COURSE GUIDE

PAD 207
STATISTICS FOR PUBLIC ADMINISTRATION

Course Team Dr. Amehojo Daniel & Dr. Musa Zakari- (Course Writer/Developer)
(NOUN)
Dr. Weinoh Oboromeni- (Content Editor)
(NOUN)

NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
University Village
Plot 91, Cadastral Zone
Nnamdi Azikiwe Expressway
Jabi, Abuja

Lagos Office
14/16 Ahmadu Bello Way
Victoria Island, Lagos

E-mail : centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Printed 2024

ISBN: 978-978-786-239-1

INTRODUCTION

This course, PAD 207: Statistics for Public Administration holds a significant position within the academic curriculum as it carries a weight of three (3) credit units and is considered compulsory for students pursuing a Bachelor of Science (B.Sc) degree in Public Administration or any closely related field of study. The relevance of this course extends beyond its credit value, as it plays a pivotal role in equipping students with crucial statistical skills essential for their academic and professional pursuits.

Students enrolled in B.Sc Public Administration and other related disciplines find themselves directed toward this course to enhance their statistical competencies and analytical proficiencies. The overarching goal of PAD 207 is to provide students with a fundamental understanding of statistical methodologies and tools, enabling them to effectively analyze data, draw informed conclusions, and make evidence-based decisions within the public administration realm.

By exploring into the core principles of statistics, students are not only exposed to theoretical concepts but are also provided with practical applications that bridge the gap between classroom learning and real-world scenarios. Through a dynamic curriculum that covers topics such as data collection, analysis, interpretation, and presentation, students are empowered to navigate the complexities of statistical analysis with precision and confidence.

As students' progress through PAD 207, they are challenged to think critically, problem-solve creatively, and engage with statistical models in a structured and meaningful manner. The course serves as a stepping stone for students to develop a strong statistical foundation, laying the groundwork for future academic endeavors and professional endeavors within the dynamic field of Public Administration.

COURSE GUIDE

The purpose of this course is to present an examination of quantitative methods and techniques that are used in the public sector. Areas of focus include Statistical analysis, sampling, forecasting and time-series analysis, and design issues in generating and testing research questions and hypotheses. Quasi-experimental and non-experimental designs will be used including survey research to assess public input on government service quality and applied to practice and policy issues. IBM SPSS software will be used to analyze government datasets using descriptive and inferential statistics including correlations, cross-tabulations, t-tests, ANOVA, and regression

MEASURABLE LEARNING OUTCOMES

Upon successful completion of these modules, you will be able to:

By the end of this course, students will be able to:

- Understand and apply basic statistical concepts and techniques.
- Analyze and interpret data relevant to public administration.
- Utilize statistical software to conduct statistical analyses.
- Make data-driven decisions and policy recommendations.
- Communicate statistical findings effectively to a non-technical audience.

COURSE OBJECTIVES

PAD207: Statistics for Public Administration provides an introduction to statistical methods and their applications in the field of public administration. The course covers fundamental concepts and techniques necessary for analyzing and interpreting data to make informed decisions in public policy and administration.

By the end of this course, students will be able to:

- Understand and apply basic statistical concepts and techniques.
- Analyze and interpret data relevant to public administration.
- Utilize statistical software to conduct statistical analyses.
- Make data-driven decisions and policy recommendations.
- Communicate statistical findings effectively to a non-technical audience.

ASSESSMENT METHODS

- **Self-Assessment-Exercise (SAEs)**

Two Self-assessment Exercises each are incorporated in the study material for each unit. Self-assessment Exercise helps students to be a realistic judge of their own performance and to improve their work.

Promotes the skills of reflective practice and self-monitoring; Promotes academic integrity through student self-reporting of learning progress; Develops self-directed learning; Increases student motivation and Helps students develop a range of personal, transferrable skills

- **Tutor Mark Assignment (TMAs)**

In academic settings, Tutor Marked Assignments (TMAs) play a crucial role in evaluating students' understanding and application of course materials. These assignments serve as practical opportunities for students to demonstrate their knowledge, critical thinking skills, and ability to communicate effectively through writing. TMAs are designed to challenge students and encourage them to explore deeper into the subject matter by analyzing, synthesizing, and presenting their findings in a structured and cohesive manner. As such, Tutor Marked Assignments are not only assessments but also learning tools that contribute significantly to the overall educational experience.

- **Final Exam: Comprehensive exam covering all modules.**

The final exam is a crucial assessment that evaluates students' knowledge of all modules learned throughout the course. This comprehensive examination is designed to test their understanding and retention of key concepts, theories, and applications from each unit covered in the academic term. Students are expected to demonstrate their proficiency in various subject areas, showcasing their ability to synthesize information, analyze complex problems, and apply critical thinking skills. This comprehensive evaluation serves as a culminating activity to gauge the overall comprehension and mastery of the course material.

- **Possible Answers to Self-Assessment Exercise(s) within the content**

The materials contained Possible Answers to Self-Assessment Exercise(s) within the content. The possible Self-assessments answers enable you to understand how well you're performing in the contents. It is a way of analysing your work performance and any areas for growth. Reflecting on your strengths, weaknesses, values and accomplishments can help you determine what goals to work toward next.

CONCLUSION

This course aims to equip students with the essential statistical skills needed to analyze data and make evidence-based decisions in public administration. Through a combination of theoretical knowledge and practical applications, students will gain the confidence to handle statistical challenges in their professional careers. This course aims to equip students with the essential statistical skills needed to analyze data and make evidence-based decisions in public administration. Through a combination of theoretical knowledge and practical applications, students will gain the confidence to handle statistical challenges in their professional careers.

