Type I and Type II errors.
3. What is the concept of power of a test?
4. Develop Null Hypothesis and Alternate Hypothesis for a study that aims to test whether increase in income leads to increase in savings.
5. The average annual income of people in a particular locality is assumed to be ₹15,00,000. The Income Tax Authorities feel that there is gross understatement of income. In order to check this, the tax authorities selected a sample of 100 citizens and found their average annual income to be ₹18,00,000 with a standard deviation of ₹1,50,000. Verify the claim of tax authorities at 5% level of Significance. **(Answer- Reject null hypothesis)**
6. It is claimed that an automobile is driven on the average less than 20000 kilometres per year. To test this claim, a random sample of 100 automobile owners is asked to keep the record of the kilometres they travel. Would you agree with his claim if the random sample showed an average of 23500 kilometres and a standard deviation of 3900 kilometres. Use 0.01 level of Significance. **(Answer- Reject null hypothesis)**
7. In a certain college, it is estimated that fewer than 25% of students have cars on campus. Does this seem to be a valid estimate, if, in a random sample of 90 college students, 28 are found to have cars. Use 0.05 level of Significance. **(Answer- Accept null hypothesis)**
8. The average length of time of students who registered for a certain course in certain college has been 50 minutes. A new registration procedure using modern computing machine is being tried. If a random sample of 12 students had an average registration time of 42 minutes with a S.D. of 11.9 minutes, under the new system, test the hypothesis that the new system is effective using a level of Significance of 5 % **(Ans: Accept the null hypothesis, t-test)**
9. A sample of size 400 was drawn and the sample mean was found to be 99. Test whether this sample could have come from a normal population with mean 100 and S. D - 8 at 5% level of significance. **(Ans:- 2.5, no)**
10. A sample of 900 items is taken from a population with SD 15. The mean of the sample is 25. Test whether the sample has come from a population with mean 26.8. **(Ans:- Z=3.6, no)**

SGOU

11. A stenographer claims that she can take dictation at the rate of 120 words per minute. Can we reject her claim on the basis of 100 trails in which she demonstrates a mean of 116 words with a variance of 225 words (use 5% level of significance) **(Ans:- Z= -2.67, Reject)**

12. A random sample of 200 tins of cocunut oil gives an average weight of 4.95 kg with a S.D of 0.21 kg. Do we accept the hypothesis of net weight of 5 kgs. per tin at 1% level? **(Ans:- Z=-3.37, no)**

13. Two batches of same product are tested for their mean life. Assuming that lives of the product follows a normal distribution with an un known variance, test the hypothesis that the mean life is same for both the batches given the following information

Batch II	Sample size	Mean life in hrs	SD
I	10	750	12
II	8	820	14

Choose a significant level of 5% **(Ans:- t = 14.6, no)**

14. Of the two salesmen, X claims that he has made larger sales than Y For the accounts examined which were comparable for the two men, results were

	No. of sales	Average size	SD
X	10	6200	690
Y	17	5600	600

Do these average-sized sales figures differ significantly? Explain your result. **(Ans:- t= 2.28, yes)**

15. Set a table of analysis of variance for the following data.

	Varieties			
Plots	A	B	C	D
1	200	230	250	300
2	190	270	300	270
3	240	150	145	180

Test whether the varieties are different. **(Ans:- F = 3.85, Varieties are different)**

16. The following figures relate to production in kilogram of three varieties A, B and C of wheat sown in 12 plots.

SGOU

A	B	C
14	14	18
16	13	16
18	15	16
	22	19
		20

Is there significant difference in the production of three varieties. **(Ans:- F = 1.16, No significant difference)**

Reference

1. Aggarwal, B. M. (2010). *Business statistics*. Ane Books Pvt. Ltd.
2. Chandrasekharan, N., & Umaparvathi, M. (2016). *Statistics for management*. PHI Learning Pvt. Ltd.
3. Gupta, M., & Khanna, R. (2011). *Quantitative techniques for decision making*. Prentice Hall of India.
4. Gupta, S. P. (2021). *Statistical methods*. S. Chand & Sons.
5. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International (P) Ltd.
6. Levin, R. I., & Rubin, D. S. (1999). *Quantitative approaches for management*. Pearson Education.
7. Sharma, A. (2017). *Quantitative techniques for decision making*. Himalaya Publishing House.
8. Vohra, D. V. D. (2017). *Quantitative techniques for management*. New Delhi: McGraw Hill Education

Suggested Reading

1. Arora, P. N., Arora, S., & Arora, S. (2010). *Comprehensive statistical methods*. S. Chand & Company Pvt. Ltd.
2. Beri, G. C. (2017). *Business statistics*. Tata McGraw-Hill.

SGOU

3. Chawla, D., & Sondhi, N. (2016). *Research methodology: Concepts and cases*. Vikas Publishing House.

4. Krishnaswami, O. R., Ranganathan, M., & Harikumar, P. N. (2016). *Research methodology*. Himalaya Publishing House.

5. Render, B. (2022). *Quantitative analysis for management*. Prentice Hall of India.

6. Singh, A. K. (2017). *Tests, measurements and research methods in behavioural sciences*. Bharti Bhawan.

SGOU

Unit 2 Non-Parametric Tests and Their Applications

Learning Outcomes

On completion of this unit, learners will be able to:

- ◊ explain non-parametric statistical tests, including their advantages over parametric methods in analysing ordinal data and non-normally distributed samples
- ◊ familiarise the Chi-Square Test for Goodness of Fit, Test of Independence and Test of Homogeneity to evaluate categorical data using contingency tables, and interpret the statistical Significance of observed versus expected frequencies
- ◊ make them aware of rank-based, distribution-free tests such as the Sign Test (one-sample and two-sample), Wilcoxon Signed Rank Test, and Mann–Whitney U Test, Kruskal-Wallis H Test and Wald-Wolfowitz Runs Test

Background

The development of non-parametric statistical tests emerged in response to the limitations of classical parametric methods, which often rely on stringent assumptions such as normality, homogeneity of variance, and interval-level measurement. Early contributions can be traced to the work of Karl Pearson, whose Chi-Square Test (1900) laid the foundation for analysing categorical data through observed and expected frequencies. This test became instrumental in fields such as genetics, sociology, and market research, where data rarely conform to parametric norms. Over time, the Chi-Square Test has evolved to include applications in contingency tables and tests of independence, offering researchers a robust tool for hypothesis testing that relies on minimal distributional assumptions.

SGOU

The mid-20th century witnessed a surge in rank-based and distribution-free methods through the contributions of statisticians such as Frank Wilcoxon, Henry Mann, and Donald Whitney, leading to tests like the Wilcoxon Signed Rank Test, Mann-Whitney U Test, and Kruskal-Wallis H Test. The Sign Test and Wald-Wolfowitz Runs Test further expanded the toolkit for analysing medians and randomness in sequences. These methods gained prominence in behavioural sciences, education, and business research, where data often violate parametric assumptions. This unit equips learners with the conceptual understanding and practical skills to apply non-parametric tests across diverse research contexts.



Keywords

Chi-Square Test, Contingency Tables, Goodness of Fit, Rank-Based Tests, Wilcoxon and Mann–Whitney Tests, Kruskal–Wallis H Test, Wald Wolfowitz Runs Test



Discussion

Non-parametric tests are statistical methods that do not assume a specific probability distribution for the underlying data, making them particularly valuable when dealing with ordinal data, small sample sizes, or non-normal distributions. These tests are widely used in business management research where data often arises from customer surveys, employee feedback, or market segmentation studies, contexts in which parametric assumptions may not hold. Non-parametric methods thus offer flexibility and reliability in managerial decision-making, especially when working with qualitative or non-interval data.

2.2.1 Non-Parametric Test

Statisticians have devised an alternate procedure which can be used to test hypotheses about data which are not normally or for which meaningful Sample statistics cannot be calculated. Since these tests do not depend on the shape of the distribution, they are called distribution - free tests. These tests do not depend upon the population parameters. mean and the variance; they are also such as called non-parametric tests. One of the most important non-parametric techniques is Chi-Square test formulated by Karl Pearson in 1900.

SGOU

A large number of non-parametric tests exist. Some of them are us follows:

1. The Sign Test.
2. Rank Sum Test.
3. The One-Sample Runs Test.
4. The Kruskal-Wallis H Test.
5. The Spearman Rank Correlation Procedure.

Advantages of Non-parametric Tests

1. Non-parametric Tests are distribution free ie, they do not require any assumption to be made about the population following distribution. normal or any other.
2. Generally, they are simple to understand and easy to apply when the sample sizes are small.
3. Most non-parametric tests do not require lengthy and laborious computations and hence are less time consuming. If significant results are obtained, no further work is necessary.
4. Non-parametric tests are applicable to all types of data - qualitative (nominal scaling) data in rank form (Ordinal scaling) as well as data that have been measured more precisely (interval or ratio scaling).
5. Many non-parametric methods make it possible to work with very small samples. This is particularly helpful to the researcher collecting pilot study data on to the medical researcher working with a rare disease.
6. Non-parametric methods make fewer and less stringent assumptions (that are more easily met) than the classical procedures.

2.2.2 Chi-Square Test

Chi-square test is a statistical test used to test the significance of the difference between observed frequencies and the corresponding theoretical frequencies (expected frequencies) of a distribution, without any assumption about the nature of the distribution of the population. This is the most widely used non-parametric test. It was developed by Prof. Karl Pearson. It is depicted using the symbol χ^2.

Uses of Chi-Square Test (Applications of Chi- Square Test)

Chi-Square test is mainly used for the following purposes:

a. Used to test independence: As a test of independence, χ^2 test is used to test whether the attributes of a sample are associated or not.

SGOU

b. Used to test goodness of fit: As a test for goodness of fit, χ^2 test can be used to test how far the theoretical frequencies fit to the observed frequencies.

c. Used to test homogeneity: As a test of homogeneity, χ^2 test is used to test whether different samples are homogeneous as far as a particular attribute is concerned.

d. Used to test population variance: It used to test whether there is any significant difference between sample variance and population variance. Here, the test statistic value (Chi- square value) is obtained by using the following formulae $(\frac{ns^2}{\sigma^2})$.

Conditions for applying Chi-square Test

i. The total frequencies (N) must be at least 50.

ii. Expected frequencies of less than 5 must be pooled with the preceding or succeeding frequency so that the expected frequency is 5 or more.

iii. The distributions should be of original units.

iv. They should not be of proportions or percentages.

v. All the members or items in the sample must be independent

2.2.3 Testing of Independence

Procedure

i. Set up H_0 and H_1

H_0: There is independence between observed frequencies and expected frequencies.

H_1: There is no independence between observed and expected frequencies.

ii. Decide the test statistic

Here, the test statistic is Chi-Square test.

iii. Apply the appropriate formula

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where,

O_i = Observed frequencies

E_i = Expected frequencies

Here E_i values are obtained by using the following formula:

$$E_i \text{ Value} = \frac{\text{(Row Total x Column Total)}}{\text{Grand Total}}$$

SGOU

E_i Values are computed by preparing a table called Contingency Table.

Contingency Table

A contingency table is a frequency table in which a sample from the population is classified according to two or more attributes. When there are only two attributes, the contingency table is known as a 2×2 **contingency table.**

Eg: Drinking/Smoking

	Smoking	**Non-Smoking**	**Total**
Drinking	27 (a)	18 (b)	(a+b)
Non-drinking	12 (c)	1 (d)	(c+d)
Total	(a+c)	(b+d)	N

The frequency appearing in the table is known as a cell frequency for a 2×2 contingency table.

$$\chi^2 = \frac{(ad-bc)^2 N}{(a+b)(c+d)(a+c)(b+d)}$$

where a, b, c and d are the observed **cell frequencies.**

iv. Specify the level of Significance

If nothing is mentioned, take 5% level of Significance.

v. Fix the degree of freedom

Degree of freedom = (r– 1) x (c – 1)

Where r = number of rows; c = number of columns.

vi. Obtain the table value of Chi-square at specified level of significance and fixed degree of freedom

vii. Compare the actual Value of Chi-Square with the table value

Compare the actual Value of Chi-Square with the table value and decide whether to accept or reject the null hypothesis. If the calculated value is less than the table value, the null hypothesis is accepted and otherwise rejected.

Illustration 2.2.1

From the following data, can you say there is a relation between the smoking habit and literacy?

SGOU

	Smokers	Non- Smokers
Literates	83	57
Illiterates	45	68

Solution

H_0: There is no association between smoking habit and literacy.

H_1: There is a positive association between smoking habit and literacy.

Here the Observed values (Actual values) are 83, 57, 45 and 68.

2 x 2 Contingency Table			
	Smokers	**Non- smokers**	**Total**
Literates	$\frac{[(83+57) \times (83+45)]}{253} = 71$	$\frac{(140 \times 125)}{253} = 69$	140
Illiterates	$\frac{(113 \times 128)}{253} = 57$	$\frac{(113 \times 125)}{253} = 56$	113
Total	128	125	253

Observed Values(O_i)	**Expected Values(E_i)**	$\frac{(O_i - E_i)^2}{E_i}$
83	71	2.03
57	69	2.09
45	57	2.53
68	56	2.57
		$\chi^2 = 9.22$

The E_i Values corresponding to the above 'O_i' values can be

found out by preparing a 2 x 2 contingency table: So, the E values are 71, 69, 57 and 56.

Calculated Value of $\chi^2 = 9.22$ Level of significance = 5%

Degree of Freedom = (2 – 1) x (2 – 1) = 1 x 1 = 1

Table value of χ^2 at 5% level of Significance and 1 d.f is 3.841.

Since calculated value is more than the table value, null hypothesis is rejected. We accept alternate hypothesis. So we may conclude that there is no independence between smoking habit and literacy. In other words, smoking habit and literacy are related.

SGOU

Illustration 2.2.2

In a sample study about the tea drinking habit in a town, following data are observed in a sample of size 200. 46% were male, 26% were tea drinkers and 17% were male tea drinkers. Is there any association between gender and tea habits?

Solution

H_0: There is independence between gender and tea drinking habits.

H_1: There is no independence between gender and tea drinking habits.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Here, all the Observed values (Actual values) are not directly given in the question. So, we have to find the missing figures with the help of a 2 x 2 contingency table:

2 x 2 Contingency Table ('O_i' values)			
	Tea drinkers	Non- teadrinkers	Total
Male	$\frac{(200 \times 17)}{100} = 34$	58	$\frac{(200 \times 46)}{100} = 92$
Female	18	90	108
Total	$\frac{(200 \times 26)}{100} = 52$	148	200

"O_i" values are 34, 58, 18 and 90

The E_i Values corresponding to the above 'O_i' values can be found out by preparing a 2 x 2 contingency table:

2 x 2 Contingency Table ('E_i' values)			
	Tea drinkers	Non-tea drinkers	Total
Male	$\frac{(92 \times 52)}{200} = 24$	$\frac{(92 \times 148)}{200} = 68$	92
Female	$\frac{(108 \times 52)}{200} = 28$	$\frac{(108 \times 148)}{200} = 80$	108
Total	52	148	200

So, the 'E_i' values are 24, 68, 28 and 80.

SGOU

Computation of Chi-square Value			
ObservedValues (O_i)	**ExpectedValues (E_i)**	**(O_i-E_i) 2**	$\frac{(O_i - E_i)^2}{E_i}$
34	24	100	4.17
58	68	100	1.47
18	28	100	3.57
90	80	100	1.25
			$\chi^2 = 10.46$

Calculated Value of $\chi^2 = 10.46$ Level of significance = 5%

Degree of Freedom = (2 – 1) x (2 – 1) = 1 x 1 = 1

The table value of χ^2 at 5% significance level and 1 d.f. is 3.841.

Since calculated value is more than the table value, null hypothesis is rejected. We accept alternate hypothesis. So, we may conclude that there is no independence between gender and smoking habit. In other words, gender and smoking habit are closely associated.

2.2.4 Testing of Goodness of Fit

Procedure

i. Set up H_0 and H_1

H_0: There is goodness of fit between observed frequencies and expected frequencies.

H_1: There is no goodness of fit between observed frequencies and expected frequencies.

ii. Decide the test statistic

Here, the test statistic is Chi-Square test.

iii. Apply the appropriate formula

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where,

O_i = Observed frequencies

E_i = Expected frequencies

iv. Specify the level of Significance

If nothing is mentioned, take 5% level of significance.

SGOU

v. Fix the degree of freedom

Degree of freedom = n – r –1 Where

n = number of pairs of observations.

r = number of parameters computed from the given data to find the expected frequencies.

vi. Obtain the table value

Obtain the table value of Chi-square at specified level of significance and fixed degree of freedom.

vii. Compare the actual Value of Chi-Square with the table value

Compare the actual Value of Chi-Square with the table value and decide whether to accept or reject the null hypothesis. If calculated value is less than the table value, null hypothesis is accepted and otherwise it is rejected.

Illustration 2.2.3

The numbers of road accidents per week in a certain city were as follows:

12, 8, 20, 2, 14, 10, 15, 6, 9, 4

Are these frequencies in agreement with the belief that the accidents occurred were the same during the 10-week period?

Solution

H_0: There is goodness of fit between observed frequencies and expected frequencies.

H_1: There is no goodness of fit between observed frequencies and expected frequencies.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Here the Observed values (Actual values) are 12, 8, 20, 2, 14, 10, 15, 6, 9 and 4.

i.e., O_i = 12, 8, 20, 2, 14, 10, 15, 6, 9, 4

If accidents occurred are same, then the number of accidents per week which we may expect is 10 (i.e., the average of the given values).

$$E_i = \frac{100}{10} 10 = 10$$

Now we can find the Value of Chi-square as follows:

SGOU

Computation of Chi-square Value		
Observed Values (O_i)	**Expected Values (E_i)**	$\frac{(O_i - E_i)^2}{E_i}$
12	10	0.4
8	10	0.4
20	10	10.0
2	10	6.4
14	10	1.6
10	10	0.0
15	10	2.5
6	10	1.6
9	10	0.1
4	10	3.6
		$\chi^2 = 26.6$

Calculated Value of $\chi^2 = 26.6$ Level of significance = 5%

Degree of Freedom = n - 1 = 10 – 0 -1 = 9

Table value of χ^2 at 5% level of significance and d.f of 9 is 16.919.

Since calculated value is more than the table value, null hypothesis is rejected. We accept alternate hypothesis. So we may conclude that the given figures do not agree with the belief that accident occurred were same during the 10 weeks period.

Illustration 2.2.4

The principal of a college made a sample analysis of an examination result of 200 students. It was found that 24 students had got first class, 62 second class, 68 third class and the rest were failed. Are these figures commensurate with the general examination result which is in the ratio of 2:3:3:2 for various categories respectively?

Solution

H_0: There is goodness of fit between the given figures and the figures expected in general examination

H_1: There is no goodness of fit between the given figures and the figures expected in general examination

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Here the Observed values (Actual values) for first, second, third and failed categories of students are 24, 62, 68 and 46 respectively.

SGOU

If results are in the ratio of 2:3:3:2, then the number of students for above categories may be expected as follows:

First Class	$200 \times \frac{2}{10}$	40
Second Class	$200 \times \frac{3}{10}$	60
Third Class	$200 \times \frac{3}{10}$	60
Failed	$200 \times \frac{2}{10}$	40
Total		200

So, the E_i Values are 40, 60, 60 and 40.

Now we can find the Value of Chi-square as follows:

Computation of Chi-square Value			
Observed Values (O_i)	**Expected Values(E_i)**	$(Oi - Ei)^2$	$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$
24	40	256	6.400
62	60	4	0.067
68	60	64	1.067
46	40	36	0.900
			$\chi^2 = 8.434$

Calculated Value of $\chi^2 = 8.434$ Level of significance = 5%

Degree of Freedom = n -1 = 4 -1 = 3

Table value of χ^2 at 5% level of significance and d.f. of 3 is 7.815.

Since calculated value is more than the table value, null hypothesis is rejected. We accept alternate hypothesis. So, we may conclude that the given figures do not commensurate with the general examination result which is in the ratio of 2:3:3:2.

SGOU

2.2.5 Test of Homogeneity

Procedure

i. Set up H_0 and H_1

H_0: There is homogeneity between the samples on the basis of the attribute.

H_1: There is no homogeneity between the samples on the basis of the attribute.

ii. Decide the test statistic

Here, the test statistic is Chi-Square test.

iii. Apply the appropriate formula

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i}$$

where,

O_i = Observed frequencies

E_i = Expected frequencies

Here 'E_i' values are obtained by using the following formula:

$$\text{'}E_i\text{' Value} = \frac{\text{(Row Total x Column Total)}}{\text{Grand Total}}$$

'E_i' Values are computed by preparing a table called Contingency Table.

iv. Specify the level of Significance

If nothing is mentioned, take 5% level of significance.

v. Fix the degree of freedom

Degree of freedom= (r– 1) x (c – 1)

Where r = number of rows; c = number of columns

vi. Obtain the table value of Chi-square

Obtain the table value of Chi-square at specified level of significance and fixed degree of freedom.

vii. Compare the actual Value of Chi-Square with the table value

Compare the actual Value of Chi-Square with the table value and decide whether to accept or reject the null hypothesis. If the calculated value is less than the table value, the null hypothesis is accepted and otherwise rejected.

SGOU

Illustration 2.2.5

In a diet survey the following results were obtained:

	Male	Female
No. of families drinking tea	124	16
No. of families not drinking tea	56	10

Is there any difference between the gender in the matter of tea drinking?

Solution

H_0: There is homogeneity between gender in the matter of tea drinking.

H_1: There is no homogeneity between gender in the matter of tea drinking.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Here the Observed values (Actual values) are 124, 16, 56, and 10

The 'E_i' values corresponding to the above 'O_i' values can be found out by preparing a 2 x 2 contingency table:

2 x 2 Contingency Table			
	Male	**Female**	**Total**
No. of families drinking tea	$\frac{(140 \times 180)}{206} = 122$	$\frac{(140 \times 26)}{206} = 18$	140
No. of families not drinking	$\frac{(66 \times 180)}{206} = 58$	$\frac{(66 \times 26)}{206} = 8$	66
Total	**180**	**26**	**206**

So, the 'E' values are 122, 18, 58 and 8.

Observed Values (O_i)	**Expected Values (E_i)**	**$(O_i - E_i)^2$**	$\frac{(O_i - E_i)^2}{E_i}$
124	122	4	0.033
16	18	4	0.222
56	58	4	0.069
10	8	4	0.500
			$\chi^2 = 0.824$

SGOU

Calculated value of $\chi^2 = 0.824$

Level of significance =5%

Degree of Freedom = (2 – 1) x (2 – 1) = 1 x 1 = 1

Table value of χ^2 at 5% level of significance and 1 degree of freedom is 3.841.

Since calculated value is less than the table value, null hypothesis is accepted. So, we may conclude that there is homogeneity between gender in the matter of tea drinking.

2.2.6 Test of Variance Population

Procedure

i. Set up H_0 and H_1

H_0: There is no significant difference between sample variance and population variance.

H_1: There is significant difference between sample variance and population variance.

ii. Decide the test statistic

Here, the test applicable is Chi-square test.

iii. Apply the appropriate formula for computing the value of test statistic

$$\chi^2 = \frac{ns^2}{\sigma^2}$$

where,

n = sample size,

s^2 = sample variance,

σ^2 = population variance.

iv. Specify the level of significance

Take 5%, unless specified otherwise.

v. Fix the degree of freedom

d.f = n –1.

vi. Locate the table value of Chi-square at the specified level of significance and fixed degree of freedom

vii. Compare the actual Value of Chi-Square with the table value

Compare the actual Value of Chi-Square with the table value and decide whether to accept or reject the null hypothesis. The null hypothesis is accepted and otherwise rejected if the calculated value is less than the table value.

SGOU

Illustration 2.2.6

A sample is drawn from a population which follows a normal distribution. The size of the sample and S.D. are, respectively 10 and 5. Test whether this is consistent with the hypothesis that the S. D. of the population is 5.3.

Solution

H_0**:** There is no significant difference between sample S.D. and population S.D. (i.e., H_0:S.D of population = 5.3)

H_1**:** There is a significant difference between sample S.D. and population S.D. (i.e., H_1 : S.D. of population ≠ 5.3)

The test applicable is the Chi-square test.

$$\chi^2 = \frac{ns^2}{\sigma^2}$$

Where,

$n = 10, s^2 = 5^2, \sigma^2 = 5.3^2$

$$= \frac{10 \times 5^2}{5.3^2}$$

$$= \frac{250}{28.09}$$

$$= 8.899$$

Specify the level of significance. Take 5%, unless specified otherwise.

Fix the degree of freedom. d.f = 10 –1 = 9

Table value of Chi-square at 5% level of significance and 9 degree of freedom is 16.919.

The null hypothesis is accepted since the calculated value is less than the table value. So we may conclude that there is no significant difference between sample S.D. and population S.D.

Illustration 2.2.7

A sample group of 10 students are selected randomly from a class. Their weights (in K.g) are 49, 40, 53, 38, 52, 47, 48, 45, 55, and 43. Can we say that the population variance is 20 Kg?

Solution

H_0: There is no significant difference between sample variance and population variance.(i.e., H_0: Variance of population = 20)

SGOU

H_1: There is significant difference between sample variance and population variance. (i.e., H_1: Variance of population $\neq 20$)

The test applicable is Chi-square test.

$$\chi^2 = \frac{ns^2}{\sigma^2}$$

Here, n = 10, $\sigma^2 = 20$, Sample variance is to be computed from the given data.

Computation of sample variance (s^2)		
Weight (x)	**$(x - \bar{x})$**	**$(x - \bar{x})^2$**
49	2	4
40	-7	49
53	6	36
38	-9	81
52	5	25
47	0	0
48	1	1
45	-2	4
55	8	64
43	-4	16
x = 470		**$(x - \bar{x})^2 = 280$**

$$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{470}{10} = 47$$

$$\text{Sample variance } (S^2) = \frac{\sum (x - \bar{x})^2}{n}$$

$$= \frac{280}{10} = 28$$

$$\chi^2 = \frac{ns^2}{\sigma^2}$$

$$\frac{10 \text{ x } 28}{20} = \frac{280}{20} = 14$$

Level of significance =5%

Degree of freedom (d.f) = 10 –1 = 9

SGOU

Table value of Chi-square at 5% level of significance and 9 degrees of freedom is 16.919. The null hypothesis is accepted since the calculated value is less than the table value. So, we may conclude that there is no significant difference between sample and population variance.

∴The population variance = 20 Kg.

2.2.7 Sign Tests

The Sign Test is a non-parametric statistical test used to evaluate hypotheses about the median of a single sample or to compare two related samples when normality cannot be assumed. It considers only the direction (positive or negative sign) of differences rather than their magnitude and is based on the binomial distribution under the null hypothesis.

There are two types of sign tests. They are

a. One sample sign test and

b. Two sample sign test.

a. One sample sign test

A number of methods have been developed for judging the randomness of a sample on the basis of the order in which the observations are taken. It is possible to determine whether the samples that look suspiciously non-random may be attributed to chance. The technique discussed below is based on the theory of runs. A run is a succession of identical letters which is followed or preceded by different letters or no letters at all.

The number of runs or 'r' is a statistic with its own special sampling distribution and its own test.

$$\mu = \frac{2\,n_1\,n_2}{n_1 + n_2} + 1$$

where μ is the mean of the 'r' statistic.