RECOMMENDED TEXTBOOKS AND RESOURCES

- Statistics for Business and Economics by Paul Newbold, William L. Carlson, and Betty Thorne
- Essentials of Statistics for the Behavioral Sciences by Frederick J. Gravetter and Larry B. Wallnau
- Statistical Methods for the Social Sciences" by Alan Agresti and Barbara Finlay
- Online Resources: Khan Academy, Coursera, edX for supplementary learning materials.

COURSE MATERIAL STRUCTURE

The course material package is comprising of Five (5) Modules and 23-unit structures as follows:

MAIN COURSE

Module 1

Module 2

Module 3

Module 4

Module 5

MODULE 1

UNIT 1 STATISTICS IN SOCIAL RESEARCH FOR PUBLIC ADMINISTRATION

Unit Structure

1.1 Introduction

This unit explains the Statistics in Social Research for Public Administration. By examining into statistical techniques commonly used in social research, it equips learners with the necessary tools to effectively collect, analyze, and interpret data essential for evidence-based decision-making in a public administration setting. Through a series of in-depth modules, participants will gain a deep understanding of how statistical methods are applied within the context of social research, enabling them to draw meaningful insights and conclusions from complex datasets.

1.2 Learning Outcome

At the end of the unit, you should be able to:

- Define Social statistics for Public Administration
- Describe the term descriptive Statistics in Social Research
- Explain the Inferential Statistics in Social Research
- demonstrate the Regression Analysis in Social Research

- Describe the Data Collection and Processing
- Explain Use of Statistical Software

1.3 Definition of Social statistics for Public Administration.

Social statistics for Public Administration involve the collection, analysis, and interpretation of data related to human social behavior and societal trends. They are essential for understanding and addressing various social issues. By providing quantitative insights, social statistics help researchers, policymakers, and social scientists make evidence-based decisions.

1.4 Data Collection Methods

- Surveys and Questionnaires:
- Online Surveys: Using tools like Google Forms, SurveyMonkey, etc.
- Paper Surveys: Traditional method using printed questionnaires.
- Telephone Surveys: Conducting surveys over the phone.
- Interviews:
- Structured Interviews: Predefined set of questions
- Semi-Structured Interviews: Some predefined questions, but open to exploring new topics.
- Unstructured Interviews: No predefined questions, more of a conversation.
- Observations:
- Participant Observation: Researcher actively participates in the setting.
- Non-Participant Observation: Researcher observes without interacting.
- Focus Groups: Group discussions to gather diverse perspectives.
- Case Studies: In-depth analysis of a single case or a small number of cases.
- Experiments: Controlled studies to determine cause-and-effect relationships.
- Secondary Data Collection: Using existing data from sources like:
- Government publications
- Research articles
- Online databases
- Online Data Collection:
- Web scraping
- Social media analysis
- Document Review: Analyzing existing documents like reports, minutes of meetings, etc.

1.4.1 Data Processing Steps

- Data Entry: Manual entry of data into spreadsheets or databases and automated data entry using software tools.
- Data Cleaning: Removing duplicates; correcting errors and handling missing data
- Data Transformation: Normalizing data, aggregating data and data discretization
- Data Integration: Combining data from different sources and ensuring consistency across datasets
- Data Coding: Assigning codes to qualitative data for analysis.
- Data Analysis: Quantitative Analysis;
- Descriptive Statistics: Mean, median, mode, etc.
- Inferential Statistics: Hypothesis testing, regression analysis, etc.
- Qualitative Analysis:
- Thematic Analysis: Identifying patterns or themes
- Content Analysis: Quantifying the presence of certain words or concepts.
- Narrative Analysis: Examining the story content.
- Data Visualization:
- Creating charts, graphs, and other visual representations of data.
- Data Storage and Management:
- Storing data in databases or data warehouses.
- Using data management tools to organize and maintain data integrity.

1.4.2 Tools and Software for Data Collection and Processing

- Survey Tools: Survey Monkey, Google Forms, Qualtrics
- Statistical Software: SPSS, R, SAS, STATA
- Data Analysis Tools: Excel, Python, MATLAB
- Qualitative Analysis Tools: NVivo, ATLAS.ti
- Database Management: SQL, Microsoft Access, Oracle
- Data Visualization Tools: Tableau, Power BI, D3.js

This list covers the essential methods and steps in data collection and processing, providing a robust framework for gathering and analyzing data effectively.

1.5. Data Analysis in Social Research

1.5.1 Descriptive Statistics in Social Research

Descriptive statistics summarize and describe the features of a dataset. They include measures of central tendency like the mean, median, and mode, which indicate the central point of a data distribution (Field, 2018). Measures of dispersion, such as range, variance, and standard deviation, describe the spread of data points (Pallant, 2020). Descriptive statistics are crucial for initial data analysis and provide a clear overview of social phenomena.
Inferential Statistics in Social Research

1.5.2 Inferential statistics in Social Research

Inferential statistics allow researchers to draw conclusions about a population based on a sample. This involves sampling methods and techniques to ensure representativeness (Creswell, 2014). Hypothesis testing, including the formulation of null and alternative hypotheses, is central to inferential statistics. Researchers use statistical tests to determine the likelihood that their findings are due to chance (Cohen, 1988). Confidence intervals provide a range within which the true population parameter is likely to fall.

1.6 Types of Inferential statistics in Social Research

1.6.1 Parametric Statistics

Regression analysis examines the relationship between variables. Linear regression analyzes the relationship between two variables, while multiple regressions involve more than two predictors (Tabachnick & Fidell, 2019). The regression coefficient indicates the strength and direction of this relationship. Regression analysis is widely used in social research to predict outcomes and understand underlying patterns. Parametric statistics are statistical techniques that assume the data come from a type of probability distribution and involve parameters such as mean and standard deviation.

List of common parametric statistics:

- Mean: The average of a set of values.
- Standard Deviation: A measure of the dispersion or spread of a set of values.
- Variance: the Square of the standard deviation; a measure of the spread of a set of values.
- t-Test:

- One-sample t-test: Tests whether the mean of a single sample differs from a known or hypothesized population mean.
- Independent samples t-test: Compares the means of two independent groups.
- Paired samples t-test: Compares the means of two related groups.
- o Analysis of Variance (ANOVA):
- One-way ANOVA: Tests for significant differences between the means of three or more independent groups.
- Two-way ANOVA: Tests for significant differences between means with two independent variables.
- o Regression Analysis:
- ▪ Simple Linear Regression: Models the relationship between two continuous variables.
- ▪ Multiple Regressions: Models the relationship between one continuous dependent variable and two or more independent variables.
- o Pearson Correlation Coefficient: Measures the strength and direction of the linear relationship between two continuous variables.
- o Confidence Intervals: Provides an estimated range of values which is likely to include the population parameter.
- o Z-Test: Used to determine whether there is a significant difference between sample and population means, or between the means of two samples when the sample size is large.
- o F-Test: Compares the variances of two populations.
- 11. Chi-Square Test (when assumptions of normality are met): Tests for independence or goodness of fit when dealing with categorical data, provided the sample size is sufficiently large.

These parametric statistics assume that the underlying data follow a specific distribution, most commonly the normal distribution. Using these techniques typically requires that the data meet certain conditions or assumptions, such as homogeneity of variance and linearity.

1.6.2 Non-Parametric Methods

Non-parametric methods are used when data do not meet the assumptions required for parametric tests, such as normal distribution. Examples include the Chi-square test, which assesses the association between categorical variables, and the Mann-Whitney U test, which compares differences between two independent groups (Siegel & Castellan, 1988). These methods are versatile and can be applied to various types of social data. Non-parametric statistics are statistical methods that do not assume a specific distribution for the data. These methods are often used when the data does not meet the assumptions required for parametric tests. Here is a **list of some common non-parametric statistical tests:**

- Tests for One Sample
  - Sign Test: Tests whether the median of a single sample is equal to a specified value.
  - Wilcoxon Signed-Rank Test: Compares the median of a single sample to a known value or the medians of two related samples.
- Tests for Two Independent Samples
  - Mann-Whitney U Test (Wilcoxon Rank-Sum Test): Compares differences between two independent groups.
  - Kolmogorov-Smirnov Test: Compares the distribution of two independent samples.
- Tests for Two Related Samples
  - Wilcoxon Signed-Rank Test: Used for comparing two related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks differ.
  - Sign Test: Tests the difference between paired observations.
- Tests for More Than Two Independent Samples
  - Kruskal-Wallis H Test: An extension of the Mann-Whitney U Test for comparing more than two independent groups.
  - Friedman Test: Used for detecting differences in treatments across multiple test attempts.
- Tests for More Than Two Related Samples
  - Friedman Test: A non-parametric alternative to the repeated measures ANOVA.
  - Cochran's Q Test: Used for binary data across multiple trials or conditions.
- Tests for Categorical Data
  - Chi-Square Test: Tests for independence or goodness of fit for categorical data.
  - Fisher's Exact Test: Used for small sample sizes to test the independence of two categorical variables.
- Correlation and Association
  - Spearman's Rank Correlation Coefficient: Measures the strength and direction of the association between two ranked variables.
  - Kendall's Tau: Measures the ordinal association between two measured quantities.
- Survival Analysis
  - Kaplan-Meier Estimate: Estimates the survival function from lifetime data.
  - Log-Rank Test: Compares the survival distributions of two samples.
- Others
  - Median Test: Tests whether two samples come from populations with the same median.

These non-parametric tests are useful when dealing with ordinal data, non-normally distributed data, or when sample sizes are small. They provide robust alternatives to parametric tests that rely on assumptions about the data's distribution.

Self-Assessment Exercise 1

i. Define the term the Social statistics for Public Administration
ii. Outline the Data Collection Methods
iii. Highlight the Tools and Software for Data Collection and Processing

1.7 Use of Statistical Software

Statistical software like SPSS, Stata, and R streamline data analysis processes, allowing researchers to handle large datasets efficiently (Pallant, 2020). These tools offer a range of statistical functions and visualizations, enhancing the analytical capabilities of social researchers.

1.7.1 The steps to run Regression, Chi-square, and T-Test in SPSS:

- Regression Analysis
  - Linear Regression
  - Open SPSS and load your dataset.
  - Click on `Analyze` > `Regression` > `Linear`.
  - Select the dependent variable (the outcome you are predicting) and move it to the Dependent box.
  - Select the independent variables (the predictors) and move them to the Independent(s) box.
  - Optional: Click on `Statistics` to choose additional statistics to display (e.g., R-squared, confidence intervals).
  - Click on `OK` to run the analysis.
  - Multiple Regression
  - Open SPSS and load your dataset.
  - Click on `Analyze` > `Regression` > `Linear`.
  - Select the dependent variable and move it to the Dependent box.
  - Select multiple independent variables and move them to the Independent(s) box.
  - Optional: Click on `Statistics` and select additional statistics you want to include.
  - Click on `OK` to run the analysis.
- Chi-Square Test
  - Chi-Square Test of Independence
  - Open SPSS and load your dataset.
  - Click on `Analyze` > `Descriptive Statistics` > `Crosstabs`.