The standard Error (SE) of the 'r' statistic is calculated by the formula

$$\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

It may be noted that the sampling distribution of 'r' can be closely approximated by the normal distribution if either n_1 or n_2 is larger than 20.

$$Z = \frac{r - \mu}{\sigma}$$

It is also called single sample sign test and is used to test the hypothesis concerning the median for one population. One sample sign test is used to test whether the sample belongs to a particular population.

SGOU

b. Two sample sign test (Paired sample sign test)

The Two-Sample Sign Test is a non-parametric test used to compare two related or paired samples to determine whether there is a significant difference between their medians when the assumption of normality is not satisfied. It is commonly applied to before–after studies or matched pairs data, where the differences between paired observations are calculated and only the direction of each difference is considered. For each pair, a positive sign (+) is assigned if the first observation exceeds the second, and a negative sign (−) if it is smaller, while zero differences are discarded.

Let n be the number of non-zero differences and x be the number of positive signs; under the null hypothesis that the median difference is zero, the number of positive signs follows a binomial distribution with parameters n and $p = 0.5$. Thus, the probability of obtaining exactly x positive signs is given by $P(X = x) = \binom{n}{x}(0.5)^n$.

For large samples (n > 25), the normal approximation may be used with the test statistic $Z = \frac{x - n/2}{\sqrt{n/4}}$, applying a continuity correction if necessary. The null hypothesis is rejected if the calculated probability or Z value indicates a significant deviation from the expected equal number of positive and negative signs.

2.2.8 Signed Rank Test (Wilcoxon Matched Pairs Test)

In the case of two related samples (when we can determine both direction and magnitude of difference between matched values) to find the significance of difference, we can use the non-parametric test known as Signed Rank Test.

When this technique is employed we first find the difference between each pair of values and assign ranks to the differences from the smallest to the largest without regard to sign. The actual sign of each difference are then put to corresponding ranks and test statistic 'T' is calculated. 'T' is the smaller of the two, namely the sum of the negative ranks and the sum of the positive ranks. When the given number of matched pair is less than or equal to 25.

When 'T' is greater than or equal to the table value, we accept that there is difference. While applying this test we may have situations where some matched pairs are equal i.e., difference between values is zero. In this case we drop out these pairs. Sometimes two or more pairs may have same differences. In such cases, we assign the average of ranks to each of these pairs.

2.2.9 Rank Sum Tests

Rank sum tests are another type of tests used for testing whether the populations are identical. Here, various samples are taken together, and ranks are assigned. There are two important types of rank sum tests.

SGOU

They are

a. Wilcoxon-Mann- Whitney Test (U- test), and

b. Kruskal – Wallis Test (H- test).

2.2.9.1 Mann-Whitney U Test

Rank sum test is a whole family of test. In this test the null hypothesis $H0 = \mu_1 = \mu_2$ without assuming whether the population sampled have roughly the shape of normal distribution. This test help us to determine whether the two samples have come from identical population.

The test of the null hypothesis that the two samples come from identical population may either be based on R_1, the sum of the ranks of the values of the first sample or on R_2,the sum of the ranks of the values of the second sample.

If the sample sizes are n_1 and n_2 , the sum of R_1 and R_2 , is simply the sum of the first $n_1 + n_2$ positive integers which is known to be:

$$\frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2}$$

This formula enable us to find R_2 if R_1 is known and vice versa.

When the use of the rank sums was first proposed as a non-parametric alternative to the two sample t-test, the decision was based on R_1 and R_2, but now the decision is usually based on either of the related statistics.

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

where n_1 and n_2 are the size of samples and R_1 and R_2 are the rank sums of the corresponding samples.

$$Z = \frac{U - \frac{n_1 \times n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

2.2.9.2 Kruskal – Wallis Test (H- test)

Kruskal-Wallis one way analysis of variance or the H-test helps in testing the null hypothesis that k independent random samples come from identical population against the alternative hypothesis that the means of these samples are not all equal.

SGOU

In the Mann-Whitney U-test all data are ranked as if they were in one sample, from lowest to highest, the rank sums of each sample are calculated. The H-statistic is calculated from the formula:

$$H = \frac{12}{N(N+1)}\left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \ldots + \frac{R_k^2}{n_k}\right] - 3(N+1)$$

where $n_1, n_2, \ldots n_k$ are the number in each k samples, $N = n_1 + n_2 + \ldots + n_k$ and $R_1, R_2, \ldots R_k$ are the rank sums of each sample. If there are ties, the usual procedure is followed, but H is fairly sensitive to ties, so if there are very many of them a correction should be made. The effect of correction is to increase slightly the value of H, so its use is not imperative. For small samples H is approximately distributed as Chi-square with k-1 degrees of freedom.

$H = \frac{12}{n(n+1)} \cdot \frac{R_i^2}{n_i} - 3(n+1)$ where R_i is sum of the ranks assigned to n_i observations in the i^{th} sample.

2.2.10 Wald-Wolfowitz' Runs Test

The Wald–Wolfowitz Runs Test, developed by Abraham Wald and Jacob Wolfowitz, is a non-parametric statistical test used to determine whether a sequence of observations is random. A *run* is defined as a sequence of identical elements (such as +/−, success/failure, male/female, above/below median) that is preceded and followed by a different element or no element at all. The test examines whether the observed number of runs in a sequence differs significantly from the number expected under randomness.

Let n_1 and n_2 be the number of observations in the two categories, and $n = n_1 + n_2$, with R representing the observed number of runs.

The expected number of runs is given by $E(R) = \frac{2n_1n_2}{n_1+n_2} + 1$, and

the variance is $Var(R) = \frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$.

For large samples, the test statistic is calculated using the normal approximation:

$Z = \frac{R-E(R)}{\sqrt{Var(R)}}$, with a possible continuity correction of ±0.5.

The null hypothesis states that the sequence is random, and it is rejected if the calculated Z value exceeds the critical value at the chosen significance level. The runs test is widely applied in quality control, finance, psychology, and business research because it does not require assumptions of normality and is suitable for ordinal or categorical data.

SGOU

Summarised Overview

The study of non-parametric tests provides a foundational understanding of statistical techniques that do not rely on assumptions of normality or equal variances. These methods are particularly useful when dealing with ordinal data, small sample sizes, or categorical variables, making them indispensable in fields such as business management, social sciences, and behavioural research. Unlike parametric tests, which require interval-level data and specific distributional assumptions, non-parametric tests offer flexibility and robustness in real-world applications where such conditions are rarely met.

The Chi-Square Test is a cornerstone of categorical data analysis. It includes two major applications: the Test of Goodness of Fit, which assesses whether observed frequencies match expected distributions, and the Test of Independence, which evaluates the relationship between two categorical variables using contingency tables. These tests are widely used in market research, consumer behaviour studies, and HR analytics to examine patterns and associations in survey responses, demographic data, and organisational metrics.

It also covers a suite of rank-based and distribution-free tests designed for comparing medians and assessing differences across groups. In both one-sample and two-sample formats, the Sign Test is used to test hypotheses about medians without relying on data symmetry. The Wilcoxon Signed Rank Test extends this approach by incorporating the magnitude of differences in paired samples, offering a more sensitive alternative to the Sign Test. For independent samples, the Mann–Whitney U Test serves as a non-parametric counterpart to the independent t-test, enabling comparisons of central tendency between two groups.

To analyse differences across multiple independent groups, the Kruskal–Wallis H Test is used, and the Wald–Wolfowitz Runs Test is introduced to assess the randomness of sequences, which is particularly relevant in quality control, financial forecasting, and behavioural pattern analysis. These methods equip learners with a comprehensive toolkit for conducting rigorous statistical analysis in diverse business and policy contexts.

SGOU

Self-Assessment Question

1. What do you mean by non-parametric tests?
2. What are the situations under which non-parametric tests are applied?

3. What are the important assumptions of non-parametric tests?
4. What are the important merits of non-parametric tests?
5. What are the important drawbacks of non-parametric tests?
6. What are the different types of non-parametric tests?
7. Distinguish between parametric tests and non-parametric tests.
8. Briefly explain sample sign test?
9. Discuss briefly Wilcoxon matched pair test and Wilcoxon Mann Whitney U-test?
10. Describe Kruskal-Wallis H-test?
11. What do you mean by Chi-square value?
12. What are the important uses of Chi-square test?
13. What do you mean by goodness of fit?
14. Explain the procedure for testing goodness of fit.
15. What do you mean by contingency table?



Assignments

1. A sample of size 50 has S.D. of 10.5. Can you contradict the hypothesis that the population S.D. is 12 ? **(Answer- $\chi^2 = 7.65$, Table value 9.488)**
2. A poker-dealing machine is supposed to deal cards at random, as if from an infinite deck. In a test, you counted 1600 cards, and observed the following:

Spades	404
Hearts	420
Diamonds	400
Clubs	376

Could it be that the suits are equally likely? Or are these discrepancies too much to be random? **(Answer- $\chi^2 = 2.480$)**

3. A die is thrown 120 times and the frequencies of various faces are as follows:

SGOU

Face No:	1	2	3	4	5	6
Frequency:	10	15	25	25	18	27

Test whether the die was fair. **(Answer: $\chi^2 = 11.40$)**

4. Explain the procedure to compute Kruskal Wallis test and Mann-Whitney U test. Compare and contrast the two tests.

5. In in an experiment with the immunisation of cattle from tuberculosis following results were obtained

	Affected	Un affected
Innoculated	12	26
Non innoculated	16	6

Explain the effect of the vaccine in controlling susceptibility to tuberculosis. **(Ans:- $\chi^2 = 9.48$, Good Effect)**

6. In a course of anti-malarial work in a certain city over a period of time quinine was administered to 606 adults out of a total population of 3540. The data regarding incidence of malarial fever is given below Examine whether the Quinine has the effect of preventing fever

	Fever	No fever	Total
Quinine	19	587	606
No quinine	193	1741	2934
	212	3328	3540

(Ans:- $\chi^2 = 18.16$, Preventive)

SGOU

Reference

1. Aggarwal, B. M. (2010). *Business statistics*. Ane Books Pvt. Ltd.
2. Chandrasekharan, N., & Umaparvathi, M. (2016). *Statistics for management*. PHI Learning Pvt. Ltd.
3. Gupta, M., & Khanna, R. (2011). *Quantitative techniques for decision making*. Prentice Hall of India.

4. Gupta, S. P. (2021). *Statistical methods*. S. Chand & Sons.

5. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International (P) Ltd.

6. Levin, R. I., & Rubin, D. S. (1999). *Quantitative approaches for management*. Pearson Education.

7. Sharma, A. (2017). *Quantitative techniques for decision making*. Himalaya Publishing House.

8. Vohra, D. V. D. (2017). *Quantitative techniques for management*. New Delhi: McGraw Hill Education

Suggested Reading

1. Arora, P. N., Arora, S., & Arora, S. (2010). *Comprehensive statistical methods*. S. Chand & Company Pvt. Ltd.

2. Beri, G. C. (2017). *Business statistics*. Tata McGraw-Hill.

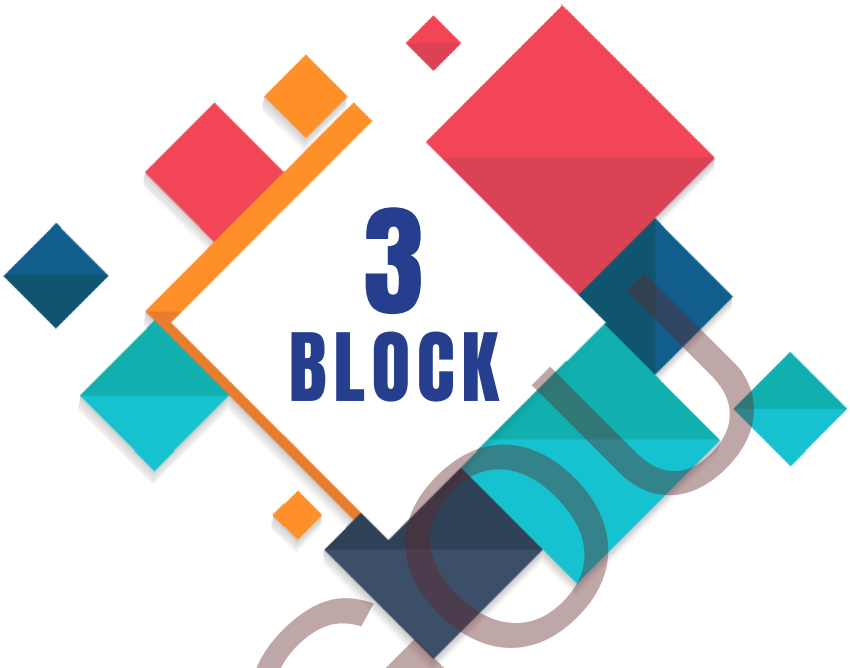
3. Chawla, D., & Sondhi, N. (2016). *Research methodology: Concepts and cases*. Vikas Publishing House.

4. Krishnaswami, O. R., Ranganathan, M., & Harikumar, P. N. (2016). *Research methodology*. Himalaya Publishing House.

5. Render, B. (2022). *Quantitative analysis for management*. Prentice Hall of India.

6. Singh, A. K. (2017). *Tests, measurements and research methods in behavioural sciences*. Bharti Bhawan.

SGOU



Correlation and Regression Analysis

Unit 1

Correlation

Learning Outcomes

On completion of this unit, learners will be able to:

- ◊ explain the concept of correlation, including its meaning, significance in statistical analysis, and relevance to business and social research
- ◊ distinguish between different types of correlation (positive, negative, linear, non-linear, and zero correlation), and interpret their implications in real-world datasets
- ◊ introduce various methods for correlation, including scatter diagrams, Karl Pearson's coefficient, Spearman's rank correlation, and the concurrent deviation method
- ◊ calculate Karl Pearson's and Spearman's correlation coefficients, using appropriate formulas and datasets, and assess the strength and direction of associations
- ◊ evaluate the reliability of correlation results through the concept of probable error, and determine the statistical significance of observed correlations in empirical studies

SGOU

Background

Correlation emerged in the late 19th century to quantify the relationship between variables. Its intellectual roots trace back to Sir Francis Galton, who introduced the idea of "correlation" while studying heredity and anthropometry. Galton's use of scatter diagrams and regression lines laid the foundation for the study of statistical association. However, French physicist Auguste Bravais had already hinted at the mathematical groundwork in 1844 when the product-moment formula for correlation was derived. Galton's work inspired Karl Pearson, who formalised the Pearson product-moment correlation coefficient in 1896,

transforming correlation into a rigorous statistical tool. Pearson's contributions extended to regression analysis, the chi-squared test, and the development of the first university statistics department at University College London in 1911.

Parallel to Pearson's work, Charles Spearman developed the rank correlation coefficient in 1904 to address relationships between ordinal variables, especially in psychological testing. Spearman's method was particularly valuable for non-parametric data and remains widely used in behavioural and social sciences. Other statisticians such as George Udny Yule and Francis Edgeworth contributed to refining correlation techniques, including partial and spurious correlation concepts. Over time, correlation analysis became a cornerstone of empirical research across disciplines, enabling scholars to explore associations, test hypotheses, and interpret complex data relationships more precisely. This unit explains the meaning, significance, types, and methods of studying correlation, including scatter diagrams, Karl Pearson's and Spearman's coefficients, the concurrent deviation method, and the concept of probable error.



Keywords

Correlation, Scatter Diagram, Karl Pearson's Coefficient, Spearman's Rank Correlation, Concurrent Deviation Method, Probable Error



Discussion

Correlation is a fundamental statistical tool used to measure and analyse the degree of association between two or more variables. It helps researchers understand whether and how variables change together, whether positively, negatively, or not at all. Understanding correlation helps decision-makers identify patterns and dependencies that influence performance in business management. For instance, a positive correlation between employee engagement scores and customer satisfaction ratings may suggest that motivated staff contribute to better service delivery. Conversely, a negative correlation between product price and sales volume might indicate that higher prices reduce demand. By quantifying these relationships, typically using a correlation coefficient ranging from -1 to +1, managers can make informed choices about resource allocation, strategy adjustments, and performance improvement initiatives.

3.1.1 Meaning of Correlation

The term correlation refers to the relationship between two variables. When two variables move together, we say that they are correlated. When two variables are

SGOU

correlated, the change in one variable results in a corresponding change in the other variable. When the values of two variables move in the same direction, Correlation is said to be direct, ie, an increase in the value of one variable results in an increase in the value of the other variable. Also or a decrease in the value of one variable results in the decrease of the other variable. The Correlation is said to be +ve.

If on the other hand the values of the two variables move in the opposite direction so that an increase in the value of variable results in a decrease in the value of other variable or the decrease in value of one variable results in the increase in value of other variable, the correlation is said to be -ve or inverse. Generally price and supply are positively co-related because when price increases supply also increases and vice versa Correlation between price and demand is said to be -ve because with a fall in price demand goes up and vice versa Eg: when the cost of commodities increase expense also increases. It is +ve correlation. When expense increases savings decreases- It is -ve correlation.

Correlation analysis focuses on examining the strength and direction of relationships between two or more variables. It explores how variables tend to move together, either in the same direction (positive correlation) or opposite directions (negative correlation).

Examples include:

1. A price rise often leads to a fall in demand.
2. Higher prices may encourage increased supply.
3. As household income grows, so does expenditure.
4. Sales of woollen clothing rise with the onset of winter.
5. More students in college lead to higher canteen sales; during holidays, sales drop.
6. Increased smoking may be linked to higher lung cancer cases.

These examples illustrate how variables can be interrelated. However, not all variables show such patterns; some may vary independently, with no consistent relationship.

Correlation analysis involves comparing two data series to assess how they vary together. It requires a sufficiently large number of paired observations to draw reliable conclusions. Each value in one series must correspond to a value in the other; incomplete pairing undermines the analysis.

3.1.2 Definition of Correlation

Correlation has been defined in different ways. Some of the main definitions are as follows:

1. According to ***Croxton and Cowden***, *"When the relationship is quantitative, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula is known as correlation."*

SGOU

2. Boddington says, *"Whenever some definite connection exists between the two or more groups, classes or series or data, there is said to be correlation."*

3. According to ***L.R. Connor***, *"If two or more quantities vary in sympathy, so that movements in the one tend to be accompanied by corresponding movements in the other(s), then they are said to be correlated."*

3.1.3 Significance of Correlation

◊ The correlation coefficient provides a single numerical value quantifying the degree of association between two variables.

◊ Identifying relationships among variables allows for future predictions. For instance, if wheat production rises, assuming other conditions remain unchanged, a decline in its market price may be anticipated.

◊ When two variables exhibit a strong relationship, the value of one can be estimated based on the known value of the other. This estimation is carried out using regression equations.

◊ Correlation analysis supports strategic decision-making in business settings. Forecasts regarding the behaviour of key variables often rely on insights derived from correlation.

3.1.4 Types of Correlation

Generally, correlation is classified into the following types.

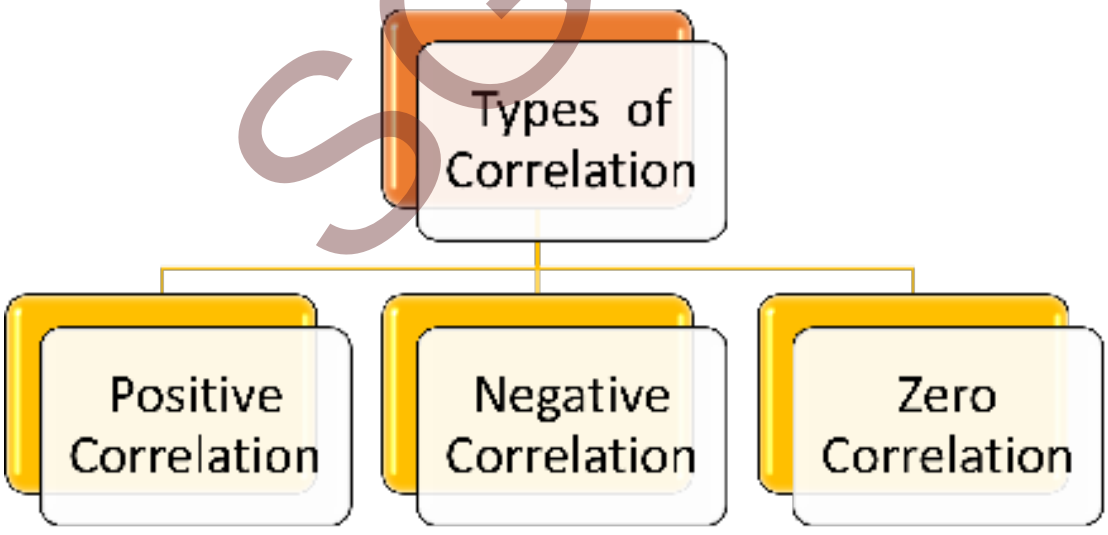


Fig. 3.1.1 Types of Correlation

3.1.4.1 Positive Correlation

Consider the variables income and expenditure. An increase in income is typically accompanied by a rise in expenditure, and a decrease in income leads to reduced

SGOU

spending. When two variables move in the same direction, increasing or decreasing, their relationship is called positive or direct correlation.

Examples of positive correlation

1. The more time spent running on a treadmill, the more calories burned.
2. An increase in the money supply often leads to a rise in commodity prices.
3. Taller individuals generally wear larger shoe sizes, while shorter individuals tend to have smaller ones.

Negative Correlation

In contrast, when one variable increases while the other decreases, or vice versa, the relationship is termed negative or inverse correlation. This indicates that the variables move in opposite directions.

Examples of negative correlation

1. As a train's speed increases, the time to reach its destination decreases.
2. If a car slows down, the travel time to its destination increases.
3. With greater experience, the number of typing errors tends to decline.

Zero Correlation

In some cases, the variables have no correlation or there is no correlation between the variables. For eg, the height of the student and the marks he got in the examination have no correlation.

A positive perfect correlation is +1, and a negative perfect correlation is -1. Correlation lies between 1 and +1.

3.1.5 Methods of Studying Correlation

There are various techniques available to determine the relationship between two variables. The choice of method depends on factors such as the type of data, the purpose of the analysis, and the specific research question being addressed. Researchers and analysts select the most appropriate approach based on these considerations. Below are some of the commonly used methods for measuring correlation.

i. Scatter Diagram

ii. Karl Pearson's Coefficient of Correlation

iii. Spearman's Rank Correlation Coefficient

iv. Concurrent Deviation Method

SGOU

3.1.5.1 Scatter Diagram

A scatter plot is one of the simplest and most effective ways to visualise the relationship between two continuous variables. It is the diagrammatic representation of the relationship between two variables. One variable is represented along the X axis, and the second variable along the Y axis. Each data point is plotted on the graph with one variable on the X axis and the other on the Y axis. We put a dot in the plane for each pair of observations of two variables. The pattern of points can provide a visual indication of the strength and direction of the correlation. If the points tend to form a line sloping upward from left to right, it suggests a positive correlation. If the points form a line sloping downward, it suggests a negative correlation. If the points are scattered randomly, there may be no significant correlation. (See the figures below).

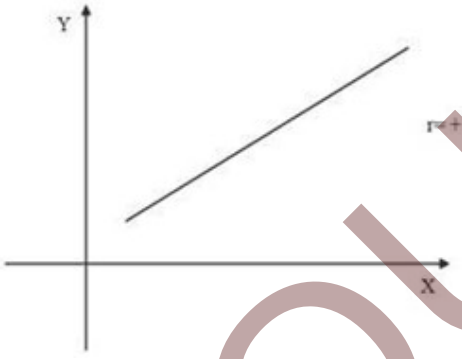


Fig. 3.1.2 Perfect positive correlation

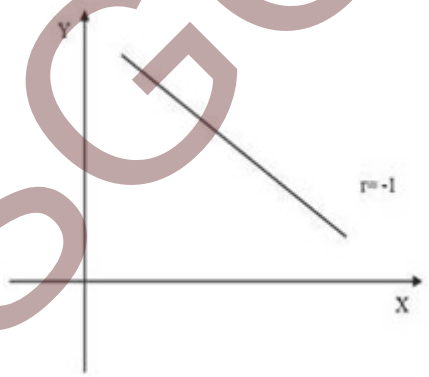


Fig. 3.1.3 Perfect negative correlation

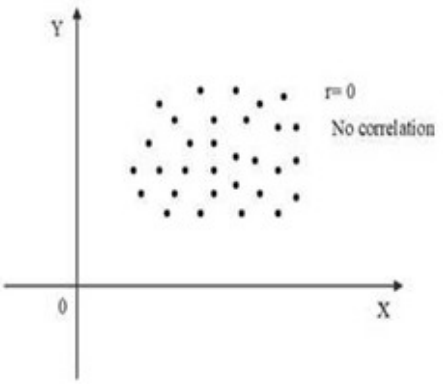


Fig. 3.1.4 No correlation

SGOU

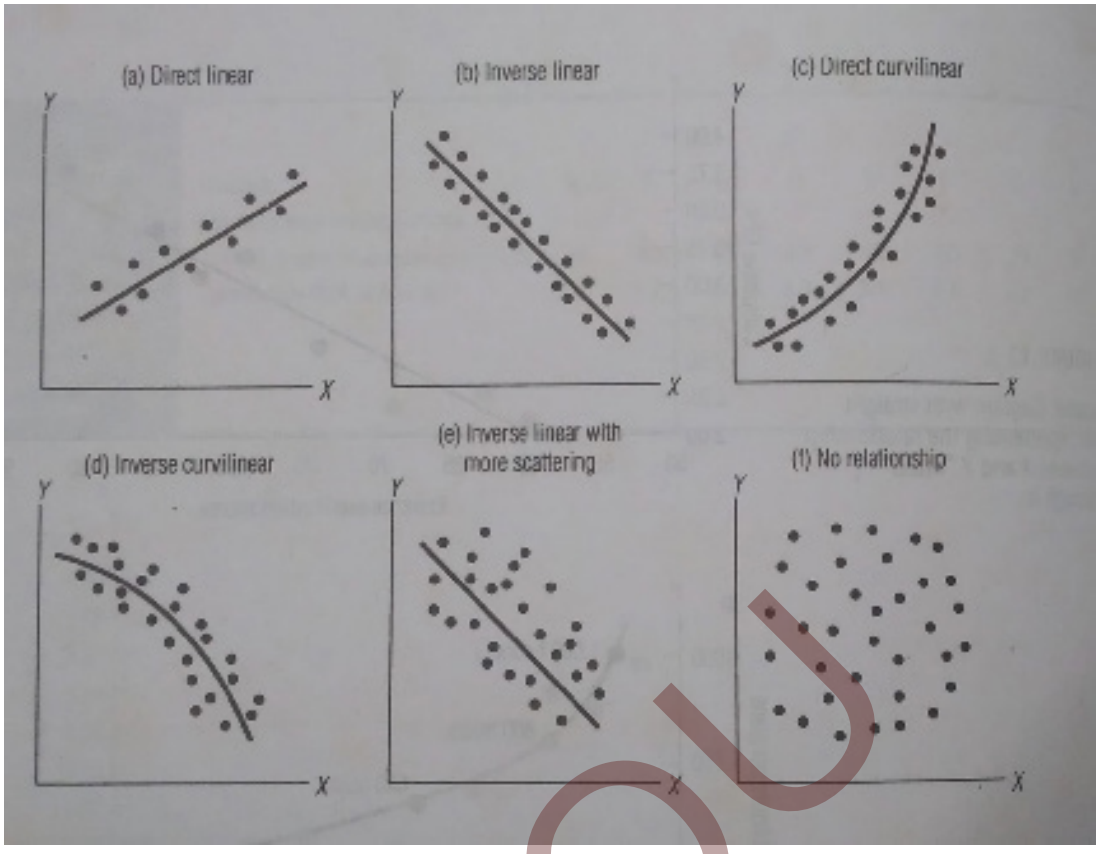


Fig. 3.1.5 Scatter Diagram

a. **Direct linear:** This graph shows a strong positive linear relationship, indicating that as the value of X increases, the value of Y increases consistently in a straight-line pattern.

b. **Inverse linear:** This graph shows a strong negative linear relationship, indicating that as the value of X increases, the value of Y decreases consistently in a straight-line pattern.

c. **Direct curvilinear:** This graph shows a positive non-linear relationship where both variables generally increase together, but the data points follow a curved line rather than a straight one.

d. **Inverse curvilinear:** This graph shows a negative non-linear relationship where as X increases, Y decreases, with the data points following a distinct curved path.

e. **Inverse linear with more scattering:** This graph shows a moderate negative linear relationship, similar to (b), but the data points are more widely scattered around the line of best fit, indicating less predictability.

f. **No relationship:** This graph shows no correlation between X and Y, as the data points are randomly dispersed across the plot with no discernible trend or pattern.

SGOU

3.1.5.2 Karl Pearson's Coefficient of Correlation (Product Moment Correlation)

Of the several methods of measuring Correlation, Karl Pearson's method is the most widely used in practice. The degree to which the variables in a bivariate distribution are related to each other is said to be Coefficient of correlation.

Karl Pearson's Coefficient of Correlation between two variables x and y

$$r = \frac{n\sum xy - (\sum x \,.\, \sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\ \sqrt{n\sum y^2 - (\sum y)^2}}$$

Illustration 3.1.1

Find Karl Pearson's correlation coefficient between x and y for the following data.

x	10	12	15	23	20
y	14	17	23	25	21

Solution

Karl Pearson's Coefficient of Correlation

$$r = \frac{n\sum xy - (\sum x \,.\, \sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\ \sqrt{n\sum y^2 - (\sum y)^2}}$$

x	y	xy	x^2	y^2
10	14	140	100	196
12	17	204	144	289
15	23	345	225	529
23	25	575	529	625
20	21	420	400	441
80	100	1684	1398	2080

$$r = \frac{5 \times 1684 - (80 \times 100)}{\sqrt{5 \times 1398 - (80)^2}\ \sqrt{5 \times 2080 - (100)^2}}$$

$$= 0.86$$

Changes of origin do not affect the Correlation Coefficient. We can replace x and y with d_x and d_y, respectively.

$$r = \frac{n\sum d_x d_y - (\sum d_x \,.\, \sum d_y)}{\sqrt{n\sum {d_x}^2 - (\sum d_x)^2}\ \sqrt{n\sum {d_y}^2 - (\sum d_y)^2}}$$

SGOU

The formula for computing Pearson's Coefficient of Correlation is

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y}$$

Illustration 3.1.2

Find Karl Pearson's correlation coefficient between x and y for the following data,

$$n = 1000, \sigma x = 4.5, \sigma y = 3.6, \Sigma (x - \bar{x}) \times (y - \bar{y}) = 4800$$

Solution

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y}$$

$$= (4800)/(1000 \times 4.5 \times 3.6)$$

$$= (4800)/16200$$

$$= 0.296$$

A Pearson correlation coefficient of 0.296 indicates a weak positive linear relationship between the two variables. This suggests that as one variable increases, the other variable also tends to increase, but not very strongly. The value signifies a mild and less direct association between the variables.

Illustration 3.1.3

Given : $\sum x = 125$, $\sum y = 100$, $\sum x^2 = 650$, $\sum y^2 = 436$, $\sum xy = 520$ *and* $n = 25$, obtain the value of Karl Pearson's correlation coefficient $r(X, Y)$.

Solution

$$r = \frac{n\sum xy - (\sum x \,.\, \sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\ \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$r = \frac{25 \times 520 - (125 \times 100)}{\sqrt{25 \times 650 - (125)^2}\ \sqrt{25 \times 436 - (100)^2}}$$

$$= 0.67$$

Illustration 3.1.4

Calculate the coefficient of correlation for the following table by Karl Pearson's coefficient of correlation.

SGOU

x	1	2	3	4	5
y	3	1	2	5	4

Solution

$$\bar{x} = \Sigma x/n = 15/5 = 3$$

$$\bar{y} = \Sigma y/n = 15/5 = 3$$

x	y	d_x	d_y	$d_x \times d_y$	d_x^2	d_y^2
1	3	-2	0	0	4	0
2	1	-1	-2	2	1	4
3	2	0	-1	0	0	1
4	5	1	2	2	1	4
5	4	2	1	2	4	1
Total =15	15	0	0	6	10	10

i.e., Karl Pearson's coefficient of correlation formula is

$$r = \frac{n\sum d_x d_y - (\sum d_x . \sum d_y)}{\sqrt{n\sum {d_x}^2 - (\sum d_x)^2}\sqrt{n\sum {d_y}^2 - (\sum d_y)^2}}$$

$$= \frac{6}{(\sqrt{10} \times \sqrt{10})}$$

$$= 6/10$$

$$= 0.6$$

Illustration 3.1.5

Calculate the coefficient of correlation for the following table using Karl Pearson's coefficient of correlation.

X	70	69	68	67	66	65	64
Y	72	68	70	68	65	67	66

Solution

$$\bar{x} = \Sigma x/n = 469/7 = 67$$

$$\bar{y} = \Sigma y/n = 476/7 = 68$$

SGOU

x	y	d_x	d_y	d_x^2	d_y^2	$d_x \times d_y$
70	72	3	4	9	16	12
69	68	2	0	4	0	0
68	70	1	2	1	4	2
67	68	0	0	0	0	0
66	65	-1	-3	9	9	3
65	67	-2	-1	1	1	2
64	66	-3	-2	4	4	6
469	476	0	0	28	34	25

Karl Pearson's coefficient of correlation formula is

$$r = \frac{n\sum d_x d_y - (\sum d_x \cdot \sum d_y)}{\sqrt{n\sum {d_x}^2 - (\sum d_x)^2}\ \sqrt{n\sum {d_y}^2 - (\sum d_y)^2}}$$

$$= \frac{25}{(\sqrt{28} \times \sqrt{34})}$$

$$= 25/30.8547$$

$$= 0.81$$

3.1.5.3 Probable Error

In correlation analysis, the probable error refers to estimating the possible deviation between the sample correlation and the actual correlation in the population. It helps to determine the accuracy and reliability of the correlation coefficient calculated from a sample, indicating the likelihood of the sample correlation differing from the actual correlation within the entire population.

Generally, Probable error is used to measure the validity of the value of the correlation coefficient. The correlation coefficient may vary for different samples drawn from the same population. However, the numerical value of such variations is expected to be less than the probable error.

The actual correlation coefficient lies between the limits (Coefficient of correlation ± Probable error)

$$Probable\ Error\ = 0.6745 \times \frac{(1-r^2)}{\sqrt{n}}$$

where r is the correlation coefficient and n' number of observation pairs.

SGOU

In correlation analysis, the standard error measures the accuracy of the sample correlation coefficient by estimating the variability or dispersion of the calculated correlation from the accurate correlation in the population. It measures how the sample correlation may vary from the actual correlation if the study were repeated multiple times. The standard error (SE) indicates the precision of the sample correlation coefficient and is calculated using the formula: $SE = \frac{(1-r^2)}{\sqrt{n}}$

Illustration 3.1.6

Find Probable error and Standard error if r = 0.8 and n= 100

$$\text{Probable Error} = 0.6745 \times \frac{(1-r^2)}{\sqrt{n}}$$

$$= 0.6745 \times \frac{(1-0.64)}{\sqrt{100}}$$

$$= 0.024$$

$$\text{Standard Error} = \frac{(1-r^2)}{\sqrt{n}}$$

$$SE = \frac{(1-0.64)}{\sqrt{100}}$$

$$= 0.36/10 = 0.036$$

Illustration 3.1.7

The correlation coefficient between two variables for 20 items is 0.96. Find the probable error also determine the limits for population r.

Solution

Given $r = 0.96, \ n = 20$

$$Probable\ Error = 0.6745 \times \frac{(1-r^2)}{\sqrt{n}}$$

$$= 0.6745 \times \frac{(1-0.96^2)}{\sqrt{20}}$$

$$= 0.6745 \times 0.0175$$

$$= 0.0118$$

SGOU

limits for population $r = r \pm P.E.(r)$

$$0.96 \pm 0.0118$$

$$(0.96 - 0.0118,\quad 0.96 + 0.0118)$$

$$(0.9482,\quad 0.9718)$$

3.1.5.4 Spearman's Rank Correlation Coefficient

The Spearman Rank Correlation Coefficient, often denoted as ρ (rho), is a statistical measure used to assess the strength and direction of the monotonic relationship between two variables. Unlike Pearson's correlation coefficient, which measures linear relationships, Spearman's rank correlation works with ranked or ordinal data and is based on the ranks of the observations rather than their actual values.

Spearman's Rank Correlation coefficient is valuable for non-normally distributed and ordinal data. It detects non-linear relationships, making it versatile for various datasets, especially with outliers. It is crucial in psychology and social sciences where variables can be ranked but not precisely measured.

i. Calculating Spearman's Rank Correlation Coefficient

Let us use a simple array of data to illustrate the use of Spearman's rank correlation:

Suppose you have data representing the amount of time (in hours) spent commuting to work (X) and the corresponding levels of job satisfaction (Y) for a group of employees:

Commute Time (X)	Job Satisfaction (Y)
1	4
2	3
3	1
4	2
5	5

Let us check the steps for calculating Spearman's Rank Correlation Coefficient:

Step 1: Rank the data for both X and Y, separately, from lowest to highest values. Assign the same rank for tied values.

Commute Time (X)	Rank of X	Job Satisfaction (Y)	Rank of Y
1	1	4	2
2	2	3	3

SGOU

3	3	1	5
4	4	2	4
5	5	5	1

Step 2: Calculate the differences between the ranks of X and the ranks of Y for each data point.

Commute Time (X)	Rank of X	Job Satisfaction(Y)	Rank of Y	Rank Difference (d)
1	1	4	2	1
2	2	3	3	1
3	3	1	5	2
4	4	2	4	0
5	5	5	1	4

Step 3: Square the rank differences (d) and calculate the sum of squared differences (Σd^2).

$$\Sigma d^2 = 1^2 + 1^2 + 2^2 + 0^2 + 4^2 = 22$$

Step 4: Calculate the Spearman Rank Correlation Coefficient (ρ) using the formula:

$$\rho = 1 - \frac{6\Sigma d^2}{n\,(n^2 - 1)}$$

Where 'n' is the number of data points, which in this case is 5.

$$\rho = 1 - \frac{6 \times 22}{5\,(5^2 - 1)}$$

$$= 1 - 132/120$$

$$= 1 - 1.1$$

$$= -0.1$$

The resulting ρ value is approximately -0.1. Since Spearman's rank correlation ranges from -1 to 1, this suggests a very weak negative correlation between commute time and job satisfaction in this dataset. In other words, as commute time increases, job satisfaction tends to slightly decrease, but the relationship is not very strong.

Illustration 3.1.8

Following are the marks obtained by 10 students of two subjects Mathematics and Physics in a class test. Estimate Spearman's Rank Correlation.

SGOU

Name of students	Mathematics	Physics
A	7	6
B	3	2
C	1	1
D	4	5
E	6	8
F	8	7
G	2	3
H	5	4

Solution

Name of students	Mathematics	Rank in Mathematics	Physics	Rank in Physics	d	d^2
A	7	2	6	3	1	1
B	3	6	2	7	1	1
C	1	8	1	8	0	0
D	4	5	5	4	1	1
E	6	3	8	1	2	4
F	8	1	7	2	1	1
G	2	7	3	6	1	1
H	5	4	4	5	1	1
					Σd^2	10

$$\rho = 1 - \frac{6\Sigma d^2}{n\,(n^2 - 1)}$$

$$\rho = 1 - \frac{6 \times 10}{8\,(8^2 - 1)}$$

$$= 1 - 60/504$$

= 1- 0.12

= 0.88

A rank correlation coefficient of 0.88 suggests a strong positive relationship between the ranked variables, indicating that they generally change together in the same direction.

SGOU

Illustration 3.1.9

The data given below relates to the price and demand of a commodity over a period. Compute the correlation coefficient between the Price and Demand.

Price	50	75	60	70	95	90	88
Demand	100	140	110	115	150	134	120

Solution

Price	Rank of Price	Demand	Rank of Demand	Rank Difference(d)	d^2
50	7	100	7	0	0
75	4	140	2	2	4
60	6	110	6	0	0
70	5	115	5	0	0
95	1	150	1	0	0
90	2	134	3	1	1
88	3	120	4	1	1
					$\Sigma d^2=6$

Spearman's Rank correlation Coefficient $= \rho = 1 - \frac{6\Sigma d^2}{n\,(n^2-1)}$

$$\rho = 1 - \frac{6 \times 6}{7\,(7^2 - 1)}$$

$$= 1 - 36/336$$

$$= 1 - 0.1071$$

$$= 0.8929$$

A rank correlation of 0.8929 means a strong tendency for the ranked variables to change together in a similar direction, but not at an even pace. This shows a solid positive relationship between them. The closer the value is to 1, the stronger the connection.

ii. Computation for tied observations (Repeated Rank)

If there exist two or more same or equal values the ranking is said to be tied. In such case an average rank is to be given to each item. For example, if the value is repeated twice at the second rank, the common rank is to be assigned to each item. $(2 + 3)/2 = 2.5$, which is the average of 2 and 3.

When equivalent ranks are assigned to multiple entries, certain adjustments in the formula become necessary for calculating the Rank Correlation coefficient. This adjust-

SGOU

ment involves adding. $\frac{m^3-m}{12}$ to the sum of squared differences $\sum D^2$, where m stands for the number of items with the common rank. If there is more than one such group of items with the same rank, the value is added as many times as the number of such groups. The formula in that case is written as

$$r = 1 - \frac{6(\sum D^2 + \frac{m^3 - m}{12})}{n\,(n^2 - 1)}$$

where,

'm' stands for the number of each value repeat.

$\frac{m^3 - m}{12}$ is called Correction factor.

Illustration 3.1.10

Compute the correlation coefficient between the two variables x and y

x	2.7	2.5	2.7	2.5	2.5	2.7	3.5	4.9	5.4	5.4	3.8
y	30	27	50	76	37	14	34	75	60	121	74

Solution

x	y	Rank of x	Rank of y	d	d^2
2.7	30	7	9	2	4
2.5	27	10	10	0	0
2.7	50	7	6	1	1
2.5	76	10	2	8	64
2.5	37	10	7	3	9
2.7	14	7	11	4	16
3.5	34	5	8	3	9
4.9	75	3	3	0	0
5.4	60	1.5	5	3.5	12.25
5.4	121	1.5	1	0.5	0.25
3.8	74	4	4	0	0
					$\Sigma d^2 = 115.50$

In X series 5.4 occurs twice, $m^3 - m = 2^3 - 2 = 6$

Similarly, 2.7 occurs thrice, $m^3 - m = 3^3 - 3 = 24$

2.5 occurs thrice, $m^3 - m = 3^3 - 3 = 24$

SGOU

Total $m^3 - m = 6 + 24 + 24 = 54$

$$r = 1 - \frac{6(\sum D^2 + \frac{m^3 - m}{12})}{n\,(n^2 - 1)}$$

$$= 1 - (6 \times 120)/1320$$

$$= 1 - 720/1320$$

$$= 1 - 0.545$$

$$= 0.454$$

Illustration 3.1.11

Given the following aptitude and I.Q. scores for a group of students. Find the coefficient of rank correlation

Aptitude Score	57	58	59	59	60	61	60	64
I.Q. Score	97	108	95	106	120	126	113	110

Solution

X	Y	*Rank in X*	*Rank in Y*	d	d^2
57	97	8	7	1	1
58	108	7	5	2	4
59	95	5.5	8	-2.5	6.25
59	106	5.5	6	-0.5	0.25
60	120	3.5	2	1.5	2.25
61	126	2	1	1	1
60	113	3.5	3	0.5	0.25
64	110	1	4	-3	9
					24

Here two correction factors are to be added to the equation, for X series,

59 is repeated twice, $m^3 - m = 2^3 - 2 = 6$

60 is repeated twice, $m^3 - m = 2^3 - 2 = 6$

Total $m^3 - m = 6 + 6 = 12$

SGOU

$$r = 1 - \frac{6(\sum D^2 + \frac{m^3 - m}{12})}{n\,(n^2 - 1)}$$

$$= 1 - \ 150/504$$

$$= 1 - \ 0.2976$$

$$= 0.7024$$

3.1.5.5 Concurrent Deviation Method

The concurrent deviation method finds the correlation coefficient between two variables. This method involves finding the deviations of the values of the variable from its preceding value and does not consider the exact magnitude of the values of the variables. Thus, we put a plus (+) sign, minus (–) sign or Zero for the deviation if the value of the variable is greater than, less than or equal to the preceding value, respectively. It is a straightforward way to calculate the correlation between two sets of data points. In this method, only the directions of deviations and magnitudes are taken. If the two variables' deviations are concurrent, they move in the same direction; otherwise in the opposite direction.

The formula for finding the coefficient of concurrent deviation is:

$$r = \pm\sqrt{\pm\frac{(2C - n)}{n}}$$

Where,

C = The number of pairs of concurrent deviation and

n = The number of pairs of deviations.

It should be clearly noted that here n is not the number of pairs of observations but it is the number of pairs of deviations and as such it is one less than the number of pairs of observations.

Since $-1 \leq r \leq 1$, the quantity inside the square root, $\pm\frac{(2C - n)}{n}$ must be positive, otherwise r will be imaginary which is not possible. Thus if $(2c - n)$ is positive, we take positive sign in and outside the square root in $\pm\sqrt{\pm\frac{(2C-n)}{n}}$ and if $(2c - n)$ is negative, we take negative sign in and outside the square root in $\pm\sqrt{\pm\frac{(2C-n)}{n}}$.

Illustration 3.1.12

Find out the correlation between two arrays of variables X and Y using the concurrent deviation method.

SGOU

X	13	18	23	8	21	25	28	10	22
Y	23	11	17	3	23	18	8	23	20

Solution

X	Y	Sign of deviations from the proceeding value in X dx	Sign of deviations from the proceeding value in Y dy	$dx \times dy$
13	23	-------	-----	------
18	11	+	-	-
23	17	+	+	+
8	3	-	-	+
21	23	+	+	+
25	18	+	-	-
28	8	+	-	-
10	23	-	+	-
22	20	+	-	-

C = Number of + signs in $dx \times dy$ column = 3

$n = 9 - 1 = 8$ (first pair of observation is not compared)

$$r = \pm\sqrt{\pm\frac{(2C - n)}{n}}$$

$$= \pm\sqrt{\pm\frac{(2 \times 3 - 8)}{8}}$$

$$= \pm\sqrt{(\pm(-2)/8)}$$

$$= \pm\sqrt{(\pm(-0.25))}$$

$$= -0.5$$

Since $2c - n = -0.25$, i.e., (negative), we take negative sign inside and outside the square root to get, $r = -\sqrt{(-(-0.25))} = -0.5$.

Illustration 3.1.13

Calculate the coefficient of concurrent deviations for the following data

SGOU

Supply	65	40	35	75	63	80	35	20	80	60	50
Demand	60	55	50	56	30	70	40	35	80	75	80

Solution

x	y	Direction of change in x dx	Direction of change in y dy	$dx \times dy$
65	60	-------	-----	------
40	55	-	-	+
35	50	-	-	+
75	56	+	+	+
63	30	-	-	+
80	70	+	+	+
35	40	-	-	+
20	35	-	-	+
80	8	+	+	+
60	75	-	-	+
50	80	-	+	-

C = Number of + signs in $dx \times dy$ column = 9

$n = 11 - 1 = 10$ (first pair of observation is not compared)

$$r = \pm\sqrt{\pm\frac{(2C - n)}{n}}$$

$$= \pm\sqrt{\pm\frac{(2 \times 9 - 10)}{10}}$$

$$= \pm\sqrt{(\pm 8/10)}$$

$$= \pm\sqrt{(\pm(0.8))}$$

$$= 0.89$$

Illustration 3.1.14

Calculate the coefficient of concurrent deviations from the data given below:

SGOU

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001
Supply	160	164	172	182	166	170	178	192	186
Demand	292	280	260	234	266	254	230	190	200

Solution

Year	x	y	Direction of change in x dx	Direction of change in y dy	$dx \times dy$
1993	160	292	-------	-----	------
1994	164	280	+	-	-
1995	172	260	+	-	-
1996	182	234	+	-	-
1997	166	266	-	+	-
1998	170	254	+	-	-
1999	178	230	+	-	-
2000	192	190	+	-	-
2001	186	200	-	+	-

C = Number of + signs in $dx \times dy$ column = 0

$n = 9 - 1 = 8$ (first pair of observation is not compared)

$$r = \pm\sqrt{\pm\frac{(2C - n)}{n}}$$

$$= \pm\sqrt{\pm\frac{(2 \times 0 - 8)}{8}}$$

$$= \pm\sqrt{(\pm(-8)/8)}$$

$$= \pm\sqrt{(\pm(-1))}$$

Since $2c - n = -8$, i.e., (negative), we take negative sign inside and outside the square root to get, $r = -\sqrt{(-(-1))} = -1$. Hence, there is perfect negative correlation between the supply and the price.

Illustration 3.1.15

Calculate the coefficient of concurrent deviations from the following data.

No. of pairs of observations = 96

No. of pairs of concurrent deviations = 36

SGOU

Solution

C = Number of pairs of concurrent deviations = 36

$n = 96 - 1 = 95$ (first pair of observation is not compared)

$$r = \pm\sqrt{\pm\frac{(2C - n)}{n}}$$

$$= \pm\sqrt{\pm\frac{(2 \times 36 - 95)}{95}}$$

$$= \pm\sqrt{(\pm\,(72 - 95)/95)}$$

$$= \pm\sqrt{(\pm\,(-23)/95)}$$

$$= \pm\sqrt{(\pm(-0.242)\,)}$$

$$= -\sqrt{(-(-0.242)\,)}$$

$$= -\,0.492$$

Note:

1. Increase in the value is denoted by + sign and decrease by – sign.
2. r is positive when $2C > n$ and r is negative when $2C < n$.

Summarised Overview

Correlation refers to the statistical measure that expresses the extent to which two variables are related. It helps in understanding whether an increase or decrease in one variable corresponds to an increase or decrease in another. The significance of correlation lies in its ability to reveal patterns, guide predictions, and support decision-making across disciplines. Correlation can be classified into types such as positive, negative, and zero correlation, and further into linear and non-linear forms. To study correlation, several methods are employed: the scatter diagram offers a visual representation of data relationships; Karl Pearson's coefficient quantifies the strength and direction of linear correlation; Spearman's rank correlation is used for ordinal data or non-linear relationships; and the concurrent deviation method provides a simplified approach for assessing directional agreement. Additionally, the concept of probable error helps evaluate the reliability of the correlation coefficient, indicating whether the observed relationship is statistically significant or likely due to chance.

SGOU

Self-Assessment Question

1. What is correlation analysis, and what is its purpose in statistical studies?
2. Describe the idea of a perfectly positive correlation between two variables.
3. Differentiate between linear and non-linear types of correlation.
4. In practical terms, how can a negative correlation be understood and interpreted?
5. What does it mean when two variables exhibit no correlation? Share a real-life example where this might occur.
6. How can recognising a positive correlation assist in making better decisions?
7. What is meant by non-linear correlation? Give a real-world example where this type of relationship is evident.
8. What is a correlation coefficient, and how is it defined?
9. What does a correlation coefficient value of zero indicate about the relationship between variables?
10. Outline the procedure for computing the Pearson correlation coefficient
11. Find the correlation coefficient between the x and y series for the following data,

$$n = 15, \sigma x = 3.01, \sigma y = 3.03, \sum(x - \bar{x})(y - \bar{y}) = 122$$

12. The correlation coefficient between two variables x and y is 0.48. The covariance is 36, and the variance of x is 16. Find the standard deviation of y.
13. Find Karl Pearson's correlation coefficient between x and y for the following data,

$$n = 10, \Sigma x = 12, \ \Sigma x^2 = 450, \Sigma y = 15, \Sigma y^2 = 320, \Sigma xy = 285$$

14. What is a Scatter Diagram? How is it useful in the study of correlation?

SGOU

Assignments

1. If a study finds a strong positive link between how often individuals exercise and their level of cardiovascular fitness, how should this relationship be understood?

2. What does a moderate negative association between regional temperature and ice cream sales suggest, and how might it be interpreted?

3. In examining student attendance and academic achievement, how would you interpret a slight positive relationship between these two factors?

4. When a strong positive correlation is observed between time spent playing video games and increased aggression in children, what considerations should this raise for parents and educators?

5. Compute Karl Pearson's Coefficient of Correlation

Price (Rs.)	11	12	13	14	15	16	17	18	19	20
Demand (Rs.)	30	29	29	25	24	24	24	21	18	15

(Ans: -0.962)

6. Find Karl Pearson's coefficient of correlation between x and y from the following data giving test scores of 10 candidates in mathematics and statistics and interpret

Scores in mathematics	98	70	40	20	85	75	95	80	10	5
Scores in statistics	85	65	32	30	80	60	80	70	20	10

(Ans:- 0.99, There is very good positive correlation)

7. Calculate coefficient of correlation by Spearman's method from the following data.

Roll no.	1	2	3	4	5	6	7	8	9	10
Marks in statistics	45	56	39	54	45	40	56	60	30	35
Marks in law	40	56	30	44	36	32	45	42	20	36

(Ans:- 0.86)

8. Apply concurrent deviation method to determine the value of coefficient of correlation from the data given below.

X	100	110	110	120	122	125
Y	120	140	160	160	130	110

(Ans: -0.58)

SGOU

9. Coefficient of correlation between two variates X and Y is 0.48, there covariation is 36. The variance of x=16. Find the standard deviation of Y series. **(Ans:- 18.75)**

10. If r = 0.8 and N = 0.81, find out the probable error of the coefficient of correlation. **(Ans: 0.027)**

Reference

1. Aggarwal, B. M. (2010). *Business statistics*. Ane Books Pvt. Ltd.
2. Chandrasekharan, N., & Umaparvathi, M. (2016). *Statistics for management*. PHI Learning Pvt. Ltd.
3. Gupta, M., & Khanna, R. (2011). *Quantitative techniques for decision making*. Prentice Hall of India.
4. Gupta, S. P. (2021). *Statistical methods*. S. Chand & Sons.
5. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International (P) Ltd.
6. Levin, R. I., & Rubin, D. S. (1999). *Quantitative approaches for management*. Pearson Education.
7. Sharma, A. (2017). *Quantitative techniques for decision making*. Himalaya Publishing House.
8. Vohra, D. V. D. (2017). *Quantitative techniques for management*. New Delhi: McGraw Hill Education

Suggested Reading

1. Arora, P. N., Arora, S., & Arora, S. (2010). *Comprehensive statistical methods*. S. Chand & Company Pvt. Ltd.
2. Beri, G. C. (2017). *Business statistics*. Tata McGraw-Hill.
3. Chawla, D., & Sondhi, N. (2016). *Research methodology: Concepts and cases*. Vikas Publishing House.