- Select the variables you want to test for independence and move them to the Rows and Columns boxes.
- Click on `Statistics` and check `Chi-square`.
- Optional: Click on `Cells` to select additional cell display options (e.g., observed, expected values).
- Click on `OK` to run the analysis.

○ T-Test

- Independent Samples T-Test
- Open SPSS and load your dataset.
- Click on `Analyze` > `Compare Means` > `Independent-Samples T Test`.
- Select the dependent variable and move it to the Test Variable(s) box.
- Select the grouping variable and move it to the Grouping Variable box.
- Click on `Define Groups` and specify the values for the groups.
- Click on `Continue` and then `OK` to run the analysis.
- Paired Samples T-Test
- Open SPSS and load your dataset.
- Click on `Analyze` > `Compare Means` > `Paired-Samples T Test`.
- Select the pairs of variables you want to compare and move them to the **Paired Variables box.
- Click on `OK` to run the analysis.
- These steps should help you run Regression, Chi-Square, and T-Test in SPSS efficiently.

Self-Assessment Exercises 2

Outline The steps to run Regression, Chi-square, and T-Test in SPSS

1.8 Summary

In summary, the unit covers a comprehensive range of topics related to social statistics in the field of Public Administration. It begins by defining Social statistics and delves into various aspects such as Data collection methods, data processing Steps, and the utilization of Tools and Software for data collection and Processing. The unit further explores Data Analysis in Social Research, focusing on Descriptive Statistics and Inferential statistics. In the realm of inferential statistics, the unit discusses different types and provides insights into Parametric Statistics, listing common parametric statistical methods. Additionally, Non-Parametric Methods are explored, accompanied by a list of common non-parametric statistical tests. The unit also touches upon the significance of using Statistical Software, demonstrating the steps involved in running

Regression, Chi-square, and T-Test in SPSS. Overall, the unit offers a comprehensive understanding of statistical methods and tools essential for conducting efficient and accurate social research within the realm of Public Administration.

1.9 References/Further Reading/Web Resources

Bryman, A. (2016). Social Research Methods. Oxford University Press.
Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences

1.10 Possible Answers SAEs

Answer to SAEs 1

Q1. Definition of Social statistics for Public Administration

Social statistics for Public Administration involve the collection, analysis, and interpretation of data related to human social behavior and societal trends. They are essential for understanding and addressing various social issues. By providing quantitative insights, social statistics help researchers, policymakers, and social scientists make evidence-based decisions

Q2. Data Collection Methods

○ Surveys and Questionnaires:
- Online Surveys: Using tools like Google Forms, SurveyMonkey, etc.
- Paper Surveys: Traditional method using printed questionnaires.
- Telephone Surveys: Conducting surveys over the phone.

○ Interviews:
- Structured Interviews: Predefined set of questions
- Semi-Structured Interviews: Some predefined questions, but open to exploring new topics.
- Unstructured Interviews: No predefined questions, more of a conversation.

○ Observations:
- Participant Observation: Researcher actively participates in the setting.
- Non-Participant Observation: Researcher observes without interacting.

○ Focus Groups: Group discussions to gather diverse perspectives.

Q3. Tools and Software for Data Collection and Processing

- Survey Tools: Survey Monkey, Google Forms, Qualtrics
- Statistical Software: SPSS, R, SAS, STATA
- Data Analysis Tools: Excel, Python, MATLAB
- Qualitative Analysis Tools: NVivo, ATLAS.ti
- Database Management: SQL, Microsoft Access, Oracle
- Data Visualization Tools: Tableau, Power BI, D3.js

Answer to SAEs 2

Outline the steps to run Regression, Chi-square, and T-Test in SPSS

Use of Statistical Software

Statistical software like SPSS, Stata, and R streamline data analysis processes, allowing researchers to handle large datasets efficiently (Pallant, 2020). These tools offer a range of statistical functions and visualizations, enhancing the analytical capabilities of social researchers.

1.7.1 The steps to run Regression, Chi-square, and T-Test in SPSS:

- Regression Analysis
- Linear Regression
- Open SPSS and load your dataset.
- Click on `Analyze` > `Regression` > `Linear`.
- Select the dependent variable (the outcome you are predicting) and move it to the Dependent box.
- Select the independent variables (the predictors) and move them to the Independent(s) box.
- Optional: Click on `Statistics` to choose additional statistics to display (e.g., R-squared, confidence intervals).
- Click on `OK` to run the analysis.
- Multiple Regression
- Open SPSS and load your dataset.
- Click on `Analyze` > `Regression` > `Linear`.
- Select the dependent variable and move it to the Dependent box.
- Select multiple independent variables and move them to the Independent(s) box.
- Optional: Click on `Statistics` and select additional statistics you want to include.
- Click on `OK` to run the analysis.
- Chi-Square Test
- Chi-Square Test of Independence
- Open SPSS and load your dataset.
- Click on `Analyze` > `Descriptive Statistics` > `Crosstabs`.
- Select the variables you want to test for independence and move them to the Rows and Columns boxes.
- Click on `Statistics` and check `Chi-square`.
- Optional: Click on `Cells` to select additional cell display options (e.g., observed, expected values).
- Click on `OK` to run the analysis.
- T-Test

- Independent Samples T-Test
- Open SPSS and load your dataset.
- Click on `Analyze` > `Compare Means` > `Independent-Samples T Test`.
- Select the dependent variable and move it to the Test Variable(s) box.
- Select the grouping variable and move it to the Grouping Variable box.
- Click on `Define Groups` and specify the values for the groups.
- Click on `Continue` and then `OK` to run the analysis.
- Paired Samples T-Test
- Open SPSS and load your dataset.
- Click on `Analyze` > `Compare Means` > `Paired-Samples T Test`.
- Select the pairs of variables you want to compare and move them to the Paired Variables box.
- Click on `OK` to run the analysis.
- These steps should help you run Regression, Chi-Square, and T-Test in SPSS efficiently.