SGOU

4. Krishnaswami, O. R., Ranganathan, M., & Harikumar, P. N. (2016). *Research methodology*. Himalaya Publishing House.

5. Render, B. (2022). *Quantitative analysis for management*. Prentice Hall of India.

6. Singh, A. K. (2017). *Tests, measurements and research methods in behavioural sciences*. Bharti Bhawan.

SGOU

Unit 2

Regression Analysis

Learning Outcomes

On completion of this unit, learners will be able to:

- ◊ explain regression and its role in analysing relationships between dependent and independent variables in quantitative research
- ◊ distinguish between various types of regression, focusing on linear regression, and interpret regression lines to understand predictive trends
- ◊ construct regression equations and apply them to real-world business scenarios such as forecasting sales, estimating costs, or analysing consumer behaviour
- ◊ make them aware of the differences between regression and correlation, recognising their distinct purposes, assumptions, and interpretive frameworks

Background

The origins of regression analysis trace back to the late 19th century, when Sir Francis Galton introduced the concept while studying the relationship between parents' and offspring's heights. He observed a tendency for extreme traits to regress toward the mean, coining the term "regression" in his 1886 paper. Building on Galton's work, Karl Pearson formalised the mathematical foundations of correlation and regression, laying the groundwork for statistical modelling in the social sciences. These early developments were pivotal in establishing regression as a tool for quantifying relationships between variables, particularly in biological and anthropometric studies.

By the early 20th century, regression techniques had evolved beyond simple bivariate models. Ronald A. Fisher's contributions in the 1920s expanded regression into experimental design and multivariate analysis, enabling more robust

SGOU

inference in agricultural and industrial research. As computing power grew in the mid-20th century, regression analysis became central to econometrics, marketing analytics, and operations research. Today, linear regression and its variants are indispensable in business decision-making, forecasting, and performance evaluation. This unit explores regression's meaning, types, and applications, including regression lines, equations, and the distinction between regression and correlation.



Keywords

Regression, Regression Lines, Linear Regression, Dependent and Independent Variables, Line of Best Fit



Discussion

Regression analysis is a powerful statistical method used to model and analyse the relationship between a dependent variable and one or more independent variables, enabling data-driven decision-making in business management. For instance, a retail manager may use linear regression to predict monthly sales based on advertising expenditure, identifying how changes in marketing budget influence revenue outcomes. By quantifying such relationships, regression helps managers optimise resource allocation, forecast future performance, and evaluate strategic interventions with greater precision.

3.2.1 Meaning of Regression

Forecasting or estimating future outcomes is a critical concern across nearly every area of human endeavour. For business professionals and economists alike, anticipating future trends in production, consumption, pricing, investment, sales, profits, and income is essential for informed decision-making. Regression analysis is a rigorous and systematic method for generating such predictions among the various analytical tools available.

Regression analysis is a statistical technique used to examine the nature and strength of the relationship between a dependent variable and one or more independent variables. It helps identify patterns, estimate the impact of influencing factors, and make predictions based on historical data. By fitting a mathematical equation to observed data points, regression allows analysts to understand how changes in input variables affect outcomes, making it an essential tool for forecasting, strategic planning, and performance evaluation across various domains, especially in business and economics.

SGOU

Example: A researcher studying rural entrepreneurship in Kerala uses regression analysis to examine how access to microcredit, training programmes, and infrastructure affects the income levels of women-led enterprises. By analysing survey data from 500 entrepreneurs across different districts, the regression model identifies that access to training has the most substantial positive impact on income, followed by microcredit availability. Infrastructure shows a weaker but still significant effect. These findings help policymakers prioritise interventions and allocate resources more effectively to support women's economic empowerment in rural areas.

3.2.2 Definition of Regression

According to **M. M. Blair,** "*Regression analysis is a mathematical measure of the average relationship between two or more variables in terms of the original data units".*

According to **George Udny Yule (1897),** *"Regression is the measure of the average relationship between two or more variables, expressed in terms of the expected value of one variable given the values of others."*

According to **John Tukey (1961),** *"Regression is not merely a tool for prediction, but a way to understand the structure of data and the influence of variables, especially when used with caution and graphical insight."*

David Cox (2006) states**,** *"Regression models are used to describe how the conditional distribution of a response variable depends on explanatory variables, often assuming a parametric form."*

3.2.3 Dependent and Independent Variables

In regression analysis, we study how one variable changes in response to another. These variables are classified as:

- **Dependent Variable (Response Variable):** This is the outcome or result we aim to explain or predict. Its value depends on other variables.
- **Independent Variable (Predictor or Explanatory Variable):** We use this variable to explain or predict changes in the dependent variable. It is assumed to influence or cause variation in the dependent variable.

Examples

Sales Forecasting

- **Independent Variable:** Advertising expenditure
- **Dependent Variable:** Monthly sales revenue
- **Explanation:** We use advertising spend to predict how much revenue the company will generate. More advertising is expected to lead to higher sales.

SGOU

Skill Development Impact

- ◊ **Independent Variable:** Number of training hours provided to rural entrepreneurs
- ◊ **Dependent Variable:** Increase in monthly income
- ◊ **Explanation:** The training hours are expected to influence income levels. The more training received, the higher the income may be.

3.2.4 Types of Regression

The regression analysis can be classified into:

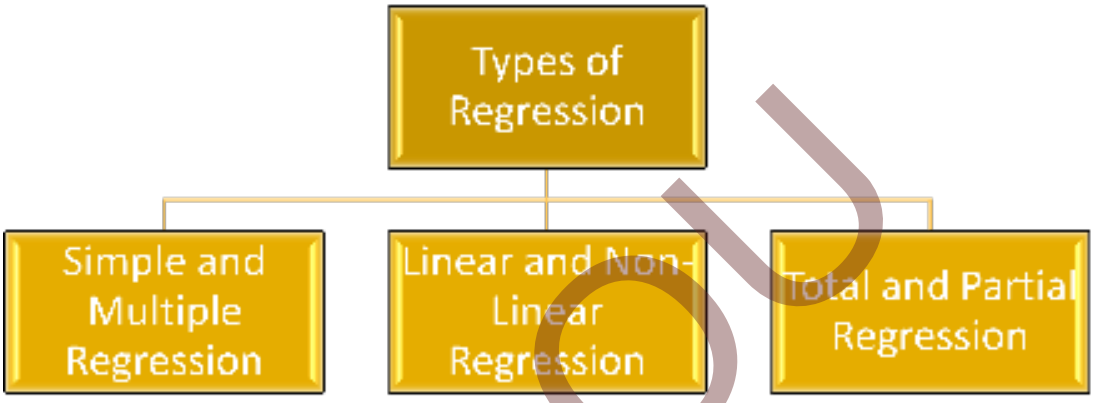


Fig. 3.2.1 Types of Regression

a. Simple vs. Multiple Regression

Simple regression explores the relationship between one independent and one dependent variable. For example, the price of a car (dependent), may be predicted based on its age (independent). Similarly, sales turnover can be modelled as a function of advertising expenditure: if y is turnover and x is ad spend, then y = f(x) represents a simple regression equation.

Multiple regression examines how two or more independent variables influence a single dependent variable. For instance, an individual's income (dependent) may depend on both education level and years of experience (independent variables). In this case, turnover z could be expressed as a function of advertising expenditure x and consumer income y, written as z = f (x, y).

b. Linear vs. Non-linear Regression

Linear regression applies when the relationship between variables follows a straight-line pattern. For example, sweetness increases proportionally with added sugar, reflecting a linear trend.

Non-linear regression is used when the relationship curves, such as weight increasing with height but eventually plateau. While both types can be applied in

SGOU

multiple or straightforward regression contexts, the key distinction lies in the shape of the relationship, straight line vs. curve.

c. Total vs. Partial Regression

In multiple regression, total regression assesses the combined effect of all independent variables on the dependent variable. For example, house prices are influenced by size, number of bedrooms, and location.

Partial regression isolates the impact of one independent variable while controlling for others. It helps identify the unique contribution of each factor, such as evaluating how house size alone affects price, independent of other variables.

3.2.5 Regression Line/ Line of Best Fit

A regression line is a straight line that best captures the relationship between a dependent variable and one or more independent variables within a dataset. It forms the core of a linear regression model, which seeks to minimise the gap between actual data points and the values predicted by the model. This line is derived from a regression equation that mathematically expresses how the variables are related.

Put, the regression line offers a visual summary of how changes in one variable correspond to changes in another. It reveals the underlying trend in the data and allows us to estimate the value of the dependent variable based on known values of the independent variable(s). By fitting this line to the data, we gain insights into the overall pattern of association and can apply this understanding to make accurate forecasts or informed decisions. As noted earlier, various methods for fitting a regression line depend on the context and data characteristics.

Properties of the Regression Line

- The regression line is constructed to minimise the sum of squared residuals (the differences between observed and predicted values), ensuring the line best fits the data.
- The two regression lines cut each other at the point of average of x and average of y (i.e.,x and y)
- When $r = 1$, the two regression lines coincide and give one line.
- When $r = 0$, the two regression lines are mutually perpendicular.
- The independent variable is not influenced by the errors in the dependent variable, ensuring that causality is addressed correctly.
- The regression line can be used to predict values of the dependent variable based on the known values of the independent variable(s).
- The regression line assumes a linear relationship between the variables. The regression line might not accurately represent the relationship if the actual relationship is non-linear.

SGOU

3.2.6 Methods of drawing Regression Lines

The regression lines can be drawn by the two methods

1. Freehand curve method
2. The method of least squares.

Equation of the Regression Line

In regression analysis, there are two types of variables . The variable whose value is influenced or is to be predicted is called the dependent variable and the variable which influences the value or is used for prediction is called the independent variable .

In the case of two variables, x and y, we shall have two lines of regression. One of them is "y on x" and the other "x on y".If y is the independent variable and x depends on y, the regression line is of "x on y".

The equation of the regression line of x on y is: $x - \bar{x} = b_{xy}(y - \bar{y})$

If x is the independent variable and y depends on x, the regression line is "y on x".

The equation of the regression line of y on x is: $y - \bar{y} = b_{yx}(x - \bar{x})$

where the regression coefficients are calculated using the following formulas:

$$b_{xy} = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2}$$

$$b_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

Also, we have

$$\bar{x} = \frac{\sum x}{n}$$

and

$$\bar{y} = \frac{\sum y}{n}$$

The regression coefficients are also related to the correlation coefficient (r) and standard deviations (σ_x, σ_y):

$$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$$

and

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$$

SGOU

$\bar{x}$ and $\bar{y}$ are the means of x and y values;

b_{yx} and b_{xy} are known as regression coefficients.

Q. There are two lines of regression. Why?

There are always two lines of regression: one of y on x and the other of x on y. The line of regression of y on x is used to estimate or predict the value of y for any given value of x (when y is the dependent variable and x is the independent variable). Similarly, to estimate or predict x for any given value of y, we use the regression line of x on y (where x is the dependent variable and y is the independent variable). The two regression equations are not reversible or interchangeable because the basis and assumptions for deriving these equations differ significantly.

Illustration 3.2.1

Find the regression equation of x on y for the following data:

x	5	6	7	3	2
y	4	5	8	2	1

Solution

The Regression Equation of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

where,

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{y} = \frac{\sum y}{n}$$

$$b_{xy} = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2}$$

x	**y**	**xy**	***y*2**
5	4	20	16
6	5	30	25
7	8	56	64
3	2	6	4
2	1	2	1
23	**20**	**114**	**110**

$$\bar{x} = \frac{\sum x}{n} = \frac{23}{5} = 4.6$$

SGOU

$$\bar{y} = \frac{\sum y}{n} = \frac{20}{5} = 4$$

$$b_{xy} = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2} = \frac{5 \times 114 - 20 \times 23}{5 \times 110 - (20)^2} = \frac{110}{150} \approx 0.73$$

Regression equation of x on y:

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 4.6 = 0.73(y - 4)$$

$$x - 4.6 = 0.73y - 4 \times 0.73$$

$$x = 0.73y + 1.68$$

Illustration 3.2.2

From the following data on the age of husband (x) and the age of wife (y), form the regression line and find the wife's age when the husband's age is 40.

Husband (x):	36	23	27	28	28	29	30	31	33	35
Wife (y):	29	18	20	22	27	21	29	27	29	28

Solution

Husband's Age (x)	Wife's Age (y)	xy	x^2
36	29	1044	1296
23	18	414	529
27	20	540	729
28	22	616	784
28	27	756	784
29	21	609	841
30	29	870	900
31	27	837	961
33	29	957	1089
35	28	980	1225
$\Sigma x = 300$	$\Sigma y = 250$	$\Sigma xy = 7623$	$\Sigma x^2 = 9138$

Since we have to find the age of the wife (y) when the husband's age (x) is 40, we use the regression equation of y on x

Since we have to find the age of the wife (y) when the husband's age (x) is given, we need the regression equation of y on x .

SGOU

Step 1: Calculate means

$$\bar{x} = \frac{\Sigma x}{n} = \frac{300}{10} = 30$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{250}{10} = 25$$

Step 2: Calculate the regression coefficient b_{yx}

The formula for b_{yx} is: $b_{yx} = \frac{n\Sigma xy - \Sigma x.\Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$

Substitute the values:

$$b_{yx} = \frac{10 \times 7623 - 300 \times 250}{10 \times 9138 - (300)^2} = \frac{76230 - 75000}{91380 - 90000} = \frac{1230}{1380} \approx 0.89$$

Step 3: Formulate the regression equation of y on x

The equation of the regression line of y on x is:

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

Substitute the calculated values:

$$y - 25 = 0.89(x - 30)$$

$$y - 25 = 0.89x - (30 \times 0.89)$$

$$y - 25 = 0.89x - 26.7$$

$$y = 0.89x - 26.7 + 25$$

$$y = 0.89x - 1.7$$

Step 4: Estimate wife's age when husband's age (x) is 40

Substitute $x = 40$ into the regression equation:

$$y = 0.89 \times 40 - 1.7$$

$$y = 35.6 - 1.7$$

$$y = 33.9$$

The estimated age of the wife when the husband's age is 40 is approximately 33.9 years.

Illustration 3.2.3

Find the coefficient of correlation between the variables of x & y using the regression coefficient.

SGOU

x	y	xy	x^2	y^2
2	3	6	4	9
3	5	15	9	25
4	4	16	16	16
5	8	40	25	64
6	9	54	36	81
$\Sigma x = 20$	$\Sigma y = 29$	$\Sigma xy = 131$	$\Sigma x^2 = 90$	$\Sigma y^2 = 195$

$n = 5$

Step 1: Calculate the regression coefficient b_{xy}

$$b_{xy} = \frac{n\Sigma xy - \Sigma x.\Sigma y}{n\Sigma y^2 - (\Sigma y)^2}$$

$$b_{xy} = \frac{5 \times 131 - 20 \times 29}{5 \times 195 - (29)^2} = \frac{655 - 580}{975 - 841} = \frac{75}{134} \approx 0.56$$

Step 2: Calculate the regression coefficient b_{yx}

$$b_{yx} = \frac{n\Sigma xy - \Sigma x.\Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$b_{yx} = \frac{5 \times 131 - 20 \times 29}{5 \times 90 - (20)^2} = \frac{655 - 580}{450 - 400} = \frac{75}{50} = 1.5$$

Step 3: Calculate the correlation coefficient r

$$r = \pm\sqrt{b_{xy} \times b_{yx}}$$

$$r = \pm\sqrt{0.56 \times 1.5} = \pm\sqrt{0.84} \approx \pm 0.92$$

Since both regression coefficients (0.56 and 1.5) are positive, the correlation coefficient is positive.

Answer: The coefficient of correlation (r) is 0.92

Illustration 3.2.4

Find the correlation coefficient using the regression coefficient for the following data.

SGOU

x	y	xy	x^2	y^2
4	2	8	16	4
2	3	6	4	9
3	2	6	9	4
4	4	16	16	16
2	4	8	4	16
$\Sigma x = 15$	$\Sigma y = 15$	$\Sigma xy = 44$	$\Sigma x^2 = 49$	$\Sigma y^2 = 49$

Step 1: Calculate the regression coefficient b_{yx}

$$b_{yx} = \frac{n\Sigma xy - \Sigma x.\Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$b_{yx} = \frac{5 \times 44 - 15 \times 15}{5 \times 49 - (15)^2} = \frac{220 - 225}{245 - 225} = \frac{-5}{20} = -0.25$$

Step 2: Calculate the regression coefficient b_{xy}

$$b_{xy} = \frac{n\Sigma xy - \Sigma x.\Sigma y}{n\Sigma y^2 - (\Sigma y)^2}$$

$$b_{xy} = \frac{5 \times 44 - 15 \times 15}{5 \times 49 - (15)^2}$$

$$b_{xy} = -0.25$$

Step 3: Calculate the correlation coefficient r

$$r = \pm\sqrt{b_{yx} \times b_{xy}}$$

$$r = \pm\sqrt{(-0.25) \times (-0.25)} = \pm\sqrt{0.0625} = \pm 0.25$$

Since both regression coefficients are negative, the correlation is negative.

Answer: The coefficient of correlation (r) is -0.25

Illustration 3.2.5

The two regression equations are $\mathbf{8x - 10y + 66 = 0}$ and $\mathbf{40x - 18y - 214 = 0}$. Find the Coefficient of Correlation.

Step 1: Assume equation (1) is x on y and equation (2) is y on x

Equation (1): $8x - 10y + 66 = 0$

SGOU

Solve for x to get b_{xy}

$$8x = 10y - 66$$

$$x = \frac{10}{8}y - \frac{66}{8}$$

$$b_{xy} = \frac{10}{8}$$

Equation (2): $40x - 18y - 214 = 0$

Solve for y to get b_{yx}

$$-18y = -40x + 214$$

$$y = \frac{-40}{-18}x + \frac{214}{-18} = \frac{40}{18}x - \frac{214}{18}$$

$$b_{yx} = \frac{40}{18}$$

Calculate r with these coefficients:

$$r = \pm\sqrt{b_{xy} \times b_{yx}} = \pm\sqrt{\frac{10}{8} \times \frac{40}{18}} = \pm\sqrt{\frac{400}{144}} \approx \pm\sqrt{2.77} \approx \pm 1.67$$

The correlation coefficient r cannot exceed 1, so our initial assumption was wrong.

Step 2: Assume equation (1) is y on x and equation (2) is x on y

Equation (1): $8x - 10y + 66 = 0$

Solve for y to get b_{yx}

$$-10y = -8x - 66$$

$$y = \frac{-8}{-10}x - \frac{66}{-10} = \frac{8}{10}x + \frac{66}{10}$$

$$b_{yx} = \frac{8}{10}$$

Equation (2): $40x - 18y - 214 = 0$

Solve for x to get b_{xy}

SGOU

$$40x = 18y + 214$$

$$x = \frac{18}{40}y + \frac{214}{40}$$

$$b_{xy} = \frac{18}{40}$$

Calculate r with these coefficients:

$$r = \pm\sqrt{b_{yx} \times b_{xy}} = \pm\sqrt{\frac{8}{10} \times \frac{18}{40}} = \pm\sqrt{\frac{144}{400}} = \pm\sqrt{0.36} \approx \pm 0.60$$

This value of r is valid. Since both regression coefficients (8/10 and 18/40) are positive, the correlation coefficient is positive.

Answer: The coefficient of correlation (r) is 0.60.

Illustration 3.2.6

Find two regression lines and estimate marks in B for Candidates who secured 50 marks in A.

	Subject A (x)	**Subject B (y)**
Mean Marks	$\bar{x} = 39.5$	$\bar{y} = 47.5$
S.D.	$\sigma_x = 10.8$	$\sigma_y = 16.8$
Coe. of Correlation	$r = 0.42$	

Step 1: Calculate the regression coefficient b_{xy}

$$b_{xy} = r.\frac{\sigma_x}{\sigma_y}$$

$$b_{xy} = 0.42 \times \frac{10.8}{16.8} \approx 0.27$$

Step 2: Formulate the regression equation of x on y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 39.5 = 0.27(y - 47.5)$$

$$x - 39.5 = 0.27y - 12.825$$

$$x = 0.27y + 26.675$$

SGOU

Step 3: Calculate the regression coefficient b_{yx}

$$b_{yx} = r.\frac{\sigma_y}{\sigma_x}$$

$$b_{yx} = 0.42 \times \frac{16.8}{10.8} \approx 0.65$$

Step 4: Formulate the regression equation of y on x

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 47.5 = 0.65(x - 39.5)$$

$$y - 47.5 = 0.65x - 25.675$$

$$y = 0.65x + 21.83$$

Step 5: Estimate marks in B (y) when marks in A (x) are 50

Substitute $x = 50$ into the regression equation of y on x

$$y = 0.65 \times 50 + 21.83$$

$$y = 32.5 + 21.83$$

$$y = 54.33$$

Answer:

The two regression equations are $\mathbf{x = 0.27y + 26.675}$ and $\mathbf{y = 0.65x + 21.83}$

The estimated marks in Subject B for a candidate who secured 50 marks in Subject A are 54.33.

3.2.7 Difference between Correlation and Regression

Although closely connected, correlation and regression are distinct statistical concepts. Below is a brief overview highlighting the key differences between them.

Table 3.2.1 Difference between Correlation and Regression

Point of Difference	Correlation	Regression
Purpose	Measures the strength and direction of a relationship between two variables	Predicts the value of the dependent variable based on the independent variable(s)
Focus	Focuses on the association between variables	Focuses on the degree of causal relationship between variable(s)

SGOU

Analysis	Analyses the degree of relationship between variables without establishing cause and effect	Analyses the nature of the relationship and estimates the effect of the independent variable(s) on the dependent variable
Result Range	Ranges from -1 to 1, indicating the strength and direction of the relationship	No specific range for the results; the emphasis is on the estimation and prediction of the dependent variable
Output	Provides a correlation coefficient (r) that indicates the strength and direction of the relationship	Provides an equation that represents the relationship between the variables
Application	Commonly used to examine the relationship between two continuous variables	Commonly used to make predictions, understand cause and effect, and control one variable to understand its effect on another

3.2.8 Application of Regression Equations in Business

Regression equations are powerful statistical tools used to model relationships between variables. In business, they help quantify how changes in one factor (independent variable) affect another (dependent variable), enabling data-driven decisions across various functions.

1. Sales Forecasting

Regression helps predict future sales based on influencing factors such as advertising spend, pricing, or seasonal trends.

Example: A retail company uses regression to model the relationship between monthly advertising budget and sales revenue. The equation might look like:

Sales = 5000 + 3.2 × Advertising Spend

If the company plans to spend ₹10,000 on advertising, the predicted sales would be ₹5000 + (3.2 × ₹10,000) = ₹35,000.

2. Pricing Strategy

Businesses use regression to understand how price changes affect demand or revenue, helping optimise pricing decisions.

SGOU

Example: A smartphone manufacturer analyses how different price points affect units sold. The regression equation shows that for every ₹100 increase in price, sales drop by 50 units. This insight helps set competitive prices without losing market share.

3. Cost Estimation and Budgeting

Regression can estimate costs based on production volume or other operational variables, aiding in budgeting and financial planning.

Example: A factory uses regression to estimate monthly electricity costs based on machine hours:

Electricity Cost = ₹2000 + ₹0.75 × Machine Hours

If machines run for 4000 hours, the estimated cost is ₹2000 + (₹0.75 × 4000) = ₹5000.

4. Human Resource Planning

Regression helps analyse employee performance, turnover rates, or training effectiveness based on various factors.

Example: An HR team models employee productivity based on training hours:

Productivity Score = 60 + 1.5 × Training Hours

This helps determine optimal training investment to boost performance.

5. Inventory Management

Regression can forecast inventory needs based on sales trends, seasonal demand, or promotional activities.

Example: A grocery store uses regression to predict weekly demand for milk based on temperature:

Milk Demand = 300 + 20 × Temperature (°C)

Warmer weeks lead to higher demand, helping the store stock appropriately.

6. Customer Behaviour Analysis

Businesses use regression to understand how customer demographics, income, or engagement levels influence purchasing decisions.

Example: An e-commerce platform models customer spending based on age and income:

Spending = ₹1000 + ₹0.5 × Income – ₹10 × Age

This helps tailor marketing strategies to different customer segments.

SGOU

7. Risk Assessment and Credit Scoring

Financial institutions use regression to assess credit risk by modelling the likelihood of default based on income, debt level, and credit history.

Example: A bank uses regression to predict loan default probability:

Default Risk Score = 0.3 × Debt – 0.2 × Income + 0.5 × Credit History Score

This supports informed lending decisions.

8. Market Research and Demand Analysis

Regression helps identify market demand factors, guiding product development and market entry strategies.

Example: A beverage company analyses how temperature and advertising affect demand for cold drinks:

Demand = 1000 + 25 × Temperature + 10 × Ad Spend

This informs seasonal marketing and distribution planning.

9. Operational Efficiency and Process Improvement

Regression is used to identify bottlenecks or inefficiencies in operations by linking output to input variables.

Example: A logistics firm models delivery time based on distance and traffic conditions:

Delivery Time = 30 + 0.5 × Distance + 10 × Traffic Index

This helps optimise routing and scheduling.

Summarised Overview

Regression is a statistical technique used to examine the relationship between a dependent variable and one or more independent variables. It helps predict outcomes and identify trends based on historical data. The syllabus begins with the foundational meaning of regression. It introduces its major types, primarily simple regression, which involves one independent variable, and multiple regression, which includes two or more predictors. These types are essential for analysing complex business scenarios where multiple factors influence outcomes. A regression line is a graphical representation of the relationship between variables, typically derived using the least squares method. In linear regression, this line is a straight line that best fits the data points, indicating the direction and strength of the relationship. The syllabus focuses on linear regression due to its simplicity and wide applicability in business decision-making. It also introduces

SGOU

the concept of slope and intercept, which help quantify the impact of changes in independent variables on the dependent variable.

Regression equations are mathematical expressions derived from data to model relationships and make predictions. These equations are vital tools in business analytics, enabling applications such as sales forecasting, cost estimation, demand analysis, and risk assessment. By applying regression models, businesses can optimise operations, allocate resources efficiently, and support strategic planning with data-driven insights. While both regression and correlation deal with relationships between variables, they serve distinct purposes. Correlation measures the degree and direction of association between variables but does not imply causation. In contrast, regression quantifies the relationship and allows for prediction and causal inference.



Self-Assessment Question

1. What is an independent variable? Illustrate with a relevant example.
2. Define a dependent variable and support your explanation with an example.
3. Describe the concept of linear regression and demonstrate its practical application with a real-world scenario.
4. Why is non-linear regression important for analysing complex variable relationships? Provide a detailed explanation.
5. How do multiple and straightforward regression differ? Discuss their respective benefits and uses in practical data analysis.
6. Explain the role of multiple regression in disciplines like economics, where multiple factors affect a single outcome. Include a suitable example.
7. Clarify the distinction between total regression and partial regression.
8. Provide a clear definition of regression as a statistical concept.
9. Enumerate the key characteristics of regression coefficients.
10. Outline the main features of regression lines.
11. Prove that $r = \sqrt{(byx \times bxy)}$
12. Write your inference when the regression lines are perpendicular.

SGOU

13. Mention the differences between correlation and regression analysis

14. Find the regression equation of y on x

X	2	4	6	8	10
Y	5	7	9	8	11

(Ans:- y= 0.65x + 4.1)

15. From the data given below, obtain the regression equation of x on y

X	2	3	7	8	10
Y	10	9	11	8	12

(Ans:- 0.8y – 2)

16. The following data shows the monthly income and expenditure on food of 10 families

Income	120	90	83	150	130	140	110	95	75	105
Expenditure	40	36	40	45	40	44	45	38	50	35

Calculate the linear regression of Expenditure of food(y) on Income(x)

(Ans:- y = 0.019x + 39.2)

17. Price index numbers of wheat x and cereals y at 12 successive seasons (quarters) are given below

X	87	84	88	102	101	84	72	84	83	98	97	100
Y	88	79	83	97	96	90	82	84	88	100	80	102

i. Fit a line of regression of Y on X

ii. Suggest what value of Y will be when X is expected to be 110.