UNIT 2 POPULATION IN SOCIAL RESEARCH

Unit Structure

2.1 Introduction

In our previous the unit, we discussed the social statistics in the field of Public Administration. It defined Social statistics and delves into various aspects such as Data collection methods, data processing Steps, and the utilization of Tools and Software for data collection and Processing. The unit further explores Data Analysis in Social Research, focusing on Descriptive Statistics and Inferential statistics. The unit will be discussing the concept of population, types of population, the general population, sample, characteristics of Population, importance of population in research and Application in Research

2.2 Learning Outcome

At the end of the unit, you should be able to:

- Define Population
- Identify the Types of Population
- Explain the general population
- Discuss the Target population
- Explain Accessible Population
- Outline the element of Sample
- Highlight the Characteristics of Population
- Explain the Importance of Population in Research
- Explain the Application in Research

2.3 Define Population

A population refers to the complete set of individuals, objects, or data that have some common characteristic defined by the researcher. A population, which is a fundamental concept in research and statistical analysis, encompasses the entire group of individuals, objects, or data points that share a specific trait designated by the researcher. In the realm of research methodology, a population provides the foundation for drawing conclusions and making inferences about a particular phenomenon or issue being studied. It serves as the starting point for data collection, analysis, and interpretation, offering a comprehensive view of the group characteristics under investigation. Understanding the population under scrutiny is crucial, as it shapes the scope and generalizability of research findings. Researchers often define the population based on specific criteria or attributes of interest, such as age, gender, location, or behavior, to ensure that their study results accurately reflect the target group. By defining and delimiting the population, researchers can effectively apply research methods and statistical tools to draw meaningful insights and make informed decisions. The concept of a population extends beyond just a group of individuals; it encompasses a diverse range of elements that form the basis for scientific inquiry and exploration. Therefore, having a clear understanding of the population being studied is essential for conducting rigorous and valid research that contributes to the advancement of knowledge in various domains.

2.4 Types of Population

2.4.1 The general population

The general population refers to the comprehensive assembly of individuals or elements exhibiting characteristics or qualities that have piqued the curiosity of the researcher, encompassing a vast array of diverse subjects who form the focal point of the study and are the subject of investigation in the research project or academic inquiry. The general population typically serves as the foundation for data collection and analysis, embodying a broad spectrum of individuals from various backgrounds, demographics, and contexts, providing a rich and varied source of information that can offer valuable insights and contribute significantly to the researcher's understanding of the subject matter being explored. By including representatives from the general population in the study cohort, researchers can obtain a more comprehensive and holistic perspective on the topic under examination, allowing for a thorough examination of trends, patterns, and relationships within the broader population group, leading to more informed conclusions and robust findings that hold greater validity and generalizability. In essence, the general population serves as the cornerstone of research endeavors, defining the scope and boundaries of the study while offering a diversified and representative sample that reflects the complexities and nuances present within the broader societal fabric, thereby enhancing the rigor and credibility of the research outcomes and paving the way for meaningful contributions to the academic community and beyond.

2.4.2 Target population

The target population refers to a particular subset of individuals or groups within the larger general population that researchers choose to focus on for their studies. This specific demarcation allows for a more precise examination of characteristics, behaviors, or trends within this defined group. By honing in on this targeted segment, researchers can gain deeper insights and draw more accurate conclusions that may not be possible when studying the general population as a whole. Understanding the unique attributes and dynamics of the target population is crucial for ensuring the research findings are applicable and relevant to that specific group, aiding in the development of targeted interventions or strategies to address any identified issues or challenges. Ultimately, the process of selecting and defining the target population plays a pivotal role in shaping the research design, methodology, and interpretation of results, contributing to the overall quality and effectiveness of the research outcomes.

2.4.3 Accessible Population

Accessible Population refers to the subgroup within the broader target population that can be practically reached and included in a study by the researcher. It signifies the specific individuals or groups that are within the easy reach of the researcher for data collection and analysis purposes. The concept of an accessible population is fundamental as it delineates the boundary within which the researcher can effectively investigate and draw valid conclusions in their research. This subset of the target population is crucial for ensuring the feasibility and manageability of a study, as it enables researchers to focus their efforts and resources on a more defined and attainable group of subjects. By identifying and defining the accessible population, researchers can optimize their study design and methodology to generate reliable findings that accurately represent this subset of the target population. In essence, the accessible population acts as the core foundation for researchers to establish the scope and boundaries of their research endeavors, guiding them in selecting the most appropriate strategies for data collection and analysis within the constraints of practicality and accessibility.

2.4.4 Sample

In a research study, a sample refers to a specific group of individuals that is carefully chosen from the larger population to participate in the research. This selection process is a crucial step in ensuring that the findings and conclusions drawn from the study can be applied to the entire population with a certain level of confidence. By selecting a subset of individuals to represent the larger group, researchers aim to gather data that is both meaningful and representative of the characteristics and diversity present within the population. The sample serves as a microcosm of the population, allowing researchers to make inferences and generalizations about the broader group based on the observed behaviors and trends within the sample. The process of selecting the sample involves various considerations, such as randomization, stratification, and sample size, all of which are designed to minimize bias and ensure that the sample accurately reflects the population. Ultimately, the goal of selecting a sample is to obtain reliable and valid data that can be analyzed to draw meaningful conclusions and contribute to the existing body of knowledge within a particular field of study.

2.5. Characteristics of Population

Demographics encompass various essential factors that provide insights into the composition of a population. These factors include age, which can shed light on different life stages and specific needs, gender, influencing societal roles and opportunities, ethnicity, which reflects cultural diversity and traditions, education level, which correlates with skills and knowledge, and occupation, indicating the type of work individuals do. Socio-economic status is another crucial aspect that gives a broader perspective on individuals' well-being and opportunities. Income levels play a significant role in determining the financial stability and access to resources, while employment status reflects the level of job security and stability. Additionally, housing conditions contribute to the quality of life and overall living standards. Geographic distribution further adds to the complexity of understanding populations by considering the settings in which they reside. This can range from urban areas, characterized by dense populations and diverse services, to rural settings, known for their quieter pace and closer community ties. Regional distribution highlights variations in cultural norms and infrastructure availability, while mobility patterns reveal how people move within and across these locations, influencing their daily experiences and opportunities. Overall, these factors paint a comprehensive picture of a population's composition, well-being, and lifestyles, providing valuable insights for research, policy-making, and social interventions.

2.6 Importance of Population in Research

Understanding the population is crucial for several reasons:

1. Basis for Sampling: It allows for the creation of a representative sample that mirrors the characteristics of the population.
2. Ensuring Representativeness: It ensures that the findings can be generalized to the broader population.

Example Scenario

Context: A study on the impact of remote work on employee productivity in the technology sector.

Identification of Population: In the realm of research, it is important to clearly delineate the different segments of the population under study. In this case, the general population consists of all individuals working in the expansive field of technology. Narrowing the focus, the target population specifically comprises those individuals who are remote workers within tech companies. Building on this, the accessible population refers to the subset of remote workers in tech companies situated within a defined geographic area, such as a particular city or region.

Self-Assessment Exercise 1

i.	Define Population
ii.	Outline the types of Population
iii.	Explain the characteristics of Population

1.7 Application in Research

When conducting research, it is crucial for investigators to carefully consider which population they are examining to ensure the validity and generalizability of their findings. In this scenario, researchers would primarily gather data from the accessible population, encompassing remote workers within tech firms in the specified locality. By selecting this subset, researchers aim to have a sample that is representative of the broader target population of remote workers in the tech industry. This approach enables researchers to derive meaningful insights and make informed conclusions regarding the potential influence of remote work setups on employee productivity, thereby contributing to the body of knowledge in this field.

Population in social science research refers to the entire group of individuals or entities to which a researcher intends to generalize the findings of a study. Understanding the application of population in social science research is essential for accurate data collection, analysis, and interpretation. Here are the key aspects:

Application in Research

- o Identifying the Population: Clearly defining the population in terms of characteristics such as age, gender, occupation, geographic location, etc.
- o Sampling: Selecting a representative sample from the population to make the study feasible. Common sampling methods include:
- ▪ Random Sampling: Each member of the population has an equal chance of being selected.
- ▪ Stratified Sampling: The population is divided into subgroups (strata) and random samples are taken from each stratum
- ▪ Cluster Sampling: The population is divided into clusters, some of which are randomly selected, and all members of selected clusters are studied.
- ▪ Convenience Sampling: Selecting a sample based on ease of access and availability.
- o Data Collection: Gathering data from the sample through various methods such as surveys, interviews, observations, or secondary data analysis.
- o Data Analysis: Using statistical methods to analyze the data collected from the sample and make inferences about the population.
- o Generalization: Applying the findings from the sample to the larger population, taking into consideration the sampling method and potential biases.
- ▪ Sampling Bias: Errors that occur when the sample does not accurately represent the population.
- ▪ Sample Size: Larger samples generally provide more accurate and reliable results but require more resources.
- ▪ Ethical Considerations: Ensuring the rights and privacy of the population are respected during research.

Examples

Census: A complete enumeration of the population, such as the national census conducted by governments to gather demographic data.

Surveys: Research studies that use questionnaires to collect data from a sample, such as public opinion polls.

Case Studies: In-depth studies of a single individual or small group, often used to explore complex issues in detail.

Self-Assessment Exercises 2

i.	Describe the Importance of Population in Research
ii.	Explain the Application in Research

1.8 Summary

The application of population in social science research is crucial for designing studies that yield valid and generalizable results. By carefully defining the population, selecting appropriate samples, and using robust data collection and analysis methods, researchers can draw meaningful conclusions that contribute to the understanding of social phenomena.

1.9 References/Further Reading/Web Resources

Creswell, J. W. (2014). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage publications.

Trochim, W. M., & Donnelly, J. P. (2008). Research Methods Knowledge Base. Atomic Dog Publishing.

2.10 Possible Answers SAEs

Answer to SAEs 1

Q1 **Define Population**

A population refers to the complete set of individuals, objects, or data that have some common characteristic defined by the researcher. A population, which is a fundamental concept in research and statistical analysis, encompasses the entire group of individuals, objects, or data points that share a specific trait designated by the researcher.