(Ans:- (i) Y = ,0.576x + 37.2 (ii) Y = 100.6)

18. For 10 observations on price X and supply Y, the following data were obtained (in appropriate units)

$$\sum X = 130, \sum Y = 220, \sum XY = 3467, \sum X^2 = 2288, \sum Y^2 = 5506$$

Obtain the line of regression of Y on X and estimate the supply when the price is 16 units

(Ans :- Y = 8.8 +1.015x 25.04)

SGOU

Assignments

1. A biologist observes that a population initially grows quickly but slows down as resources become scarce. Why is non-linear regression essential for accurately modelling and forecasting this type of population growth?
2. As an agricultural researcher examining how fertiliser usage affects crop yields in a given area, how can linear regression assist in analysing this relationship? In what ways could the insights gained be applied to enhance farming practices locally?
3. If you're a real estate analyst trying to estimate property prices in a city, how can multiple regression help in this analysis? How might the outcomes of this study be valuable to stakeholders such as buyers, sellers, and developers?
4. To forecast students' exam performance, how would you incorporate study hours along with other relevant factors like class attendance and previous academic records into your prediction model?

Reference

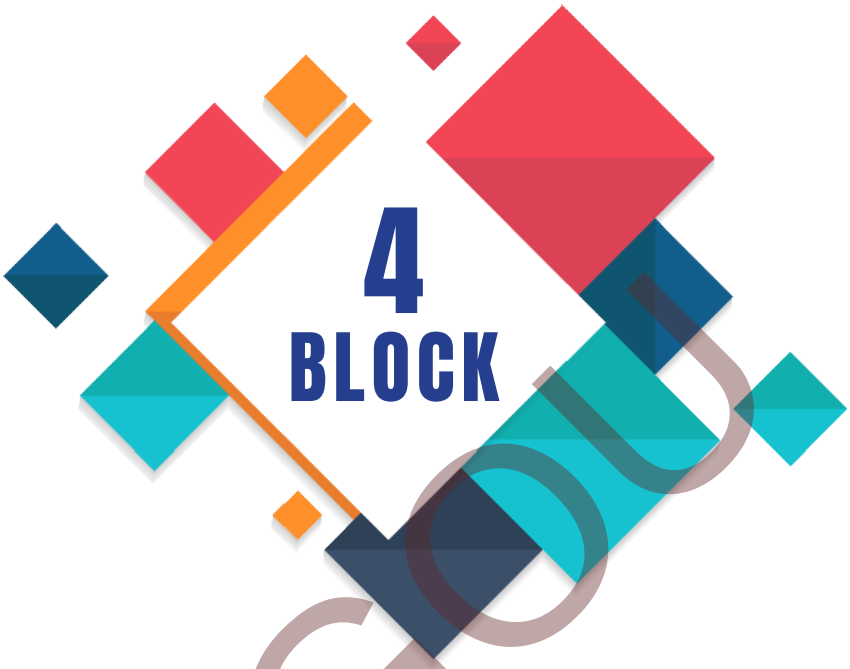
1. Aggarwal, B. M. (2010). *Business statistics*. Ane Books Pvt. Ltd.
2. Chandrasekharan, N., & Umaparvathi, M. (2016). *Statistics for management*. PHI Learning Pvt. Ltd.
3. Gupta, M., & Khanna, R. (2011). *Quantitative techniques for decision making*. Prentice Hall of India.
4. Gupta, S. P. (2021). *Statistical methods*. S. Chand & Sons.
5. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International (P) Ltd.
6. Levin, R. I., & Rubin, D. S. (1999). *Quantitative approaches for management*. Pearson Education.
7. Sharma, A. (2017). *Quantitative techniques for decision making*. Himalaya Publishing House.
8. Vohra, D. V. D. (2017). *Quantitative techniques for management*. New Delhi: McGraw Hill Education

SGOU

Suggested Reading

1. Arora, P. N., Arora, S., & Arora, S. (2010). *Comprehensive statistical methods*. S. Chand & Company Pvt. Ltd.
2. Beri, G. C. (2017). *Business statistics*. Tata McGraw-Hill.
3. Chawla, D., & Sondhi, N. (2016). *Research methodology: Concepts and cases*. Vikas Publishing House.
4. Krishnaswami, O. R., Ranganathan, M., & Harikumar, P. N. (2016). *Research methodology*. Himalaya Publishing House.
5. Render, B. (2022). *Quantitative analysis for management*. Prentice Hall of India.
6. Singh, A. K. (2017). *Tests, measurements and research methods in behavioural sciences*. Bharti Bhawan.

SGOU



Time Series Analysis and Index Numbers

SGOU

Unit 1

Time Series Analysis

Learning Outcomes

On completion of this unit, learners will be able to:

- ◊ explain the concept of time series and its relevance in quantitative analysis
- ◊ identify the key components of time series data: trend, seasonal, cyclical, and irregular variations
- ◊ apply appropriate techniques such as moving averages and the method of least squares to measure long-term trends
- ◊ evaluate the utility of time series analysis in forecasting and decision-making within business and management contexts
- ◊ integrate time series methods into real-world business scenarios, including sales forecasting, inventory planning, and financial analysis

Background

The origins of time series analysis can be traced back to the early 19th century, when economists and statisticians began systematically recording data over time to understand economic fluctuations. Early applications focused on agricultural output, commodity prices, and population growth, with scholars like William Stanley Jevons pioneering the study of periodicity in economic cycles. The development of graphical methods and rudimentary trend estimation laid the foundation for more formal techniques, including moving averages and regression-based models.

By the mid-20th century, time series analysis had evolved into a rigorous statistical discipline supported by computing and mathematical modelling advances. Its adoption in business and management accelerated with the rise of operations research and econometrics, enabling firms to forecast sales, manage

SGOU

inventories, and analyse financial performance. Today, time series methods are indispensable tools in strategic planning, risk assessment, and data-driven decision-making across industries. This unit gives an analytical framework for understanding, measuring, and applying time series techniques in business and management contexts.

Keywords

Time Series, Trend Analysis, Seasonal trend, Cyclical component, Secular Trend, Irregular Component, Moving Average, Least Squares Method, Forecasting

Discussion

Time Series Analysis is a statistical technique used to analyse data points collected or recorded at successive intervals over time, enabling the identification of patterns, trends, and seasonal variations. It is crucial in forecasting and strategic decision-making, particularly in business contexts where historical data informs future planning. For example, a retail company may use time series analysis to examine monthly sales figures over several years, identifying seasonal peaks during festive periods and long-term growth trends. By applying moving averages or least squares regression methods, the company can forecast future sales, optimise inventory levels, and align marketing strategies with anticipated demand.

4.1.1 Time Series - Meaning

Time series refers to the systematic arrangement of statistical observations recorded in sequential time order. It involves a series of data points captured at consistent intervals, allowing analysts to study patterns and variations over time. This method is primarily used for forecasting, which entails estimating future values of a variable by examining its historical behaviour. When predictions are made solely using the past values of the same variable, the process is known as time series forecasting. For instance, consider monitoring the monthly sales figures of a well-known ice cream parlour over a year, this data sequence forms a time series that can be analysed to anticipate future sales trends.

Month	Sales
Jan	5000
Feb	5200
Mar	6000

SGOU

Apr	8500
May	10000
Jun	12500
Jul	15000
Aug	14000
Sep	11200
Oct	9000
Nov	7500
Dec	6200

In this illustration, each entry corresponds to a particular month and its associated sales data, forming a time series. Analysing this sequence can assist in predicting future sales patterns and uncovering seasonal fluctuations for example, increased sales during the summer months.

Definition

According to ***Mooris Hamburg***, *"A time series is a set of statistical observations arranged in chronological order."*

According to ***Chris Chatfield***, *"A time series is a collection of observations made sequentially through time."*

4.1.2 Utility of Time Series Analysis

Time series analysis is pivotal in uncovering patterns, trends, and valuable insights embedded in chronological data. Observing how variables evolve over time facilitates predictive modelling, clarifies fluctuations, and supports the anticipation of future developments. In finance, economics, and business domains, this analytical approach informs strategic decisions, enabling organisations to forecast market movements, allocate resources efficiently, and adapt to changing conditions. It is instrumental in long-term planning, risk mitigation, and deriving actionable intelligence from temporal datasets, thereby enriching understanding of historical, current, and prospective trends.

The following points outline the key utilities of time series analysis:

i. **Trend Forecasting**: Enables the projection of future outcomes by analysing historical data patterns, supporting proactive planning and informed decision-making.

ii. **Detection of Patterns and Anomalies**: Facilitates the identification of consistent trends and irregular behaviours, offering deeper insights into temporal dynamics.

iii. **Strategic Planning and Decision-Making**: Assists businesses in optimising resource deployment, refining operational strategies, and enhancing overall performance through data-driven decisions.

SGOU

iv. **Risk Assessment and Management**: A framework for understanding variability and trends allows organisations to anticipate and mitigate risks linked to market or economic shifts.

v. **Efficient Resource Utilisation**: Supports operational planning by pinpointing demand cycles and peak activity periods, thereby improving efficiency and aligning resources with actual needs.

4.1.3 Components of Time Series

In time series analysis, data can be imagined as a narrative unfolding over time, with its components acting as central characters in the storyline. The trend is the protagonist, reflecting the overall upward, downward, or stable direction. Accompanying it is seasonality, the reliable companion, introducing consistent and predictable patterns such as those linked to holidays or weekends. Then comes the cyclical component, the unpredictable ally, representing longer-term fluctuations that do not follow a fixed schedule. Finally, random variation, or noise, adds an element of unpredictability, much like unexpected plot twists.

Grasping these components is akin to understanding the cast of a story; it equips us to interpret the data's progression over time and enhances our ability to make informed forecasts. Before delving into the actual analysis, it is essential to develop a clear understanding of each of these elements.

There are mainly four components to time series;

- ◊ Secular Trend
- ◊ Cyclic Component
- ◊ Seasonal Trend
- ◊ Irregular Component (Erratic)

This may be classified according to the time as in long term and short term as follows:

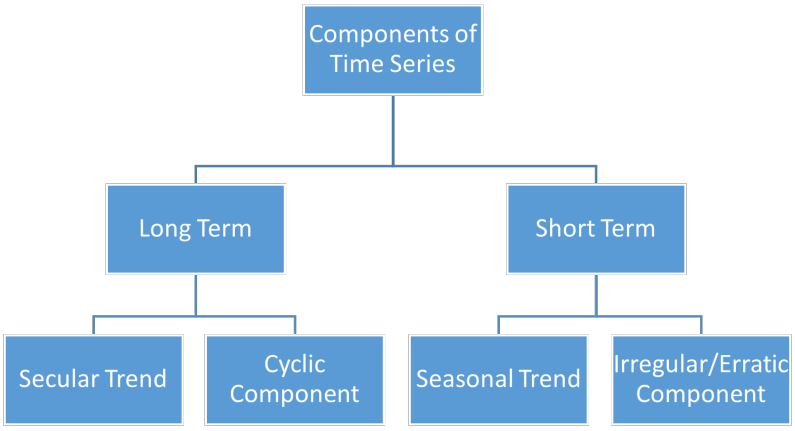


Fig. 4.1.1 Components of Time Series

SGOU

4.1.3.1 Secular Trend

A secular trend, often referred to simply as the trend, represents the long-term movement within a time series. It reflects the persistent tendency of data to rise, fall, or remain stable over an extended period. Think of it as the central narrative thread in the story of your data highlighting the overarching direction and revealing enduring shifts that unfold gradually over time.

For instance, consider tracking the population of a city across several decades. If the population steadily grows yearly, driven by factors such as urbanisation and employment opportunities, that sustained increase exemplifies a secular trend. While short-term fluctuations may occur due to economic challenges or infrastructure developments, the consistent upward trajectory signals a more profound, structural change. This long-term pattern serves as the demographic storyline of the city, guiding policymakers and planners in making informed decisions for future development.

Understanding the secular trend is essential, providing the foundational context for interpreting the data's broader journey. The accompanying figure illustrates this concept using India's population data, visually representing long-term growth.

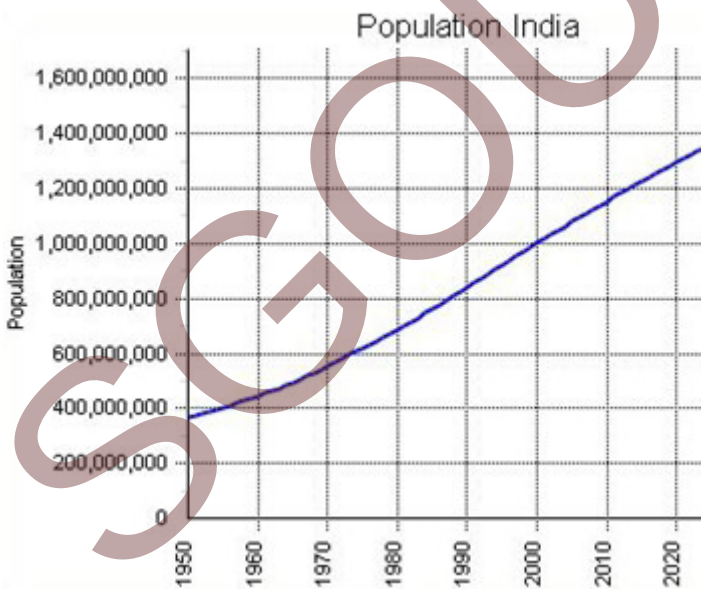


Fig. 4.1.2 Secular trend in population of India

4.1.3.2 Cyclic Component

The cyclical component in a time series represents the broader, wave-like movements in data that unfold over more extended periods. Unlike seasonal patterns, these cycles are not tied to specific calendar intervals and lack regularity. They resemble a rollercoaster, characterised by expansion and contraction phases, but without a fixed schedule.

For instance, consider the housing market. During periods of economic growth, property prices tend to rise, indicating an upward phase in the cycle. In contrast, property values may decline during recessions or economic slowdowns, reflecting a

SGOU

downward phase. Though less predictable than seasonal changes, these cyclical shifts offer valuable insights into long-term economic or business dynamics.

Recognising the cyclical component helps analysts and decision-makers understand extended fluctuations in data, enabling more informed interpretations of underlying trends and market behaviour.

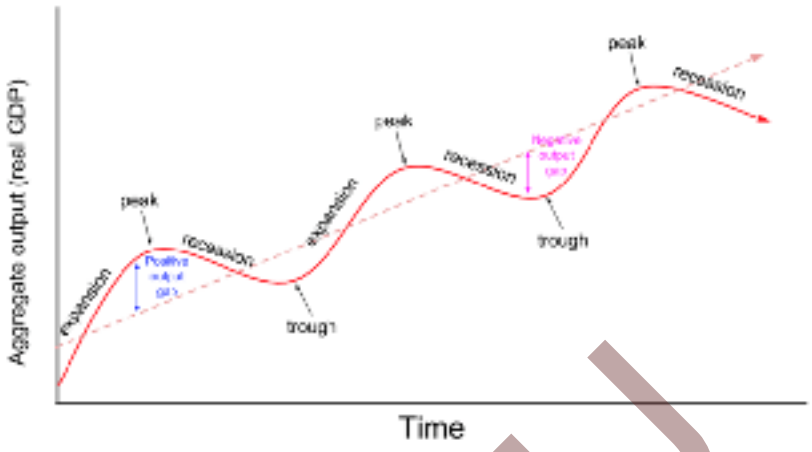


Fig. 4.1.3 Cyclic Component in Time Series

4.1.3.3 Seasonal Trend

A seasonal trend refers to a recurring pattern in data that occurs within a year, typically at regular and predictable intervals. These fluctuations are consistent over time, such as increased activity during holidays, weekends, or specific seasons, and help us anticipate changes based on historical behaviour.

For example, imagine a favourite ice cream shop experiencing higher sales every summer. This predictable rise reflects a seasonal trend. Similarly, coffee sales often surge during winter as people seek warmth and comfort. This repeated increase, like clockwork each year, illustrates how seasonal patterns shape consumer behaviour. Businesses leverage these insights to prepare in advance, such as stocking up on coffee beans, to meet the anticipated demand.

The figure below visually represents this seasonal trend, highlighting how data follows a consistent rhythm over shorter time frames.

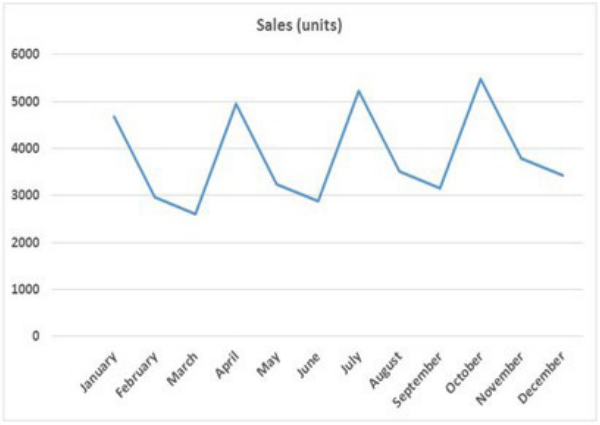


Fig. 4.1.4 Seasonal Trend in Time Series

SGOU

4.1.3.4 Irregular Component

The irregular component in a time series captures the unpredictable sudden twists that defy any consistent pattern. These are random fluctuations caused by unforeseen events or anomalies that cannot be attributed to trend, seasonality, or cycles.

For example, a sudden surge in ice cream sales triggered by an unexpected heatwave illustrates this irregular behaviour. Such variations are crucial to consider, as they reflect real-world disruptions and anomalies that influence data unexpectedly. Recognising the irregular component helps create a more realistic understanding of time series data and enables businesses to remain agile, adjusting their strategies in response to abrupt changes in market conditions or operational environments.



Fig. 4.1.5 Irregular Component in Time Series

As illustrated in the figure above, an irregular movement in a time series exhibiting an upward trend refers to a sudden and unexpected decline in observed values, such as the noticeable drop in February. This type of fluctuation represents a departure from the established growth pattern, momentarily disrupting the consistent trajectory seen over time.

Irregular movements typically arise from unforeseen influences, including economic downturns, abrupt shifts in consumer behaviour, unexpected events, or external market shocks. Recognising and analysing these anomalies is vital in time series analysis, as it enables researchers and decision-makers to explore the underlying causes of such disruptions and adapt their strategies accordingly. These deviations also underscore the importance of accounting for external variables that may impact a time series, even when the data appears to follow a stable trend.

4.1.4 Measurement of Trend

Sometimes, patterns or trends in a time series can be discerned simply through visual observation. However, as datasets grow in size and complexity, such intuitive analysis becomes less feasible. Analysts rely on established methods to systematically uncover underlying patterns and trends, especially in diverse and intricate data. In this section, we will focus on two widely used techniques for trend analysis in time series, particularly relevant in business and economic contexts.

SGOU

They are;

i. Moving Average Method and

ii. Method of Least Squares

Moving Average is an example of a graphical method of finding a trend in time series and method of least squares is an example of quantitative or mathematical trend analysis method.

4.1.4.1 Moving Average Method

The moving average method is used to smooth out fluctuations in time series data, revealing underlying trends. The process involves calculating averages for multiple sets of consecutive data points for a specific duration (like 3 years, taking Average of observations for year 1 to 3, year 2 to 4, year 3 to 5, and so on) from a time series data, or 'moving' averages, which helps in identifying patterns and trends.

This method smoothens the data, making trends more apparent and aiding in decision-making based on a more precise representation of the underlying pattern. This method is used when the trend is secular and requires high accuracy. The trend values obtained by this method are more accurate.

Simply put, the moving Average may be for three, four, five, six, or seven years. Suppose the moving Average is to be calculated for three years. We will take the Average for the first three years and place it against the middle year of the three. Similarly, the moving Average is to be calculated for five years; we will take the Average for the first five years and place it against the middle year of the five. Now, leaving the first year and adding the sixth year, take the Average of the five years. Place this Average against the middle of those five years. This way taking the Average of after leaving the first year and taking one next year i.e., if the observations are a, b, c, d, e, f and so on, then three year moving averages are (a+ b + c) /3, (b +c + d) /3, (c +d +e) /3 etc.

Five yearly moving averages are (a + b + c + d + e) /5, (b + c + d + e + f) /5 etc.

This will be much clearer once you familiarise the procedure for calculating moving averages.

a. Calculation of Moving Average

Generally, moving averages are calculated as 3 year moving Average, 5 year moving Average, etc. Let us go through the step-by-step procedure for calculating a simple moving average using a small array of data. For this example, let us say we have the following data representing monthly sales over six months:

Jan	Feb	Mar	Apr	May	June
25	30	28	35	40	32

Step 1: Decide on the period for your moving average. For simplicity, let us use a 3 month moving average.

SGOU

Step 2: Create subsets of the data based on the chosen period. For a 3-month moving average, the subsets would be:

Jan	25	
Feb	30	(25+30+28)
Mar	28	(30+28+35)
Apr	35	(28+35+40)
May	40	(35+40+32)
June	32	

Step 3: Find the Average of each subset. For the given subsets, the averages are as follows:

Jan	25		
Feb	30	(25+30+28)	(25+30+28) /3 = 27.7
Mar	28	(30+28+35)	(30+28+35)/3 = 31
Apr	35	(28+35+40)	(28+35+40)/3 = 34.3
May	40	(35+40+32)	(35+40+32)/3 = 35.7
June	32		

Step 4: Plot these averages instead of original data to observe the underlying trend in the data

Illustration 4.1.1

Give three yearly and five yearly moving average for the following series

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Sales	1	2	3	5	6	7	9	10	11	15	20

SGOU

Solution

3 Year Moving Average

Year	Sales	3 yearly moving total	3 yearly moving average
2010	1	-	-
2011	2	6	2
2012	3	10	3.33
2013	5	14	4.67
2014	6	18	6
2015	7	22	7.33
2016	9	26	8.67
2017	10	30	10
2018	11	36	12
2019	15	46	15.33
2020	20	-	-

5 Year Moving Average

Year	Sales	5 yearly moving total	5 yearly moving average
2010	1		
2011	2		
2012	3		
2013	5	17	3.4
2014	6	23	4.6
2015	7	30	6
2016	9	37	7.4
2017	10	43	8.6
2018	11	52	10.4
2019	15	65	13
2020	20		

SGOU

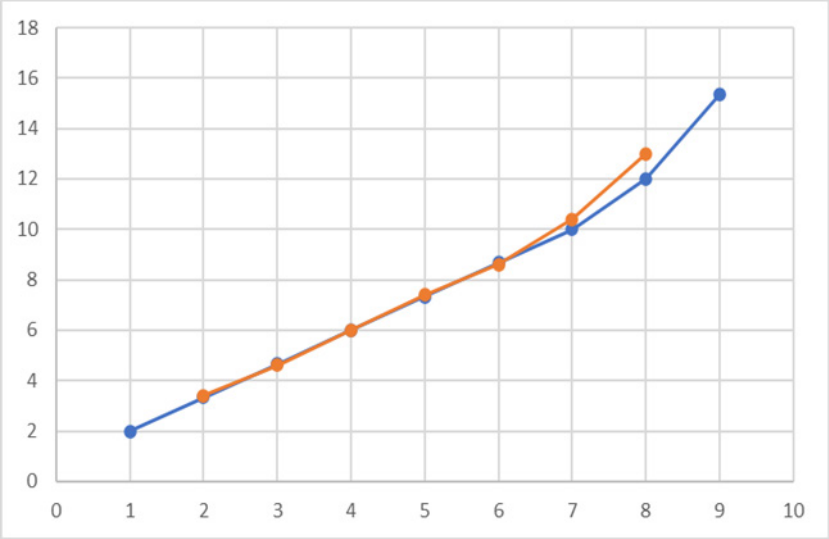


Illustration 4.1.2

Find the underlying trend in the Sales data for 10 years using 5 year moving Average.

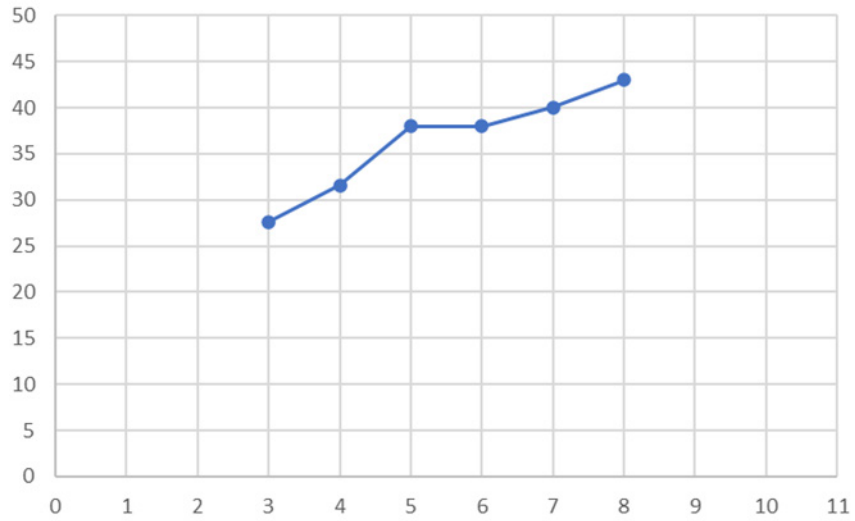
Year	1	2	3	4	5	6	7	8	9	10
Yearly Sales	20	25	30	28	35	40	45	42	38	50

Solution

5 Year Moving Average

Year	5 yearly moving sum	5 yearly moving average
1		
2		
3	138	27.6
4	158	31.6
5	178	38
6	190	38
7	200	40
8	215	43
9		
10		

SGOU



b. Method for calculating the Moving Average for an even number of years

If the number of years is even, for example, to calculate a year's moving Average, calculate the Average of the first four years. Put this value between the second and third years, i.e., middle of the first four years. Now, leave the first year, calculate the Average of the next four years and place it in the middle of these four years and so on. Then, calculate the Average of the two moving averages already calculated, taking first and second, second and third, etc. This Average is called a moving average centred. We will place the first average centred against the middle of the two moving averages. This Average will be against the third year of the original data. In this way, we calculate averages centred on the other years. The centred moving averages will be the trend values.