Q2 Types of Population

The general population

The general population refers to the comprehensive assembly of individuals or elements exhibiting characteristics or qualities that have piqued the curiosity of the researcher,

Target population

The target population refers to a particular subset of individuals or groups within the larger general population that researchers choose to focus on for their studies

Accessible Population

Accessible Population refers to the subgroup within the broader target population that can be practically reached and included in a study by the researcher

Sample

In a research study, a sample refers to a specific group of individuals that is carefully chosen from the larger population to participate in the research. This selection process is a crucial step in ensuring that the findings and conclusions drawn from the study can be applied to the entire population with a certain level of confidence

Q3 Characteristics of Population

Demographics encompass various essential factors that provide insights into the composition of a population. These factors include age, which can shed light on different life stages and specific needs, gender, influencing societal roles and opportunities, ethnicity, which reflects cultural diversity and traditions, education level, which correlates with skills and knowledge, and occupation, indicating the type of work individuals do

Answer to SAEs 2

Q1 Importance of Population in Research

Understanding the population is crucial for several reasons:

1. Basis for Sampling: It allows for the creation of a representative sample that mirrors the characteristics of the population
Ensuring Representativeness: It ensures that the findings can be generalized to the broader population

Q2 Application in Research

When conducting research, it is crucial for investigators to carefully consider which population they are examining to ensure the validity and generalizability of their findings. In this scenario, researchers would primarily gather data from the accessible population, encompassing remote workers within tech firms in the specified locality. By selecting this subset, researchers aim to have a sample that is representative of the broader target population of remote workers in the tech industry. This approach enables researchers to derive meaningful insights and make informed conclusions regarding the potential influence of remote work setups on employee productivity, thereby contributing to the body of knowledge in this field

Application in Research

- Identifying the Population: Clearly defining the population in terms of characteristics such as age, gender, occupation, geographic location, etc.
- Sampling: Selecting a representative sample from the population to make the study feasible. Common sampling methods include:
  - Random Sampling: Each member of the population has an equal chance of being selected.
  - Stratified Sampling: The population is divided into subgroups (strata) and random samples are taken from each stratum
  - Cluster Sampling: The population is divided into clusters, some of which are randomly selected, and all members of selected clusters are studied.

- Convenience Sampling: Selecting a sample based on ease of access and availability.
- Data Collection: Gathering data from the sample through various methods such as surveys, interviews, observations, or secondary data analysis

Data Analysis: Using statistical methods to analyze the data collected from the sample and make inferences about the population

UNIT 3 SAMPLING AND HYPOTHESES IN PUBLIC ADMIN RESEARCH

Unit Structure

3.1 Introduction

In our previous the unit, we discussed the concept of population, types of population, the general population, sample, characteristics of Population, importance of population in research and Application in Research. This unit will be discussing sampling in public administration research.

3.2 Learning Outcome

At the end of the unit, you should be able to:

3.3 concept of Sampling

Sampling is a crucial aspect of social research, as it allows researchers to draw conclusions about a larger population based on data collected from a subset of that population. Let's go through an example scenario to illustrate the concept of sampling in social research. Imagine you are a sociologist interested in studying the job satisfaction levels of employees in a large multinational corporation with 10,000 employees spread across various departments and locations.

3.4 Steps to Conduct Sampling:

Define the Population:

The population in this scenario is all 10,000 employees of the multinational corporation.

Choose a Sampling Method:

There are several sampling methods you can choose from, including simple random sampling, stratified sampling, cluster sampling, and systematic sampling. Let's use **stratified sampling** for this example to ensure representation from different departments.

Stratify the Population:

Divide the population into different strata (subgroups) based on a characteristic that is important to the research. In this case, you might stratify the employees by department (e.g., HR, Sales, IT, Marketing, etc.).

Determine Sample Size:

Decide on the sample size. Let's say you want to survey 500 employees. You will need to determine how many employees to sample from each department based on their proportion in the population.

Randomly Select Samples from Each Stratum:

Within each department, randomly select the required number of employees to include in your sample.

Example Calculation:

Let's assume the corporation has the following distribution of employees across departments:

HR: 1,000 employees
Sales: 3,000 employees
IT: 2,000 employees
Marketing: 4,000 employees
To ensure proportional representation, you would calculate the number of employees to sample from each department as follows:
HR: 1,00010,000×500=5010,0001,000×500=50 employees
Sales: 3,00010,000×500=15010,0003,000×500=150 employees

IT: 2,00010,000×500=10010,0002,000×500=100 employees
Marketing: 4,00010,000×500=20010,0004,000×500=200 employees
Next, you would randomly select 50 employees from HR, 150 from Sales, 100 from IT, and 200 from Marketing.

Conduct the Survey:

Administer a job satisfaction survey to the 500 selected employees. The survey might include questions about their work environment, management, job roles, and overall satisfaction.

3.5. Techniques of sample Size Determination

1. **Cochran's Formula**: Used for large populations to determine a representative sample size, accounting for desired confidence levels and margin of error.

$$n = \frac{Z^2 p(1-p)}{e^2}$$

The formula in the image is used for calculating sample size in statistics, particularly for determining the required sample size for a proportion in a population. Here's the interpretation of each variable in the formula:
n: The sample size needed.

Z: The Z-value (or Z-score) that corresponds to the desired confidence level (e.g., 1.96 for 95% confidence).

p: The estimated proportion of the population (or the expected proportion in the sample).

e: The margin of error (the desired level of precision).
2.
3. **Krejcie and Morgan's Formula**: Provides a pre-calculated table of sample sizes for different population sizes at a 95% confidence level with a 5% margin of error.

$$n = \frac{\chi^2 Np(1-p)}{e^2(N-1) + \chi^2 p(1-p)}$$

The formula in the image is the sample size formula for a finite population. It is used to determine the sample size needed when sampling from a finite population. The formula is given by:

where:

n = is the sample size.

X^2 = (chi-squared) is the critical value of the chi-square distribution with 1 degree of freedom at the desired confidence level.

N = is the population size.

p = is the estimated proportion of an attribute that is present in the population.

e = is the desired level of precision (the margin of error).