Illustration 4.1.3

Give 4 yearly moving average for the following series. Also, plot the given values and trend values on a graph

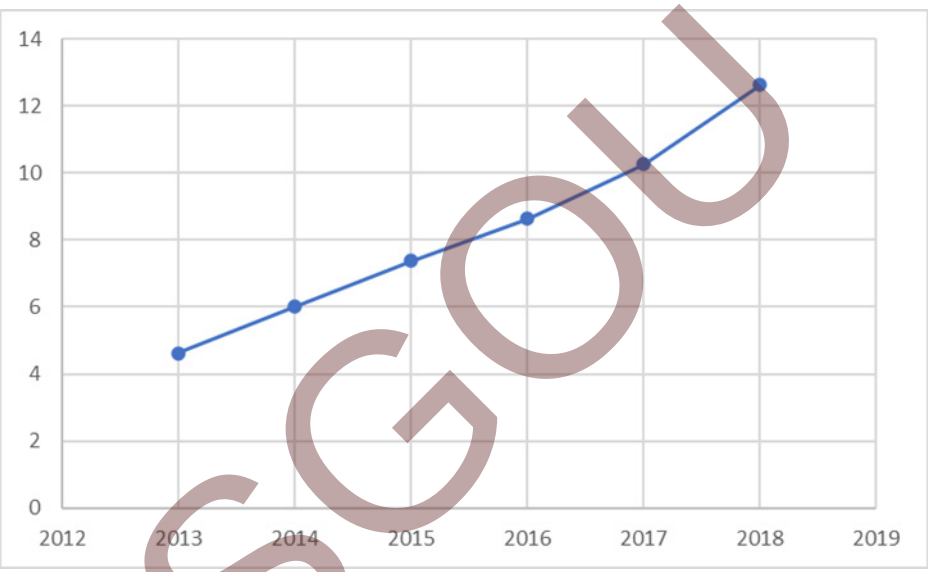
Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Sales	2	3	5	6	7	9	10	11	15	20

Solution

Year	Sales	Total of 4 years	4 yearly moving average	4 yearly moving average centered
2011	2			
2012	3			
		16	4	
2013	5			4.625
		21	5.25	

SGOU

2014	6			6
		27	6.75	
2015	7			7.375
		32	8	
2016	9			8.625
		37	9.25	
2017	10			10.25
		45	11.25	
2018	11			12.625
		56	14	
2019	15			
2020	20			



SGOU

Illustration 4.1.4

Workout centred four year moving average for the following data and plot them in graph.

Year	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968
Data	1102	1250	1180	1340	1212	1317	1452	1549	1586	1476	1614	1586

Solution

Year	Data	4 yearly totals	4 yearly moving averages	Centered average
1957	1102	-	-	
1958	1250	-	-	
1959	1180	4872	1218	1231.75
1960	1340	4982	1245.5	1253.85
1961	1212	5049	1262.2	1296.2
1962	1317	5321	1330.2	1356.35
1963	1452	5530	1382.5	1429.85
1964	1549	5904	1476	1495.85
1965	1586	6063	1515.7	1535.95
1966	1476	6225	1556.2	1560.85
1967	1614	6262	1565.5	-
1968	1586	-	-	-

4.1.4.2 Method of Least Squares

The least squares method is a statistical technique commonly employed in time series analysis to identify and quantify trends. Its primary objective is to determine the best-fitting straight line, known as the regression line, through a series of data points by minimising the sum of the squared vertical deviations (residuals) between the observed values and the line. This approach offers a systematic means of estimating the parameters of a linear trend, making it a valuable tool for forecasting and interpreting long-term movements in data.

At the core of this method lies the principle of least squares, which seeks to minimise the total squared differences between actual data points and the estimated trend line, thereby ensuring the most accurate representation of the data. This process, referred to as curve fitting, involves deriving a mathematical function that best captures the underlying pattern in the dataset.

By reducing the discrepancies between observed and predicted values, the least squares method provides a reliable framework for parameter estimation across various domains, including time series analysis, regression modelling, and economic forecasting. Its precision supports robust decision-making and enhances pattern recognition in scientific research, business analytics, and policy evaluation.

The principle of least squares provides us with a mathematical or analytical derivation to obtain a mathematical curve which will be fit to the given series. The technique of obtaining this mathematical curve by the principle of least squares is known as Curve fitting.

The principle of least squares states that the sum of the squares of the deviation between the observed values and trend values is the least. The techniques can be used to fit linear as well as non-linear trends. A linear trend is one which gives a straight line

SGOU

when plotted on a graph paper, whilst a non-linear trend is one which gives non-linear curves like parabola, exponential curve, logistics curve, etc.

Line of Best Fit

A straight trend line can be fitted to time series data by the method of curve fitting, depending on the most popular method of least squares. Such a straight line is also known as the line of best fit.

This method is an algebraic device and is widely used. It gives a straight line best suited to the given data when plotted on a graph. The total of the vertical distance from various points on one side from this line will be equal to the total of the vertical distance from various points on the other side of the straight line. When the sum of deviation on either side is equal to zero, the sum of the squares of these deviation will be least. Let the line of the best fit be described by the equation of the type, $y = a + bx$, , where a and b are two unknown constants whose values are to be determined.

To find a and b , we apply the method of least squares. The following two equations are formed and solved:

$$\sum y = na + b\sum x$$

$$\Sigma xy = a\,\Sigma\, x + b\Sigma\, x^2$$

These two equations are known as normal equations.

Solving these two equations we get the values of a and b and substituting them in $y=a+bx$ we get the equation of the line of best fit.

Illustration 4.1.5

A firm's sales (in tonnes) for 2015 to 2021 are given below. Estimate the sales for 2024 using the trend.

Years	2015	2016	2017	2018	2019	2020	2021
Sales (in tonnes)	5	6	8	6	10	10	11

Solution

$n = 7$, shift the origin t to x by $x = (t - 2018)/1 = t - 2018$

Year t	Sales(y)	x	x^2	xy	Trend $Yc = 8 + 1x$
2015	5	-3	9	-15	5
2016	6	-2	4	-12	6
2017	8	-1	1	-8	7
2018	6	0	0	0	8

SGOU

2019	10	1	1	10	9
2020	10	2	4	20	10
2021	11	3	9	33	11
n = 7	$\sum y = 56$		$\Sigma x^2 = 28$	$\sum xy = 28$	56

$$a = (\sum y)/n$$

$$= 56/7$$

$$= 8$$

$$b = (xy) / \Sigma x^2$$

$$= 28/28 = 1$$

The straight line is

$$Y_c = 8 + x$$

Estimation of sales in the year $t = 2024$, the x value is $2024 - 2018 = 6$

$$Y(2024) = 8 + 6$$

$$= 14$$

Illustration 4.1.6

Below are given the figures of production (in thousand tons) of a sugar factory :

Year	1999	2000	2001	2002	2003	2004	2005
Production	77	88	94	85	91	98	90

i. Fit a straight line by the method of 'least squares' and show the trend values.

ii. What is the monthly increase in production?

Solution

$n = 7$, shift the origin t to x by $x = (t - 2002)/1 = t - 2002$

Year t	Production y	x	x^2	xy	Trend $Yc = 89 + 2x$
1999	77	-3	9	-231	83
2000	88	-2	4	-176	85
2001	94	-1	1	-94	87
2002	85	0	0	0	89

SGOU

2003	91	1	1	91	91
2004	98	2	4	198	93
2005	90	3	9	270	95
n = 7	$\sum y = 56$	$\sum x = 0$	$\Sigma x^2 = 28$	$\sum xy = 56$	623

The straight line trend of y on x is $y = a + bx$. Since $\sum x = 0$

$$a = (\sum y)/n$$

$$= 623/7$$

$$= 89$$

$$b = (xy) / \Sigma x^2$$

$$= 56/28$$

$$= 2$$

The straight line is

$$Yc = 89 + 2x$$

From the equation, the trend values increase by a constant amount 'b' units every year. Thus, the yearly increase in production is 'b' units, i.e., 2 × 1000 = 2000 tons. Hence, the monthly increase in production = 2000/ 12 = 166·67 tons

Illustration 4.1.7

Compute the trend value by the method of least square from the data given below, also estimate the number of sheep in 2009.

Year	2000	2001	2002	2003	2004	2005	2006	2007
No. of sheep (000)	56	55	51	47	42	38	35	32

Solution

When there is a middle year, take the average of the two middle years as the assumed average and take the deviation from it. Multiply by two to avoid decimals. Take deviation from 2003.5 and multiply by 2.

Middle years: 2003 and 2004.

Average: $= (2003 + 2004)/2 = 2003.5$

Year	No. of sheep	x = 2(t -2003.5)	xy	x^2
2000	56	-7	-392	49
2001	55	-5	-275	25

SGOU

2002	51	-3	-153	9
2003	47	-1	-47	1
2004	42	1	42	1
2005	38	3	114	9
2006	35	5	175	25
2007	32	7	225	49
n = 8	356	0	-312	168

The trend equation is

$$y = a + bx$$

$$\sum y = na + b\sum x$$

$$\Sigma xy = a\,\Sigma\, x + b\Sigma\, x^2$$

Here n=8

Equation 1:

$$356 = 8 \times a + b \times 0$$

$$8a = 356$$

$$a = 356/8 = 44.5$$

Equation 2:

$$-312 = 44.5 \times 0 + b \times 168$$

$$168b = -312$$

$$b = (-312)/168 = -1.85$$

Year	***x***	**y = 44.5 – 1.85 *x***
2000	-7	44.5 – 1.85 x -7 = 57.45
2001	-5	44.5 – 1.85 x -5 = 53.75
2002	-3	44.5 – 1.85 x -3 = 50.05
2003	-1	44.5 – 1.85 x -1 = 46.35
2004	1	44.5 – 1.85 x 1 = 42.65
2005	3	44.5 – 1.85 x 3 = 38.95
2006	5	44.5 – 1.85 x 5 = 35.25
2007	7	44.5 – 1.85 x 7 = 31.55

Number of sheep at 2009 = 2 (2009 -2003.50 = 11

Trend Equation = 44.5 -1.85 x 11

SGOU

= 24.15

Number of sheep = 24.15 x 1000 = 24,150

4.1.5 Application of Time Series in Business and Management

Time series analysis is a powerful analytical tool in business and management, enabling organisations to uncover patterns, monitor trends, and interpret behaviours over time. Its applications span various operational and strategic areas, offering data-driven insights that support effective decision-making. Below are key areas where time series analysis proves particularly valuable:

i. **Sales Forecasting:** Businesses can anticipate future sales volumes by analysing historical sales data. This supports efficient inventory control, production scheduling, and resource allocation.

ii. **Financial Market Evaluation:** Investors and analysts apply time series techniques to study movements in stock prices, exchange rates, and other financial metrics. Recognising past trends aids in shaping sound investment and financial strategies.

iii. **Demand Forecasting:** Time series analysis helps predict customer demand, allowing firms to fine-tune inventory levels, streamline supply chains, and prevent shortages or excess stock.

iv. **Monitoring Employee Performance:** Organisations can track productivity trends over time, assess the impact of interventions, and make informed decisions regarding workforce planning and development.

v. **Quality Assurance and Process Optimisation:** In manufacturing and operations, time series is used to monitor product quality and detect patterns in defects or inefficiencies, leading to targeted improvements.

vi. **Evaluating Marketing Impact:** Businesses assess marketing campaigns' performance by examining sales or customer engagement fluctuations over time, enabling refinement of promotional strategies.

vii. **Customer Behaviour Insights:** Time series analysis reveals purchasing habits, peak activity periods, and evolving customer preferences, helping businesses tailor their offerings and engagement strategies.

viii. **Risk Analysis and Mitigation:** Historical data identifies potential risks and simulates future scenarios, supporting proactive risk management across sectors.

ix. **Supply Chain Optimisation:** Time series techniques assist in forecasting demand, identifying seasonal patterns, and adapting supply chain operations to dynamic market conditions.

SGOU

x. **Budgeting and Resource Planning:** Organisations use time series to project future expenditures and allocate resources effectively, ensuring financial discipline and operational efficiency.

In conclusion, time series analysis is a versatile and essential tool in business and management. By leveraging historical data, organisations can enhance forecasting accuracy, streamline processes, and remain responsive in a rapidly evolving marketplace.



Summarised Overview

Time Series Analysis involves the systematic study of data points collected or recorded at successive intervals over time, enabling the identification of patterns, trends, and fluctuations. Its utility lies in forecasting, strategic planning, and decision-making across finance, economics, and business operations. The key components of time series,trend, seasonality, cyclical variation, and irregular fluctuations,help in understanding data's underlying structure and behaviour. Techniques like the moving average method and the method of least squares are commonly used to measure the long-term trend; the former smooths short-term variations to reveal the overall direction, while the latter fits a regression line by minimising deviations. In business and management, time series analysis finds wide application in sales forecasting, demand planning, financial analysis, marketing evaluation, supply chain optimisation, and risk management, offering actionable insights for improving performance and anticipating future developments.



Self-Assessment Question

1. What is the central aim of conducting time series analysis?
2. How can a time series be defined within the framework of data analysis?
3. Why is analysing time series data valuable for businesses and data professionals?
4. What are the key elements of a time series, and how would you briefly describe each?
5. In which real-world situations can time series analysis support effective decision-making?
6. Describe a practical example that illustrates cyclical variation in time series data.

SGOU

7. What does the moving average method entail in time series analysis?
8. Given monthly sales data spanning two years, outline the procedure for computing a three-month moving average.
9. What is the core statistical concept behind the least squares method in time series analysis?
10. Using a set of data points, explain how to derive the least squares regression equation.
11. How do the moving average method and the least squares method differ, and in what ways are they similar?



Assignments

1. Present a real-life case where time series analysis proves helpful. How does this approach differ from other analytical techniques?
2. Describe a business or economic scenario where time series analysis could improve decision-making. In what ways might it enhance forecasting or strategic planning?
3. Identify and explain the trend, seasonal patterns, and any irregular or random variations for a given dataset. How does recognising these components improve data interpretation?
4. Suppose you have monthly sales data for a retail outlet over several years. How would you detect and examine the trend and seasonality? How could this insight benefit store management?
5. What are some common difficulties encountered when working with time series data? How can issues like outliers or missing values affect the reliability of analysis, and what methods can be used to overcome these challenges?
6. Interpret the resulting trends after applying the moving average and least squares methods to a time series dataset. How can these insights support business decision-making or contribute to forecasting future developments?
7. Suppose you are responsible for analysing a time series of monthly stock prices. How would you employ the moving Average and least squares methods to detect underlying trends and identify possible turning points? What specific factors should be considered when conducting such an analysis in a financial setting?

SGOU

8. Apply the method of semi-average for determining the trend:

Year	2001	2002	2003	2004	2005	2006	2007
Value	12	15	20	18	25	24	28

(Ans:- 13,15,17,20,23,25,26)

9. From the following series of annual data find the trend values by using 3 year moving average

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Value	170	231	261	264	278	302	289	297	340	273

(Ans:- 220.67, 252, 267.67, 281.33, 289.69, 296, 308.67,303.33)

10. Fit a trend to the following data by taking the average length of four.

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Production (in millions)	30	33	28	35	33	40	44	47	46	52	55

(Ans: -,-,31.875,33.125,36,39.5,42.625,45.75,48.625,-,-)

11. Determine the period of the moving average for the following data and calculate moving average for that period.

Year	Values
1	130
2	127
3	124
4	135
5	140
6	132
7	129
8	127
9	145
10	158
11	153
12	146
13	145
14	164
15	170

SGOU

(Ans: -,-, 130.25, 132.25, 133.375, 133,132,625, 136.5, 142.75, 148.125, 150.5, 151.25, 154.125,-,-)

12. Fit a straight-line trend equation to the following data by the method of least squares. Find the trend values. Estimate sales for 2009 on the following data.

Year	2004	2005	2006	2007	2008
Sales ('000)	10	12	14	10	15

(Ans:- y = 12.2 + 0.8x, Trend values: 10.6, 11.4,12.2,13,13.8.

Sale for 2009 = 4.6)

13. Fit a straight-line trend by the method of least squares and find the trend values.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008
Value	68	72	73	80	90	92	90	110	114

(Ans:- 87.67 + 5.73x.

Trend values are 64.75,70.48,76.21,81.94,87.67,93.4,99.73,104.86,110.59)

Reference

1. Aggarwal, B. M. (2010). *Business statistics*. Ane Books Pvt. Ltd.
2. Chandrasekharan, N., & Umaparvathi, M. (2016). *Statistics for management*. PHI Learning Pvt. Ltd.
3. Gupta, M., & Khanna, R. (2011). *Quantitative techniques for decision making*. Prentice Hall of India.
4. Gupta, S. P. (2021). *Statistical methods*. S. Chand & Sons.
5. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International (P) Ltd.
6. Levin, R. I., & Rubin, D. S. (1999). *Quantitative approaches for management*. Pearson Education.
7. Sharma, A. (2017). *Quantitative techniques for decision making*. Himalaya Publishing House.
8. Vohra, D. V. D. (2017). *Quantitative techniques for management*. New Delhi: McGraw Hill Education

SGOU

Suggested Reading

1. Arora, P. N., Arora, S., & Arora, S. (2010). *Comprehensive statistical methods*. S. Chand & Company Pvt. Ltd.
2. Beri, G. C. (2017). *Business statistics*. Tata McGraw-Hill.
3. Chawla, D., & Sondhi, N. (2016). *Research methodology: Concepts and cases*. Vikas Publishing House.
4. Krishnaswami, O. R., Ranganathan, M., & Harikumar, P. N. (2016). *Research methodology*. Himalaya Publishing House.
5. Render, B. (2022). *Quantitative analysis for management*. Prentice Hall of India.
6. Singh, A. K. (2017). *Tests, measurements and research methods in behavioural sciences*. Bharti Bhawan.

SGOU

Unit 2

Index Numbers

Learning Outcomes

On completion of this unit, learners will be able to:

- ◊ explain the meaning, significance, and applications of index numbers in economic and business contexts
- ◊ make them aware of the common problems in constructing index numbers, including base year selection, item selection, weighting, and data reliability
- ◊ differentiate between simple and weighted index number methods, and justify using weights to reflect the relative importance of items in composite indices
- ◊ demonstrate proficiency in constructing index numbers using aggregative and average price relative methods
- ◊ compute index numbers using Laspeyres, Paasche, Fisher's Ideal, Bowley's, Dorbish, and Marshall-Edgeworth methods

Background

The study of index numbers traces its origins to the early 18th century, when economists and statisticians began seeking systematic ways to measure changes in economic variables over time. The earliest known use of index numbers was in tracking price movements, notably by William Fleetwood in 1707, who compiled price data to assess changes in the cost of living for Oxford students. However, during the 19th and early 20th centuries, the theoretical foundation of index numbers was formalised, with significant contributions from economists such as Étienne Laspeyres and Hermann Paasche. Their respective methods, Laspeyres' using base-period weights and Paasche's using current-period weights, laid the groundwork for modern price and quantity indices.

SGOU

As industrial economies expanded and statistical thinking matured, the need for more refined and unbiased measures led to the development of composite indices. Irving Fisher's Ideal Index, introduced in the early 20th century, combined the Laspeyres and Paasche indices geometrically to address their limitations. Subsequent refinements by Bowley, Dorbish, and Marshall-Edgeworth introduced alternative formulations that balanced simplicity with statistical rigour. These methods have become integral to national income accounting, inflation measurement, and business analytics. The evolution of index number theory reflects a broader shift towards empirical precision in economic decision-making and policy formulation. This unit sheds light on the conceptual foundations, methodological approaches, and practical applications of index numbers in quantitative analysis.



Keywords

Index Numbers, Weighted Index, Aggregative Method, Price Relative Method, Laspeyres Method, Paasche Method, Fisher's Ideal Index



Discussion

Index numbers are statistical measures that track changes in variables such as prices, quantities, or values over time, relative to a chosen base period. They simplify complex data sets into a single figure that reflects overall trends, making them essential tools in economics and business analysis. For example, suppose the cost of a basic basket of goods (rice, milk, and vegetables) was ₹1,000 in 2020, and in 2025 the same basket costs ₹1,250. The price index number for 2025, with 2020 as the base year, would be 125. This means prices have risen by 25% since 2020, indicating inflation in the economy. Such indices help businesses monitor growth, assess inflationary impacts, and make informed pricing, inventory, and strategic planning decisions.

4.2.1 Meaning of Index Numbers`

Index numbers rank among economic and business analysis's most extensively applied statistical tools. They reveal whether prices fluctuate, industrial output is increasing or declining, or sales figures are improving or deteriorating compared to earlier periods. As vital indicators of economic health, index numbers help detect inflationary pressures or deflationary trends.

SGOU

Fundamentally, index numbers capture relative shifts, typically as ratios or comparative figures, in key economic variables such as output, wages, prices, and employment across time. Beyond temporal comparisons, they can also reflect spatial or categorical variations, such as crime rates across regions or traffic accident frequency. Household characteristics, like the Number of rooms in a residence, may act as proxies for economic well-being. Thus, index numbers are instruments for quantifying changes in a phenomenon over time, across locations, or by other distinguishing features. These phenomena may include commodity prices, quantities produced or consumed, or abstract attributes like intelligence, attractiveness, or performance.

Although index numbers are often associated with tracking price changes of individual goods, their practical application typically involves monitoring variations across groups of related items or data series. These may pertain to production levels, price trends, employment statistics, or cost of living. Index numbers are regarded as economic barometers in this broader context, offering insights into the overall direction and momentum.

4.2.2 Definitions

An index number can be described as a statistical tool that consolidates the fluctuations across multiple data series into a single value, effectively capturing the overall impact of changes in the individual components.

According to ***Spiegal,*** *"An index number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographical location or some other characteristic".*

According to ***Wheldon****, "An index number is a statistical device for indicating the relative movements of the data where measurement of actual movements is difficult or incapable of being made."*

According to ***Croxton and Cowden,*** *"Index Numbers are devices for measuring differences in the magnitude of a group of related variables."*

According **to Dr. A.L. Brole,** *"Index Numbers are used to measure the changes in some quantity which we cannot observe directly."*

According to ***John. I. Griffin***, "An index number is a quantity which by reference to a base period shows by its variation, the changes in the magnitudes over a period of time."

According to ***Edge Worth,*** *"Index Number shows by its variation the changes in a magnitude which is not susceptible of either accurate measurement in itself or of direct variation in practice".*

SGOU

4.2.3 Importance of Index Numbers

a. Expanded Applications of Index Numbers

Initially developed to monitor shifts in commodity prices, index numbers have evolved into versatile tools with wide-ranging applications. Today, they are employed to measure changes in physical output, production volumes, trade activity, service costs, and fluctuations in share and stock prices. Their relevance spans multiple disciplines, including economics, finance, statistics, and the social sciences, where they serve as indispensable instruments for data analysis, trend evaluation, and evidence-based decision-making.

b. Role in Economic Analysis

Index numbers are fundamental in tracking key economic indicators such as inflation, growth, and employment. For instance, the Consumer Price Index (CPI) reflects changes in the average cost of goods and services, enabling policymakers to monitor inflation and adjust monetary strategies accordingly. Similarly, GDP-related indices offer insights into national economic performance and development trajectories.

c. Temporal and Spatial Comparisons

These statistical measures facilitate consistent comparisons across periods, geographic regions, and variables. Index numbers help analysts, decision-makers, and businesses identify patterns, evaluate disparities, and respond strategically to emerging trends by providing a standardised reference point.

d. Assessing Development Levels

Index numbers are instrumental in evaluating and comparing developmental progress across nations or regions. For example, the Human Development Index (HDI) integrates life expectancy, educational attainment, and income levels to gauge societal advancement. Governments and institutions use such indices to pinpoint developmental gaps, allocate resources, and monitor long-term improvements.

e. Applications in Social Science Research

In social sciences, index numbers are widely used to quantify subjective phenomena through surveys and empirical studies. They enable researchers to analyse large datasets by converting qualitative attributes, like satisfaction, happiness, or well-being, into measurable indicators. This facilitates the identification of behavioural patterns, social correlations, and broader societal trends.

f. Insights for Business and Marketing

Businesses leverage index numbers to evaluate customer satisfaction, brand engagement, and market dynamics. These indices offer actionable insights that guide product enhancement, refine marketing strategies, and elevate consumer experiences. Index numbers support informed, data-driven business decisions by translating complex market data into digestible metrics.

SGOU

4.2.4 Problems in the Construction of Index Numbers

The construction of index numbers involves several key challenges, which include:

i. Defining the Objective

Index numbers are not universally applicable; each is designed for a specific purpose. Therefore, it is essential to clearly and precisely define the objective before constructing an index number. As Haberler aptly notes,

"Different index numbers are constructed to fulfil different objectives, and before setting out to construct a particular index number, one must clearly define the objective of the study, because it is on this basis that the nature and format of the index number depend."

ii. Selection of the Base Period

Choosing an appropriate base period is a critical step in constructing index numbers. The base period is a reference point against which changes in other periods are measured.

Morris Hamburg observes:

"It is desirable that the base period be not too far removed from the present. The further away we move from the base period, the dimmer our recollections of the economic conditions prevailing at that time. Consequently, comparisons with such remote periods tend to lose significance and become tenuous in meaning."

George Simpson and Fritz Kafka add:

"Since practical decisions are made regarding index numbers, and economic practices are often short-term, we prefer to make comparisons using a base that lies within the same general economic framework as the years of immediate interest. Therefore, we choose a base relatively close to the years being studied."

In today's rapidly changing world, the base year should ideally not be over a decade old. There is also a psychological rationale for selecting a recent base period. T. W. Lewis and R. A. Fox explain:

"A fairly recent base is advantageous simply because it keeps the values of the index close to the 100 mark, thereby aiding comprehension of changes. If, for example, the base year is too far in the past, current values might be around 400, 500 or 600. If two consecutive index values are 573 and 604, further calculation is required to appreciate the magnitude of the increase. Moreover, a psychological effect arises an index at 604 may give the impression that prices are excessively high. A recent base also ensures comparisons are based on relatively homogeneous quantities."

iii. Selection of Items (Regimen or Basket)

Including all commodities in an index number is neither feasible nor necessary. Each index is designed to measure changes within a specific group. The selection of items depends on the intended purpose of the index and should reflect widely consumed goods.

SGOU

Chosen items must be relevant, representative, reliable, and comparable. Interestingly, increasing the Number of items may reduce the index's reliability.

iv. Selection of Data Sources and Collection Methods

Data collection primarily concerns price information. Whether wholesale or retail prices are used depends on the study's objective. Price quotations should be sourced from key markets. For more accurate results, it is advisable to use a standard price that represents the commodity over the entire period under consideration.

v. Selection of an Appropriate Average

Since index numbers measure relative changes, the geometric Mean is theoretically the most suitable average. However, the arithmetic mean is commonly used in practice due to computational complexities.

vi. Weighting System

John I. Giffin explains:

"In simple terms, weighting is designed to give component series an importance in proper relation to their real significance."

A suitable weighting system should be employed to ensure that each commodity exerts an appropriate influence on the index. In unweighted index numbers, all commodities are treated equally, which may not reflect their economic impact.

For example, if the Price of wheat doubles while the Price of tobacco halves, the index may show no change. However, the rise in wheat prices affects consumers more significantly than the reduction in tobacco prices. This underscores the importance of assigning weights to different items.

Weights may be based on:

- ◊ The value or quantity produced
- ◊ The value or quantity consumed
- ◊ The value or quantity sold

For general index numbers, weights are typically based on production. For cost-of-living indices, consumption-based weights are more appropriate. Depending on the context, weights may be applied using Paasche's, Laspeyre's, or Fisher's methods.

4.2.5 Characteristics of Index Numbers

Index numbers possess several defining features that make them valuable statistical and economic analysis tools. These include:

SGOU

1. Index Numbers as Specialised Averages

Index numbers represent a unique average, typically a weighted average derived from a sample of items considered representative of a broader group. As L. R. Connor explains:

"In its simplest form, an index number represents a special case of an average, generally a weighted average compiled from a sample of items judged to be representative of the whole."

Unlike simple averages involving data measured in uniform units, index numbers aggregate variables expressed in diverse units. For instance, an index may include food, clothing, rent, and cooking gas, each measured differently, yet yield a meaningful composite measure.

2. Index Numbers as Measures of Relative Change

Index numbers quantify relative changes over time and are typically expressed in percentage terms, although the percentage symbol is not explicitly used. These changes are measured concerning a specific point in time known as the Base Period or Base Year.

For example, if the index number for the year 2000 is 150, based on a base value of 100 in 1995, this indicates a 50% increase in prices over those five years. Thus, index numbers serve as numerical indicators of relative change.

3. Basis for Comparison

Index numbers facilitate comparisons across different periods by referencing a standard base period. This allows analysts to assess trends, fluctuations, and patterns in economic variables over time.

4. Broad Applicability

Index numbers have wide-ranging applications across various sectors of the economy. They are used to track and analyse changes in:

- ◊ Price levels
- ◊ Standards of living
- ◊ Industrial and agricultural output
- ◊ National and per capita income
- ◊ Export and import volumes

Their versatility makes them indispensable tools for economists, policymakers, and researchers.

SGOU

4.2.6 Advantages or Uses of Index Numbers

Index numbers serve various analytical and practical functions across economic, business, and policy domains. Their key advantages include:

1. Measuring Changes in Price Levels and the Value of Money

One of the most significant uses of index numbers is to track changes in the value of money over time. They help assess the impact of inflation or deflation on different segments of society and support the formulation of corrective measures to stabilise the economy.

2. Assessing Changes in Living Standards

Index numbers provide insights into shifts in the standard of living. While nominal incomes may rise, a decline in the value of money as indicated by index numbers can lead to reduced real income and living standards. Thus, they are vital for evaluating changes in purchasing power.

3. Adjusting Salaries and Allowances

The cost-of-living index is crucial for governments and employers to revise wages and allowances. When the index rises, corresponding adjustments help workers maintain their standard of living. Proportional increases ensure that inflation does not erode real income.

4. Supporting Business Decision-Making

Price index numbers assist the business community in analysing market trends and economic conditions. They inform production, pricing, inventory management, and strategic planning decisions.

5. Monitoring Production Trends

Production index numbers reveal whether agricultural and industrial output is increasing or declining. These indicators are essential for evaluating sectoral performance and planning resource allocation.

6. Analysing Foreign Trade

Export and import index numbers offer valuable data on trade performance. They guide the formulation of trade policies and help assess the competitiveness of domestic industries in global markets.

7. Informing Government Policy

Governments rely on index numbers to shape monetary and fiscal policies. These indicators support the development of programmes to enhance investment, output, income, employment, trade, and consumption.

SGOU

8. Assisting Political Analysis

Politicians use index numbers to understand the economic realities of the country. This enables them to offer informed critiques and propose constructive policy alternatives.

9. Facilitating Contract Escalation

Historical trends in index numbers are used to adjust prices in long-term contracts. This ensures fairness and reflects changes in market conditions over time.

10. Indicating Market Equilibrium

Index numbers help assess the balance between aggregate supply and demand, particularly at the primary market level. This information is crucial for maintaining price stability.

11. Aiding Planning and Production

Index numbers are employed in planning plant construction, scheduling production, procuring raw materials, managing inventories, estimating costs, and guiding investment strategies.

12. Enabling Forecasting

By analysing historical price movements, index numbers support pricing, marketing, and sales forecasting. They help anticipate future trends and prepare accordingly.

13. Examining Commodity Relationships

Index numbers also study the evolving relationships among different commodities, offering insights into substitution effects, consumption patterns, and market dynamics.

Why are Index Numbers called Specialized Averages?

Index numbers are called specialized averages because they condense diverse, heterogeneous data (like prices of different goods measured in different units) into a single representative figure that shows relative change over time or across places. Unlike simple averages (mean, median, mode), index numbers deal with data expressed in different units (e.g., rice in kilograms, milk in litres, cloth in metres). To make them comparable, prices or quantities are converted into relatives (percentages or ratios) and then averaged. They are designed to show relative changes rather than absolute values. For example, a price index of 120 means prices have increased by 20% compared to the base year. Index numbers can be weighted or unweighted, depending on the importance of items in the basket. This makes them more refined than ordinary averages.

SGOU

Key Features

Table 4.2.1 Key features of Index Numbers

Aspect	Ordinary Average	Index Number (Specialized Average)
Units	Same unit required	Different units can be combined
Purpose	Shows central tendency	Shows relative change over time/place
Method	Mean, median, mode	Ratio/percentage-based averages
Application	General statistics	Economics, business, social studies

Index numbers are specialized averages because they transform diverse data into a comparable scale, then average those relatives to reflect overall change. This makes them indispensable in economics and business for measuring inflation, cost of living, trade, and production trends.

Why are Index Numbers Referred to as Economic Barometers?

Index numbers are called economic barometers because they act like measuring instruments that indicate the overall health and direction of an economy. Just as a barometer measures atmospheric pressure to forecast weather changes, index numbers measure changes in economic variables, such as prices, production, employment, or trade, to forecast and understand economic trends. For example, a rising price index signals inflation, while a falling industrial production index may indicate a slowdown in economic activity. Policymakers, businesses, and researchers rely on these indices to gauge performance, make decisions, and plan for the future. In this way, index numbers serve as vital tools for monitoring and predicting the "climate" of the economy.

4.2.7 Limitations of Index Numbers

According to Coulbourn:

"The theoretical defects of index numbers are well known and are not likely to be remedied soon. Nevertheless, they are not likely to be remedied unless we continue to construct and use index numbers."

Despite their usefulness, index numbers have several limitations:

1. Incomplete Coverage of Commodities

Index numbers typically measure price changes for a limited set of predefined commodities. They do not represent the overall price level or the full spectrum of purchasing power. Many categories of goods significantly influencing the general price level are excluded.

2. Exclusion of Retail Transactions

Most index numbers are based on wholesale prices rather than retail prices. Since retail prices reflect what consumers pay, relying solely on wholesale data may result in a distorted view of real-world price fluctuations.

SGOU

3. Challenges in International Comparisons

There is no globally standardised base period for constructing index numbers. Different countries use varying base years, which makes cross-national comparisons inconsistent and unreliable.

4. Influence of Changing Consumption Patterns

Over time, consumer preferences, habits, and purchasing capacities evolve. Index numbers based on outdated consumption data may fail to capture current economic realities, reducing their relevance and accuracy.

5. Limited Scope of Application

Index numbers are developed for specific purposes and target groups. An index designed to reflect the economic conditions of one demographic may not apply to another. Similarly, an index constructed for one objective may not be suitable for alternative uses.

6. Lack of Scientific Weighting

No universally accepted scientific method for assigning weights in most index number constructions exists. As a result, the conclusions drawn from such indices may lack precision and objectivity.

4.2.8 Types of Index Numbers

Index numbers are statistical instruments designed to capture and compare changes in both economic and non-economic variables over time. Widely applied in fields such as economics and finance, they help quantify relative prices, quantities, and value shifts. Different index numbers serve distinct analytical purposes; each tailored to specific change dimensions.

4.2.9 Price Index Number

Price Index Numbers are the most commonly used indices, designed to measure average changes in price levels over time or across locations. One can assess relative price movements by comparing index values between two periods or places. Developed initially to track cost-of-living changes and guide wage adjustments, price indices are key tools for forecasting inflation and comparing price levels across regions or countries. They express current prices relative to a fixed base period.

Price index = (Current year price / Base year price) x 100

$$P_{01} = \frac{P_1}{P_0} \times 100$$

Here, P_{01} is the current year's price index based on the base year price of a given commodity, and 100 is the common denominator.

SGOU

4.2.9.1 Types of Price Index

i. Wholesale Price Index Number (WPI)

The Wholesale Price Index (WPI) is pivotal in tracking wholesale price movements. It reflects price changes in goods traded between businesses, not consumer purchases, and is widely used to:

- ◊ Measure inflation and deflate nominal macroeconomic indicators like GDP
- ◊ Inform government policies on trade, taxation, and fiscal management
- ◊ Support price escalation clauses in long-term contracts for raw materials, machinery, and construction
- ◊ Assess supply-demand dynamics in manufacturing and industrial sectors

Economists and industry professionals recognise WPI as a reliable benchmark for wholesale price adjustments and economic trend analysis.

ii. Retail Price Index

Retail Price Indices capture changes in the prices of consumer goods, financial assets (like stocks, bonds, and deposits), and other commodities at the retail level. In India, these indices are compiled by the Labour Bureau under the Ministry of Labour, covering urban and rural centres.

A key variant of Retail Price Index is the Consumer Price Index (CPI), also known as the Cost-of-Living Index, which measures how price changes affect the purchasing power of specific social groups, such as labourers, industrial and agricultural workers, and low- to middle-income households. CPI data in India is available for:

- ◊ Central Government employees
- ◊ Middle-class consumers
- ◊ The working class

4.2.10 Methods of Construction of Price Index Numbers

The following are the different methods of construction of index numbers of prices of different commodities:

i. Simple (or unweighted) aggregative method

ii. Simple (or unweighted) average relatives

iii. Weighted aggregative method

iv. Weighted average of relatives

SGOU

4.2.10.1 Simple (or Unweighted) Aggregative Method

The unweighted aggregate index represents the most basic aggregate index number, where each item included in the calculation is treated with equal significance, implying no differential weighting is applied.

In this method, the total current year prices of the commodities are divided by the total base year prices, and the quotient is multiplied by 100.

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

Here, P_{01} is the Price Index, i.e., the current year's price index based on the base year's Price.

$\sum p_1$ = Total prices of the current year of various commodities.

$\sum p_0$ =Total prices of the base year.

Steps

- ◊ To get $\sum p_1$, add the current year prices of various commodities.
- ◊ Calculate $\sum p_0$ by adding the base year prices of several commodities.
- ◊ To acquire the index number, divide $\sum p_1$ by $\sum p_0$ and multiply the quotient by 100.

Limitations of the method

Although the Simple Aggregate Method is easy to compute, it has notable limitations.

- ◊ Items with higher unit prices tend to influence the resulting index number disproportionately.
- ◊ The method fails to reflect individual goods' varying significance or weight in the overall calculation.

Illustration 4.2.1

From the following data, calculate the Index Number using the Simple Aggregate Method.

Commodity	A	B	C	D
Price in 1990 (₹)	160	258	250	139
Price in 1991(₹)	170	165	190	144

SGOU

Solution

Computation of price index number

Commodity	Price (in Rupees)	
	1990(p_0)	1991(p_1)
A	160	170
B	258	165
C	250	190
D	139	144
Total	$\sum p_0 = 807$	$\sum p_1 = 669$

The price index number using the Simple Aggregate Method is given by:

$$= \Sigma P1 / \Sigma P0 \ x \ 100$$

$$= \frac{669}{807} \times 100$$

$$= 82.90$$

Illustration 4.2.2

The table below provides the retail prices for 1998, 2002, and 2006 for five items-soap, edible oil, sugar, rice, and bread that are part of a family's shopping basket. Using this data, compute the unweighted aggregate price index numbers for 2002 and 2006, using 1998 as the base year.

Retail prices of a family's shopping basket in 1998, 2002, and 2006

Items	1998	2002	2006
Soap (1 dozen)	80	100	120
Edible oil (1 litre)	60	75	90
Sugar (1kg)	25	27	30
Rice (1Kg)	20	22	25
Bread (250gm)	15	17	20

Solution

To compute the unweighted aggregate price index numbers, first of all we need to compute the total (aggregate) price for five items: soaps, edible oil, sugar, rice, and bread, as shown in the table below:-

SGOU

Items	1998 (P_0)	2002 (P_1)	2006
Soap (1 dozen)	80	100	120
Edible oil (1 litre)	60	75	90
Sugar (1kg)	25	27	30
Rice (1Kg)	20	22	25
Bread (250gm)	15	17	20
Total	**200**	**241**	**285**

Using 1998 as the base year, the unweighted aggregate price index numbers for 2002 can be computed as;

$$I_{2002} = \frac{\sum p_{2002}}{\sum p_{1998}} \times 100 = \frac{241}{200} \times 100 = 120.5$$

Using 1998 as the base year, the unweighted aggregate price index numbers for 2006 can be computed as;

$$I_{2006} = \frac{\sum p_{2006}}{\sum p_{1998}} \times 100 = \frac{285}{200} \times 100 = 142.5$$

The value $I_{2002} = 120.5$ indicates that the Price of the items included in the family's shopping basket has increased by 20.5% compared to 1998. Similarly, I_{2006} = 142.5 indicates that the Price of the items included in the family's shopping basket has increased by 42.5% compared to 1998.

4.2.10.2 Simple (or Unweighted) Average Relatives

In this approach, the initial step involves calculating the Price relative to each commodity. These are determined by expressing the current year's Price as a percentage of the base year's Price. To derive the overall price index for the group of commodities, an appropriate average of these price relatives is computed, typically using either the arithmetic or the geometric Mean.

This method offers certain advantages:

i. It minimises the impact of extreme values on the index.

ii. It remains unaffected by the units in which prices are measured.

However, it also presents some limitations:

1. Choosing the most suitable form of average can be challenging.

2. It assumes that all price relatives carry equal weight or significance

SGOU

Price relative for a commodity = I = $P1/(P0)$ x 100

When the Arithmetic mean is used, the formula for the price index is given by $\Sigma I/n$

When the Geometric Mean is used, its formula is Antilog $[(\Sigma \log I)/n]$

Illustration 4.2.3

Calculate Simple Index Number by Average Relative Method

Items	Price in Base Year (P_0)	Price in Current Year (P_1)
1	5	7
2	10	12
3	15	25
4	20	18
5	8	9

Solution

Items	Price in Base Year (P_0)	Price in Current Year (P_1)	Index ($I = P_1 / P_0 \times 100$)
1	5	7	140.0
2	10	12	120.0
3	15	25	166.7
4	20	18	90.0
5	8	9	112.5
			Total = 629.2

Simple Index number = $\Sigma I/n$ = $629.2/5$ = 125.84

Illustration 4.2.4

Find from the following data simple index number by average relative method using Geometric Mean as average.

Items	**A**	**B**	**C**	**D**	**E**
Price in 2005	5	12	7	10	6
Price in 2008	6	15	5	12	8

Solution

Items	Price		I = $P1/(P0)$ x 100	log I
	2005 (P_0)	2008 (P_1)		
A	5	6	120	2.0792

SGOU

B	12	15	125	2.0969
C	7	5	71.43	1.8539
D	10	12	120	2.0792
E	6	8	133.3	2.1248
				10.2340

Simple index No. by the average relative method using the Geometric Mean as the average

$= \text{Antilog } [(\Sigma \log I)/n]$

$= \text{Antilog } [10.2340/5]$

=Antilog (2.0468)

=111.4

Illustration 4.2.5

Compute a price index for the following by:

a. Simple aggregate method

b. Simple average of Price relative method using both Arithmetic Mean and Geometric Mean

Commodity	A	B	C	D	E	F
Price in 2005 (Rs.)	20	30	10	25	40	50
Price in 2008 (Rs.)	25	30	15	35	45	55

Solution

Commodity	P_0	P_1	$I = P1/(P0) \times 100$	log I
A	20	25	125	2.0969
B	30	30	100	2.0000
C	10	15	150	2.1761
D	25	35	140	2.1461
E	40	45	112.5	2.0511
F	50	55	110	2.0414
Total	175	205	737.5	12.5116

a. Simple Aggregative Index Number = $\Sigma P0/(\Sigma P1) \ x \ 100 = 205/175 \ x \ 100 =$ 117.143

b. Simple Average Relative Index Number using Arithmetic Mean = = $\Sigma I/n$ = $737.5/6$ = 122.92

c. Simple average relative Index no. by Geometric Mean as average = Antilog $[(\Sigma \log I)/n]$

= Antilog $[12.5116/6]$ = Antilog (2.0853) = 123.7

4.2.10.3 Weighted Aggregative Method

This method assigns appropriate weights to various commodities to reflect their relative importance in the group. The weights can be production figures, consumption figures or distribution figures. For constructing the price index numbers, quantity weights are used, i.e., the amount of the quantity consumed, purchased or marketed. If 'w' is the weight attached to a commodity, then the price index is given by:

$$P_{01} = \frac{\sum wp_1}{\sum wp_2} \times 100$$

There are different ways to assign weights to each item in the basket. In addition, there are different ways to use weighted aggregates to calculate an index. This section will focus on a few approaches to determine how to assign weights to different items in a basket. These methods are as follows:

- ◊ Laspeyre's Method
- ◊ Paasche's Method
- ◊ Dorbish-Bowley Method
- ◊ Fisher's Ideal Method
- ◊ Marshall- Edgeworth Method
- ◊ Kelley's Method

i. Laspeyre's Method

1871 German economist Etienne Laspeyre developed this approach for constructing price indices. The method assumes that the quantities consumed in the base year and current year are the same and that weights are determined by the quantities consumed in the base year. The formula is:

$$P_{01}(L) = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

Steps

i. Calculate p1q0 by multiplying the current year price of each commodity by the corresponding base year quantity.

ii. Calculate p0q0 by multiplying the base year price of each commodity by its base year quantity.

SGOU

iii. To calculate the price index, divide p1q0 by p0q0 and multiply the quotient by 100.

Advantages

The main advantages of the method are:

◊ It is simple to understand and compute.

◊ It is based on fixed weights since the quantity of the base year is used as the weight of the item in both years.

◊ It does not necessitate quantities consumed in the current year.

Disadvantages

The following are the main disadvantages of the method:

◊ The concept that the exact quantities are utilised in the base and current years is inappropriate.

◊ It does not allow the use of averages such as geometric Mean, median, etc.

◊ This method does not satisfy the ideal index number test.

◊ It does not use current year quantities, even though they are accessible.

ii. Paasche's Method

A German mathematician, Paasche developed this method in 1874 to improve Laspeyre's method. He used the current year's quantities as the weights of the items in this method. As a result, Paasche provides the following formula:

$$P_{01}(P) = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$$

Steps

i. To get $\sum p_1 q_1$, multiply the current year prices of various commodities by their respective current year quantities.

ii. Calculate $\sum p_0 q_1$ by multiplying the base-year prices of various commodities by their respective current year quantities.

iii. Multiply the quotient by 100 after dividing $\sum p_1 q_1$ by $\sum p_0 q_1$.

Advantages

◊ It is easy to calculate and understand.

◊ It is based on constant weights for both the current and base years.

◊ The base year quantity is not required.

◊ It meets the unit test for the adequacy of an index number formula.

SGOU

Disadvantages

- ◊ The method assumes that consumption quantities are the same in the base and current years.
- ◊ Each time an index number is constructed, weights must be determined.
- ◊ The time-reversal, factor reversal, and circular tests required for an ideal index number are unmet.
- ◊ It completely ignores the actual quantities consumed in the base year.

iii. Dorbish-Bowley Price Index

The Dorbish-Bowley Price Index is a composite index that combines the Laspeyres and Paasche price indices. It is calculated as the arithmetic mean of the Laspeyres and Paasche price indices. The formula for the Dorbish-Bowley Price Index is given by:

(Dorbish-Bowley Price Index = (Laspeyres Price Index + Paasche Price Index) / 2)

To understand the formula, let us first define the Laspeyres and Paasche price indices:

Laspeyres Price Index: The Laspeyres Price Index measures the change in the cost of a fixed basket of goods and services from the base to the current period. It uses the quantities from the base period as weights.

(Laspeyres Price Index = (Σp_1q_0 / Σp_0q_0) × 100)

Where: p1 = Price of the commodity in the current period p0 = Price of the commodity in the base period q0 = Quantity of the commodity in the base period

Paasche Price Index: The Paasche Price Index measures the change in the cost of a fixed basket of goods and services from the base period to the current period, but it uses the quantities from the current period as weights.

(Paasche Price Index = (Σp_1q_1 / Σp_0q_1) × 100)

Where: p_1 = Price of the commodity in the current period p_0 = Price of the commodity in the base period q_1 = Quantity of the commodity in the current period

Now, the Dorbish-Bowley Price Index is calculated by taking the arithmetic mean of the Laspeyres and Paasche price indices:

Dorbish-Bowley Price Index = $$P_{01}^{DB} = \frac{1}{2}\left[\frac{\sum p_1 q_0}{\sum p_0 q_0} + \frac{\sum p_1 q_1}{\sum p_0 q_1}\right] \times 100$$

The Dorbish-Bowley Price Index is considered a better measure than the Laspeyres or Paasche Index alone because it considers both the base and current period quantities as weights. It provides a compromise between the two indices. It is often used when there is a significant difference between the Laspeyres and Paasche price indices due to changes in consumption patterns or substitution effects.

SGOU

iv. Fisher's Ideal Index Number

The method for constructing index numbers proposed by Prof. Irving Fisher is Fisher's Ideal Index. It is the geometric Mean of the index numbers of Laspeyre's and Paasche.

The formula is:

$$P_{01}(F) = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$$

Characteristics

- ◊ It is based on the Geometric Mean, which is thought to be the ideal average for building index numbers.
- ◊ It satisfies time and factor reversal tests.
- ◊ It removes the bias associated with Laspeyre's and Paasche's methods by considering both the current year's and base year's weights.

Merits

- ◊ It is created using the Geometric Mean, the most appropriate average for an index number.
- ◊ It uses all available data, including p_1, p_0, q_1 and q_0.
- ◊ It is ideal since it meets time reversal, factor reversal, and unit tests.
- ◊ It reflects the impact of the current year and the base year.

Demerits

- ◊ Calculation is quite complex.
- ◊ Each time an index is constructed, it requires data on the current Price and quantity.
- ◊ It's a combination of Laspeyre's and Paasche's methods.

v. Marshall-Edgeworth Index (MEI)

The Marshall-Edgeworth Index (MEI) is a price index that seeks to measure changes in the prices of a fixed basket of goods and services over time.

Taking the arithmetic cross of the quantities in the base year and the current year as weights i.e., w = (q0 + q1)/2, we obtain the Marshall-Edgeworth (M.E.) formula given by:

$$\boldsymbol{P_{01}(ME) = \frac{\sum p_1 \ (q_0 + q_1)/2}{\sum p_0 \ (q_0 + q_1)/2} \times 100}$$

SGOU

$$= \frac{\sum p_1(q_0 + q_1)}{\sum p_0 (q_0 + q_1)} \times 100$$

$$= \frac{\sum p_1 q_0 + \sum p_1 q_1}{\sum p_0 q_0 + \sum p_0 q_1} \times 100$$

Therefore,

$$P_{01}(ME) = \frac{\sum p_1 q_0 + \sum p_1 q_1}{\sum p_0 q_0 + \sum p_0 q_1} \times 100$$

Illustration 4.2.6

From the following data calculate price index numbers for 1980 with 1970 as base by (i) Laspeyre's method, (ii) Paasche's method, (iii) Marshall- Edgeworth method, and (iv) Fisher's ideal method.

Commodities	1970		1980	
	Price	Quantity	Price	Quantity
A	20	8	40	6
B	50	10	60	5
C	40	15	50	15
D	20	20	20	25

a. It is stated that Marshall- Edgeworth index number is a good approximation to Fisher's ideal index number. Verify this for the data in Part(a).

Solution

Calculations for Price Indices in various methods

Commo- dities	1970		1980					
	p_0	q_0	p_1	q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	20	8	40	6	160	120	320	240
B	50	10	60	5	500	250	600	300
C	40	15	50	15	600	600	750	750
D	20	20	20	25	400	500	400	500
					$\sum p_0q_0$= 1660	$\sum p_0q_1$= 1470	$\sum p_1q_0$= 2070	$\sum p_1q_1$= 1790

SGOU

i. Laspeyre's Price Index

$$P^{La} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100 = \frac{2070}{1660} \times 100 = 1.24699 \times 100 = 124.699$$

ii. Paasche's Price Index

$$P^{Pa} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100 = \frac{1790}{1470} \times 100 = 1.2177 \times 100 = 121.77$$

iii. Marshall- Edgeworth Price Index

$$P^{ME} = \frac{\sum p_1 q_0 + \sum p_1 q_1}{\sum p_0 q_0 + \sum p_0 q_1} \times 100 = \frac{2070+1790}{1660+1470} \times 100 = \frac{3860}{3130} \times 100 = 123.32$$

iv. Fisher's Price Index

$$P^{F} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100 = \sqrt{\frac{2070}{1660} \times \frac{1790}{1470}} \times 100$$

$$= 100 = \sqrt{1.51846} \times 100$$

$$\sqrt{1.24699 \times 1.2177}$$

$$= 1.23226 \times 100$$

$$= 123.23$$

Aliter:

$$P^{F} = \sqrt{P01La \times P01Pa} = \sqrt{124.699 \times 121.77} = \sqrt{15184.597} = 123.23$$

(b) Since $P_{01}^{ME} = 123.32$ and $P_{01}^{F} = 123.23$, are approximately equal, Marshall Edgeworth index number is a good approximation to Fisher's ideal index number.

vi. Kelly's Price Index

Kelly's Price Index or Fixed Weights Index. This formula, named after Truman L. Kelly, requires the weights to be fixed for all periods and is also sometimes known as the aggregative index with fixed weights and is given by the formula:

$$P_{01}(K) = \frac{\sum q\, p_1}{\sum q\, p_0} \times 100$$

The weights are the quantities (q) which may refer to some period (not necessarily the base year or the current year) and are kept constant for all periods. The average (A.M. or G.M.) quantities consumed in two, three, or more years may be used as weights. Kelly's fixed base index has a distinct advantage over Laspeyre's Index because, unlike Laspeyre's Index, the change in the base year does not necessitate a corresponding change in the weights, which can be kept constant until new data becomes available to revise the index. As such, this index is currently finding great favour and becoming quite popular. The Labour Bureau's wholesale price index in the U.S.A. is based on this method.

SGOU

Illustration 4.2.7

Calculate the weighted price index from the following data;

Materials required	Unit	Quantity required	Price (Rs.)	
			1963	1973
Cement	100 lb	500 lb.	5.0	8.0
Timber	c.ft.	2,000 c.ft.	9.5	14.2
Steel sheets	cwt.	50 cwt.	34.0	42.20
Bricks	per'000	20,000	12.0	24.0

Solution

Since the quantities (weights) required of different materials are fixed for both the base and current years, we will use Kelly's formula to find the price index. Further, the cement unit is 100 lbs. The quantity required is 500 lbs. Hence, the quantity consumed per unit for cement is 500/100 = 5. Similarly, the quantity consumed per unit for bricks is 20,000/1,000 = 20.

Computation of Kelly's Index Number

Materials required	Unit	Quantity required	q	Price (Rs.)		$q p_0$	$q p_1$
				1963	1973		
				p_0	p_1		
Cement	100 lb	500 lb.	5	5.0	8.0	25	40
Timber	c.ft.	2,000 c.ft.	2000	9.5	14.2	19,000	28,400
Steel sheets	cwt.	50 cwt.	50	34.0	42.20	1,700	2,100
Bricks	per'000	20,000	20	12.0	24.0	240	480
						$\sum qp_0$=20,965	$\sum qp_1$=31,020

Kelly's Price Index is given by: $P_{01}{}^{K} = \frac{\sum q\, p1}{\sum q\, p0} \times 100 = \frac{31020}{20965} \times 100$

$= 147.96$

4.2.10.4 Weighted Average of Price Relatives Method

This approach gives suitable weights to commodities based on their relative significance within the group. The overall index for the group is then calculated by taking the weighted average of the price relatives. This average can be determined using either the Arithmetic Mean or the Geometric Mean.

SGOU

When A. M. is used, the index is $\Sigma IV/\Sigma V$, where 'V' stands for the weights.

When Geometric Mean is used, the Index is Antilog $(\Sigma(\log \llbracket I \times V \rrbracket)/\Sigma V$

Illustration 4.2.8

Calculate the Index Number of Price for 2007 based on 2006, from the data given below:

Commodities	Weights	Price (2006)	Price (2007)
A	40	16.00	20.00
B	25	40.00	60.00
C	5	0.50	0.50
D	20	5.12	6.25
E	10	2.00	1.50

Solution

Commodities	Weights (V)	P_0 (2006)	P_1 (2007)	I	IV
A	40	16.00	20.00	125	5000
B	25	40.00	60.00	150	3750
C	5	0.50	0.50	100	500
D	20	5.12	6.25	122.1	2442
E	10	2.00	1.50	75	750
Total	100				12442

Index number for 2007 = $\Sigma IV/\Sigma V$ = 12442 / 100 = 124.42

Illustration 4.2.9

Find the weighted index number by Average Relative Method

Commodities	Price 2005 (P_0)	Price 2008 (P_1)	Quantity 2005 (q_0)
A	5	12	2
B	8	10	1
C	4	5	3
D	9	18	4
E	3	4	2

Solution

(Hint: Since the problem is to be done by average relative method, we have to find I and V where:

$I = (P_1 / P_0) \times 100$ and $V = P_0 \times q_0$)

SGOU

Commodities	P_0	P_1	q_0	I	V	IV
A	5	12	2	240	10	2400
B	8	10	1	125	8	1000
C	4	5	3	125	12	1500
D	9	18	4	200	36	7200
E	3	4	2	133.3	6	799.8
Total					72	12899.8

Weighted Index Number = $\Sigma IV / \Sigma V$ = 12899.8 / 72 = 179.16

Note: Weight is equal to $P_0 \times q_0$ if not otherwise given.

Illustration 4.2.10

Find the weighted Index Number by the average relative method using the Geometric Mean as the average

Commodities	Indices	Weight
A	110	15
B	180	10
C	200	12
D	215	17
E	190	6

Solution

Commodities	I	V	log I	log I × V
A	110	15	2.0414	30.6210
B	180	10	2.2553	22.5530
C	200	12	2.3010	27.6120
D	215	17	2.3324	39.6508
E	190	6	2.2788	13.6728
Total		60		134.1096

When index Number = Antilog $(\Sigma(\log \ [I \times V)] \ )/\Sigma V$ = Antilog $134.1096/60$

= Antilog (2.2352) = 171.9

SGOU

Summarised Overview

Index numbers are statistical tools used to measure changes in economic variables over time, such as prices, quantities, or values, providing a concise representation of trends and comparisons. Their importance lies in enabling policymakers, businesses, and researchers to track inflation, cost of living, production levels, and other economic indicators. However, constructing index numbers involves challenges like selecting appropriate base years, choosing representative items, and accounting for quality changes. Construction methods include simple and weighted indices, with weighted indices offering greater accuracy by incorporating item significance. Aggregative methods sum up price or quantity changes, while the average Price relative method compares individual price changes as percentages. Key formulas include Laspeyres' method (base-period weights), Paasche's method (current-period weights), and Fisher's ideal index (geometric Mean of Laspeyres and Paasche), which balances their biases. Other notable approaches include Bowley's and Dorbish's method (arithmetic mean of Laspeyres and Paasche) and the Marshall-Edgeworth method, which uses average quantities from both periods for weighting, enhancing reliability and representativeness.

Self-Assessment Question

1. What is the primary purpose of index numbers in statistics?
2. How do index numbers simplify complex data in various fields?
3. Why is the selection of a base period important in index number calculations?
4. What is the importance of reflecting weighted importance in index numbers?
5. Name some fields where index numbers are widely used and provide examples of their applications.
6. What are the limitations associated with the use of index numbers?
7. Discuss the importance of transparency in the construction of index numbers.
8. How do price index numbers impact economic analysis, monetary policy, and salary adjustments?

SGOU

9. What is the difference between the Simple Aggregate Method and the Simple Average Relatives Method in constructing index numbers?

10. When is the Simple Aggregate Method considered less reliable, and what is its limitation?

11. What are weighted index numbers, and why are they used in constructing price indices?

12. List the types of weighted index numbers, and briefly describe each one.



Assignments

1. Compare and contrast Laspeyre's and Paasche's methods for constructing weighted index numbers.

2. What is Fisher's Ideal Method, and why is it an improvement over Laspeyre's and Paasche's methods?

3. Explain the importance of the Marshall-Edgeworth Index (MEI) in constructing price indices and its relationship with Fisher's ideal index.

4. Please write a brief essay explaining the concept of price index numbers, their purpose, and how they measure the cost-of-living changes.

5. Find simple index number by aggregative method and Average relative method (using both Arithmetic mean and Geometric mean) from the following prices of 5 commodities.

Year	Commodities					
		A	B	C	D	E
	2007	2	10	8	9	12
	2008	4	11	7	11	15

(Ans:- 117.07, 128.94, 124.08)

6. Construct an index number for 2001 taking 2000 as base.

Commodity	Price in 2000	Price in 2001
A	90	95
B	40	60
C	90	110
D	30	35

(Ans:- 123.61)

SGOU

7. You are given the following data:

	Price in Rs.	
Commodity	2005	2008
Wheat	170	200
Rice	140	180

Using simple average of price relatives prepare two price index numbers from the above data i.e

i. Price index for 2008 with base 2005 and

ii. A price index for 2005 with base 2008

(Ans:- 123.11,81.39)

8. Using the data given below calculate price index numbers for the year 2007 by (1) Laspeyer's formula (2) Paasche's formula (3) Fisher's formula (4) Marshall Edgeworth formula (5) Bowley's formula.

Commodity	Price (Rs.)		Quantity ('000 kg)	
	2000	2007	2000	2007
Rice	9.3	4.5	100	90
Wheat	6.4	3.7	11	13
Pulses	5.1	2.6	5	3

(Ans:- 49.1, 49.3, 49.2,49.2)

9. Construct Laspeyer's formula and Paasche's formula Index numbers for 2008 with 2007 as base.

Commodity	2007		2008	
	Qty Consumed	Price	Qty Consumed	Price
Wheat	3.0	2.00	2.75	4.00
Grain	0.50	4.00	0.50	6.00
Vegetable	1.00	2.00	1.25	4.00
Meat	0.50	12.00	0.75	16.00
Fish	0.50	6.00	0.50	6.00

(Ans:- 166.67, 161.9)

10. From the following data construct an appropriate Index number

Commodity	2005		2008	
	Price	Qty	Price	Qty
Wheat	30	40	35	45

SGOU

Grain	25	30	30	36
Barley	16	30	20	38

Hint:- Find Fisher's Index number

(Ans:- 119.5)

11. From the data given below construct an index number of the group of four commodities by using Fisher's Index formula.

Commodities	Base year		Current year	
	Price/unit	Expenditure (Rs.)	Price/unit	Expenditure (Rs.)
1	2	40	5	75
2	4	16	8	40
3	1	10	2	24
4	5	25	10	70

Hint:- Quantity = Expenditure/Price

(Ans:- 218.7)

12. The following figures relate to the price and quantities of certain commodities. Construct an index number from the following data for the year 2008, taking 2005 as the base year.

Commodity	Quantity consumed	Price (in Rs.) in 2005	Price (in Rs.) in 2008
A	50	32	40
B	35	30	42
C	55	16	24
D	45	40	52
E	15	35	42

Hint:- Apply Kelley's formula

(Ans:- 132.5)

SGOU

Reference

1. Aggarwal, B. M. (2010). *Business statistics*. Ane Books Pvt. Ltd.
2. Chandrasekharan, N., & Umaparvathi, M. (2016). *Statistics for management*. PHI Learning Pvt. Ltd.

3. Gupta, M., & Khanna, R. (2011). *Quantitative techniques for decision making*. Prentice Hall of India.

4. Gupta, S. P. (2021). *Statistical methods*. S. Chand & Sons.

5. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International (P) Ltd.

6. Levin, R. I., & Rubin, D. S. (1999). *Quantitative approaches for management*. Pearson Education.

7. Sharma, A. (2017). *Quantitative techniques for decision making*. Himalaya Publishing House.

8. Vohra, D. V. D. (2017). *Quantitative techniques for management*. New Delhi: McGraw Hill Education



Suggested Reading

1. Arora, P. N., Arora, S., & Arora, S. (2010). *Comprehensive statistical methods*. S. Chand & Company Pvt. Ltd.

2. Beri, G. C. (2017). *Business statistics*. Tata McGraw-Hill.

3. Chawla, D., & Sondhi, N. (2016). *Research methodology: Concepts and cases*. Vikas Publishing House.

4. Krishnaswami, O. R., Ranganathan, M., & Harikumar, P. N. (2016). *Research methodology*. Himalaya Publishing House.

5. Render, B. (2022). *Quantitative analysis for management*. Prentice Hall of India.

6. Singh, A. K. (2017). *Tests, measurements and research methods in behavioural sciences*. Bharti Bhawan.

SGOU

MODEL QUESTION PAPER SETS

SGOU

SREENARAYANAGURU OPEN UNIVERSITY

Model Question Paper (SET- A)

QP CODE: ……

Reg. No:.....

Name: ……

FIRST SEMESTER MASTER OF BUSINESS ADMINISTRATION (MBA) EXAMINATION

DISCIPLINE CORE COURSE

M26BS04DC - QUANTITATIVE TECHNIQUES

(CBCS - UG)

2026 - Admission Onwards

Time: 3 Hours Max Marks: 70

Section A

(Answer any 5, each carries 2 marks)

(5x2=10)

1. What are the key objectives of applying quantitative techniques in management studies?
2. State Bayes' Theorem with an example.
3. How does Type I error differ from Type II error?
4. Explain goodness of fit.
5. What is a correlation coefficient, and how is it defined?
6. What is an independent variable? Illustrate with a relevant example.
7. Define secular trend.
8. What are Price Index numbers?

Section B

(Answer any 6, each carries 5 marks)

(6x5=30)

9. Critically evaluate the role of quantitative techniques in strategic business decision-making.
10. In a normal distribution, 31% of the items are under 45 and 8 % are over 64. Find the mean and standard deviation of the distribution.

SGOU

11. A stenographer claims that she can take dictation at the rate of 120 words per minute. Can we reject her claim on the basis of 100 trails in which she demonstrates a mean of 116 words with a variance of 225 words use 5% level of significance.

12. In a certain college, it is estimated that fewer than 25% of students have cars on campus. Does this seem to be a valid estimate, if, in a random sample of 90 college students, 28 are found to have cars. Use 0.05 level of Significance.

13. Discuss briefly Wilcoxon matched pair test and Wilcoxon Mann Whitney U-test?

14. What is a Scatter Diagram? How is it useful in the study of correlation?

15. If r = 0.8 and N = 0.81, find the probable error of the coefficient of correlation.

16. For 10 observations on price X and supply Y, the following data were obtained (in appropriate units)

$$\sum X = 130, \sum Y = 220, \sum XY = 3467, \sum X^2 = 2288, \sum Y^2 = 5506$$

Obtain the line of regression of Y on X and estimate the supply when the price is 16 units

17. Explain the method of least squares.

18. Compare and contrast Laspeyre's and Paasche's methods for constructing weighted index numbers.

Section C

(Answer any 2, each carries 15 marks)

(2x15=30)

19. State the relationship between the Binomial distribution, Normal distribution and the Poisson distribution

20. Four varieties of a crop were grown on 3 plots, and the following yield was obtained. You are required to test whether there is a significant difference in the productivity of seeds:

Plot	Variety of Seeds			
	P	Q	R	S
I	10	7	8	5
II	9	7	5	4
III	8	6	4	4

SGOU

21. Apply concurrent deviation method to determine the value of coefficient of correlation from the data given below.

X	100	110	110	120	122	125
Y	120	140	160	160	130	110

22. Fit a straight-line trend by the method of least squares and find the trend values.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008
Value	68	72	73	80	90	92	90	110	114

SGOU

SREENARAYANAGURU OPEN UNIVERSITY

Model Question Paper (SET- B)

QP CODE: ……

Reg. No:.....

Name: ……

FIRST SEMESTER MASTER OF BUSINESS ADMINISTRATION (MBA) EXAMINATION

DISCIPLINE CORE COURSE

M26BS04DC - QUANTITATIVE TECHNIQUES

(CBCS - UG)

2026 - Admission Onwards

Time: 3 Hours Max Marks: 70

Section A

(Answer any 5, each carries 2 marks)

(5x2=10)

1. What is meant by a random experiment?
2. How does Poisson distribution differ from Binomial distribution?
3. What is two-tailed test?
4. Describe Kruskal-Wallis H-test?
5. Describe the idea of a perfectly positive correlation between two variables.
6. What are regression coefficients?
7. Define cyclical variations.
8. How do index numbers simplify complex data in various fields?

Section B

(Answer any 6, each carries 5 marks)

(6x5=30)

9. What are the main properties of the binomial distribution?
10. A box containing 100 transistors, 20 of which are defective, 10 are selected for inspection. Indicate what is the probability that (i) all 10 are defective (ii) 10 are good (iii) at least one is defective and (iv) at the most three are defective?

SGOU

11. A sample of 900 items is taken from a population with SD 15. The mean of the sample is 25. Test whether the sample has come from a population with mean 26.8.

12. The average life of 26 electric bulbs were found to be 1200 hours with a standard deviation of 150 hours. Test whether these bulbs could be considered as a random sample from a normal population with mean 1300 hours.

13. Distinguish between parametric tests and non-parametric tests.

14. What do you mean by probable error?

15. Find Karl Pearson's correlation coefficient between x and y for the following data,

$$n = 10, \Sigma x = 12., \Sigma x2 = 450, \Sigma y = 15, \Sigma y2 = 320, \Sigma xy = 285$$

16. The following data shows the monthly income and expenditure on food of 10 families

Income	120	90	83	150	130	140	110	95	75	105
Expenditure	40	36	40	45	40	44	45	38	50	35

Calculate the linear regression of Expenditure of food(y) on Income(x)

17. From the following series of annual data find the trend values by using 3 year moving average

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Value	170	231	261	264	278	302	289	297	340	273

18. Explain the importance of the Marshall-Edgeworth Index (MEI) in constructing price indices and its relationship with Fisher's ideal index.

Section C

(Answer any 2, each carries 15 marks)

(2x15=30)

19. The following data show the number of seeds germinating out of 5 ib damp filter for 80 sets of seeds. Fit a binomial distribution of this data and find the expected frequencies.

x	0	1	2	3	4	5
y	6	20	28	12	8	6

SGOU

20. In an experiment with the immunisation of cattle from tuberculosis following results were obtained

	Affected	Un affected
Innoculated	12	26
Non innoculated	16	6

Explain the effect of the vaccine in controlling susceptibility to tuberculosis.

21. Distinguish between Correlation and Regression.

22. Using the data given below calculate price index numbers for the year 2007 by (1) Laspeyer's formula (2) Paasche's formula (3) Fisher's formula (4) Marshall Edgeworth formula (5) Bowley's formula.

Commodity	Price (Rs.)		Quantity ('000 kg)	
	2000	2007	2000	2007
Rice	9.3	4.5	100	90
Wheat	6.4	3.7	11	13
Pulses	5.1	2.6	5	3

SGOU

Appendix Tables



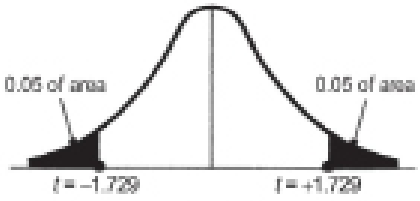
EXAMPLE: TO FIND THE AREA UNDER THE CURVE BETWEEN THE MEAN AND A POINT 2.24 STANDARD DEVIATIONS TO THE RIGHT OF THE MEAN, LOOK UP THE VALUE OPPOSITE 2.2 AND UNDER 0.04 IN THE TABLE; 0.4875 OF THE AREA UNDER THE CURVE LIES BETWEEN THE MEAN AND A z VALUE OF 2.24.

AREAS UNDER THE STANDARD NORMAL PROBABILITY DISTRIBUTION BETWEEN THE MEAN AND POSITIVE VALUES OF z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

SGOU

EXAMPLE: TO FIND THE VALUE OF t THAT CORRESPONDS TO AN AREA OF 0.10 IN BOTH TAILS OF THE DISTRIBUTION COMBINED, WHEN THERE ARE 19 DEGREES OF FREEDOM, LOOK UNDER THE 0.10 COLUMN, AND PROCEED DOWN TO THE 19 DEGREES OF FREEDOM ROW; THE APPROPRIATE t VALUE THERE IS 1.729.

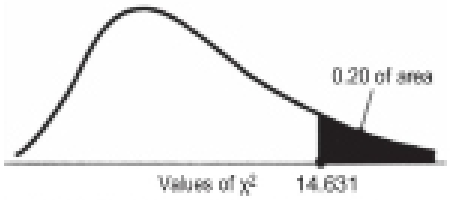


AREAS IN BOTH TAILS COMBINED FOR STUDENT'S t DISTRIBUTION

Degrees of Freedom	Area in Both Tails Combined			
	0.10	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	0.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
Normal Distribution	1.645	1.960	2.326	2.576

SGOU

EXAMPLE: IN A CHI-SQUARE DISTRIBUTION WITH 11 DEGREES OF FREEDOM, TO FIND THE CHI-SQUARE VALUE FOR 0.20 OF THE AREA UNDER THE CURVE (THE COLORED AREA IN THE RIGHT TAIL) LOOK UNDER THE 0.20 COLUMN IN THE TABLE AND THE 11 DEGREES OF FREEDOM ROW; THE APPROPRIATE CHI-SQUARE VALUE IS 14.631.



AREA IN THE RIGHT TAIL OF A CHI-SQUARE (χ^2) DISTRIBUTION

Degrees of Freedom	Area in Right Tail				
	0.99	0.975	0.95	0.90	0.800
1	0.00016	0.00098	0.00398	0.0158	0.0642
2	0.0201	0.0506	0.103	0.211	0.446
3	0.115	0.216	0.352	0.584	1.005
4	0.297	0.484	0.711	1.064	1.649
5	0.554	0.831	1.145	1.610	2.343
6	0.872	1.237	1.685	2.204	3.070
7	1.239	1.690	2.167	2.833	3.822
8	1.646	2.180	2.733	3.490	4.594
9	2.088	2.700	3.325	4.168	5.380
10	2.558	3.247	3.940	4.865	6.179
11	3.053	3.816	4.575	5.578	6.989
12	3.571	4.404	5.226	6.304	7.807
13	4.107	5.009	5.892	7.042	8.634
14	4.660	5.629	6.571	7.790	9.467
15	5.229	6.262	7.261	8.547	10.307
16	5.812	6.908	7.962	9.312	11.152
17	6.408	7.564	8.672	10.085	12.002
18	7.015	8.231	9.390	10.865	12.857
19	7.633	8.907	10.117	11.651	13.716
20	8.260	9.591	10.851	12.443	14.578
21	8.897	10.283	11.591	13.240	15.445
22	9.542	10.982	12.338	14.041	16.314
23	10.196	11.689	13.091	14.848	17.187
24	10.856	12.401	13.848	15.658	18.062
25	11.524	13.120	14.611	16.473	18.940
26	12.198	13.844	15.379	17.292	19.820
27	12.879	14.573	16.151	18.114	20.703
28	13.565	15.308	16.928	18.939	21.588
29	14.256	16.047	17.708	19.768	22.475
30	14.953	16.791	18.493	20.599	23.364

SGOU

Note: If ν, the number of degrees of freedom, is greater than 30, we can approximate χ^2_α the chi-square value leaving α of the area the right tail, by

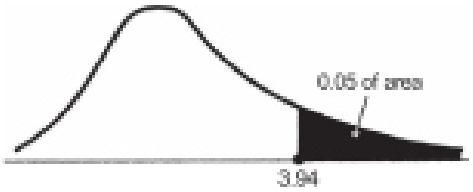
$$\chi^2_\alpha = \nu\left(1 - \frac{2}{9\nu} + z_\alpha\sqrt{\frac{2}{9\nu}}\right)^3$$

where z_α is the standard normal value (from Appendix Table 1) that leaves α of the area in the right tail.

Area in Right Tail					Degrees of Freedom
0.20	0.10	0.05	0.025	0.01	
1.642	2.706	3.841	5.024	6.635	1
3.219	4.605	5.991	7.378	9.210	2
4.642	6.251	7.815	9.348	11.345	3
5.989	7.779	9.488	11.143	13.277	4
7.289	9.236	11.070	12.833	15.086	5
8.558	10.645	12.592	14.449	16.812	6
9.803	12.017	14.067	16.013	18.475	7
11.030	13.362	15.507	17.535	20.090	8
12.242	14.684	16.919	19.023	21.666	9
13.442	15.987	18.307	20.483	23.209	10
14.631	17.275	19.675	21.920	24.725	11
15.812	18.549	21.026	23.337	26.217	12
16.985	19.812	22.362	24.736	27.688	13
18.151	21.064	23.685	26.119	29.141	14
19.311	22.307	24.996	27.488	30.578	15
20.465	23.542	26.296	28.845	32.000	16
21.615	24.769	27.587	30.191	33.409	17
22.760	25.989	28.869	31.526	34.805	18
23.900	27.204	30.144	32.852	36.191	19
25.038	28.412	31.410	34.170	37.566	20
26.171	29.615	32.671	35.479	38.932	21
27.301	30.813	33.924	36.781	40.289	22
28.429	32.007	35.172	38.076	41.638	23
29.553	33.196	36.415	39.364	42.980	24
30.675	34.382	37.652	40.647	44.314	25
31.795	35.563	38.885	41.923	45.642	26
32.912	36.741	40.113	43.194	46.963	27
34.027	37.916	41.337	44.461	48.278	28
35.139	39.087	42.557	45.722	49.588	29
36.250	40.256	43.773	46.979	50.892	30

SGOU

EXAMPLE: IN AN *F* DISTRIBUTION WITH 15 DEGREES OF FREEDOM FOR THE NUMERATOR AND 6 DEGREES OF FREEDOM FOR THE DENOMINATOR, TO FIND THE *F* VALUE FOR 0.05 OF THE AREA UNDER THE CURVE LOOK UNDER THE 15 DEGREES OF FREEDOM COLUMN AND ACROSS THE 6 DEGREES OF FREEDOM ROW; THE APPROPRIATE *F* VALUE IS 3.94.



VALUES OF *F* FOR *F* DISTRIBUTIONS WITH 0.05 OF THE AREA IN THE RIGHT TAIL

Degrees of Freedom for Denominator	Degrees of Freedom for Numerator																			
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

SGOU



EXAMPLE: IN AN *F* DISTRIBUTION WITH 7 DEGREES OF FREEDOM FOR THE NUMERATOR AND 5 DEGREES OF FREEDOM FOR THE DENOMINATOR, TO FIND THE *F* VALUE FOR 0.01 OF THE AREA UNDER THE CURVE LOOK UNDER THE 7 DEGREES OF FREEDOM COLUMN AND ACROSS THE 5 DEGREES OF FREEDOM ROW; THE APPROPRIATE *F* VALUE IS 10.5.

VALUES OF *F* FOR *F* DISTRIBUTIONS WITH 0.01 OF THE AREA IN THE RIGHT TAIL

Degrees of Freedom for Denominator	Degrees of Freedom for Numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

SGOU

LOGARITHAM

	0	1	2	3	4	5	6	7	8	9	Mean Difference								
											1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21	25	29	33	37
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19	23	26	30	34
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17	21	24	28	31
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430	3	6	10	13	16	19	23	26	29
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732	3	6	9	12	15	18	21	24	27
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14	17	20	22	25
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13	16	18	21	24
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12	15	17	20	22
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12	14	16	19	21
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11	13	16	18	20
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6345	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9

SGOU

	0	1	2	3	4	5	6	7	8	9	Mean Difference								
											1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	6	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7958	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9026	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9278	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9

ANTILOGARITHAM

	0	1	2	3	4	5	6	7	8	9	Mean Difference								
											1	2	3	4	5	6	7	8	9
.00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
.01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
.02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
.03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
.04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
.05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
.06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
.07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
.08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
.09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
.10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
.11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
.12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	2	3
.13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
.14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
.15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
.16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
.17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
.18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
.19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
.20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
.21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
.22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
.23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
.24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
.25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
.26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
.27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
.28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
.29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
.30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
.31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
.32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
.33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
.34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
.35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286	1	1	2	2	3	3	4	4	5
.36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
.37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
.38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
.39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
.40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
.41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
.42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
.43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
.44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
.45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
.46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
.47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
.48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
.49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9

SGOU

ANTILOGARITHAM

	0	1	2	3	4	5	6	7	8	9	Mean Difference								
											1	2	3	4	5	6	7	8	9
.50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
.51	3236	3243	3251	3258	3266	3273	3281	3288	3296	3304	1	2	2	3	4	5	5	6	7
.52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
.53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
.54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
.55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
.56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
.57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
.58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
.59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
.60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
.61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
.62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
.63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
.64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
.65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
.66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
.67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
.68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
.69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
.70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
.71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
.72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
.73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
.74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
.75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
.76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
.77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
.78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
.79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
.80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
.81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
.82	6607	6622	6637	6653	6668	6683	6699	6715	6730	6745	2	3	5	6	8	9	11	12	14
.83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
.84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
.85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
.86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	9	10	12	13	15
.87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
.88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
.89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	12	14	16
.90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
.91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
.92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
.93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
.94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
.95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
.96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
.97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
.98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
.99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9

SGOU

SREENARAYANAGURU OPEN UNIVERSITY

സർവ്വകലാശാലാഗീതം

വിദ്യയാൽ സ്വതന്ത്രരാകണം
വിശ്വപൗരരായി മാറണം
ഗ്രഹപ്രസാദമായ് വിളങ്ങണം
ഗുരുപ്രകാശമേ നയിക്കണേ

കൂരിരുട്ടിൽ നിന്നു ഞങ്ങളെ
സൂര്യവീഥിയിൽ തെളിക്കണം
സ്നേഹദീപ്തിയായ് വിളങ്ങണം
നീതിവൈജയന്തി പാറണം

ശാസ്ത്രവ്യാപ്തിയെന്നുമേകണം
ജാതിഭേദമാകെ മാറണം
ബോധരശ്മിയിൽ തിളങ്ങുവാൻ
ജ്ഞാനകേന്ദ്രമേ ജ്വലിക്കണേ

കുരീപ്പുഴ ശ്രീകുമാർ

SGOU

SREENARAYANAGURU OPEN UNIVERSITY

Regional Centres

Kozhikode
Govt. Arts and Science College
Meenchantha, Kozhikode,
Kerala, Pin: 673002
Ph: 04952920228
email: rckdirector@sgou.ac.in

Thalassery
Govt. Brennen College
Dharmadam, Thalassery,
Kannur, Pin: 670106
Ph: 04902990494
email: rctdirector@sgou.ac.in

Tripunithura
Govt. College
Tripunithura, Ernakulam,
Kerala, Pin: 682301
Ph: 04842927436
email: rcedirector@sgou.ac.in

Pattambi
Sree Neelakanta Govt. Sanskrit College
Pattambi, Palakkad,
Kerala, Pin: 679303
Ph: 04662912009
email: rcpdirector@sgou.ac.in

DON'T LET IT BE TOO LATE

SAY NO TO DRUGS

LOVE YOURSELF AND ALWAYS BE HEALTHY



SREENARAYANAGURU OPEN UNIVERSITY

SREENARAYANAGURU OPEN UNIVERSITY

The State University for Education, Training and Research in Blended Format, Kerala

Quantitative Techniques

COURSE CODE: M26BS04DC

SGOU

ISBN 978-81-686027-9-3

Sreenarayanaguru Open University
Kollam, Kerala Pin- 691601, email: info@sgou.ac.in, www.sgou.ac.in Ph: +91 474 2966841