1 - p = is the complement of the estimated proportion.

This formula helps in determining how many individuals should be included in a sample to ensure that the sample is representative of the population, taking into account the confidence level and the margin of error.

4. **Yamane's Formula**: Another simple method for determining sample size with a known population.

$$n = \frac{N}{1 + N(e)^2}$$

where:

n = is the sample size.

$(e)^2$ = (chi-squared) is the critical value of the chi-square distribution with 1 degree of freedom at the desired confidence level.

N = is the population size.

p = is the estimated proportion of an attribute that is present in the population.

e = is the desired level of precision (the margin of error).

5. **Sample Size Tables**: Pre-calculated tables based on different confidence levels and margins of error for quick reference.
6. **Rule of Thumb**: Common heuristics such as using a minimum of 30 participants for each subgroup in the study to ensure adequate statistical power.
7.
8. **Expert Judgment**: Relying on the expertise and experience of researchers to estimate an appropriate sample size based on the specific context and objectives of the study.

Each technique has its advantages and limitations, and the choice of method often depends on the study's specific requirements, including the population size, desired confidence level, margin of error, and available resources.

3.6 sampling methods in public administration research

In public administration research, sampling is crucial for obtaining representative data from a larger population. Various sampling techniques can be applied depending on the research objectives, resources, and the nature of the study. Commonly used sampling methods in public administration research along with their applications:

1. Probability Sampling
Probability sampling ensures that every member of the population has a known, non-zero chance of being selected. This type of sampling is often preferred for its ability to produce statistically significant results.

a. Simple Random Sampling
Application: Used when the population is homogeneous and a complete list of the population is available.
Example: Conducting a survey on citizen satisfaction with public services where every resident has an equal chance of being selected.

b. Systematic Sampling
Application: Used when the population is homogeneous and a list of the population is available.
Example: Selecting every 10th employee from a list of public sector employees for a study on job satisfaction.

c. Stratified Sampling

Application: Used when the population is heterogeneous and can be divided into different strata or subgroups.
Example: Studying the impact of administrative reforms across different levels of government employees (e.g., local, state, federal).
d. Cluster Sampling
Application: Used when the population is large and dispersed, making it impractical to sample the entire population.
Example: Conducting a survey on healthcare accessibility by randomly selecting and surveying entire clusters of households within different regions.

2. Non-Probability Sampling
Non-probability sampling does not provide every member of the population with a known chance of being selected. This approach is often used in exploratory research or when probability sampling is impractical.

a. Convenience Sampling
Application: Used when quick and easy data collection is needed, and representativeness is not the primary concern.
Example: Surveying attendees at a public administration conference about their views on recent policy changes.

b. Purposive (Judgmental) Sampling
Application: Used when the researcher wants to target a specific group of individuals who are knowledgeable about the research topic?
Example: Interviewing senior government officials to understand the implementation challenges of a new policy.

c. Snowball Sampling
Application: Used when the population is hard to reach or identify, often in qualitative research.
Example: Studying the experiences of whistleblowers in public administration by asking initial participants to refer other whistleblowers.

d. Quota Sampling
Application: Used to ensure that specific subgroups are adequately represented in the sample.

Example: Ensuring that a survey on public opinion includes a certain number of respondents from different age groups, genders, or ethnicities.

Application in Public Administration Research

Sampling in public administration research helps in making inferences about a population based on a subset of that population. Here are some specific applications:

1. Policy Evaluation:
Sampling Method: Stratified Sampling
Application: Evaluating the impact of a new education policy across different regions and socio-economic groups.

2. Service Delivery Assessment:
Sampling Method: Cluster Sampling
Application: Assessing the quality of public healthcare services by surveying households in selected clusters across different districts.

3. Employee Surveys:
Sampling Method: Simple Random Sampling
Application: Measuring job satisfaction among employees in a government department to identify areas for improvement.

4. Citizen Satisfaction Surveys:
Sampling Method: Systematic Sampling
Application: Collecting data on citizen satisfaction with municipal services by selecting every nth person from a voter registration list.

5. Case Studies:
Sampling Method: Purposive Sampling
Application: Conducting in-depth case studies on the implementation of e-governance initiatives in specific municipalities known for innovative practices.

By carefully selecting the appropriate sampling method, researchers in public administration can ensure that their studies yield reliable and valid results that can inform policy and administrative decisions.

Self-Assessment Exercise 1

1. Define the concept of Sampling
2. Outline the steps to conduct sampling
3. Highlight the techniques of sample size determination

3.7 Sampling and Hypothesis Testing

After collecting the survey responses, analyze the data to determine the overall job satisfaction levels and identify any significant differences between departments.

Make Inferences:

Use the sample data to make inferences about the job satisfaction levels of the entire population of 10,000 employees. For example, you might find that employees in the IT department have higher job satisfaction levels compared to those in the Sales department.

3.7.1 Hypothesis Testing, Confidence Intervals

Scenario:
Imagine you are a social researcher studying the average number of hours people spend on social media per week. You want to test whether the average number of hours spent on social media per week by adults in a particular city is different from the national average, which is known to be 15 hours.

Steps for Hypothesis Testing and Confidence Intervals:

1. Formulate Hypotheses:

- **Null Hypothesis (H0):** The average number of hours spent on social media per week by adults in the city is equal to the national average (15 hours).
- **Alternative Hypothesis (H1):** The average number of hours spent on social media per week by adults in the city is different from the national average (15 hours).

2. Collect Data:

- You conduct a survey and collect data from a random sample of 100 adults in the city. The sample mean ($\bar{X}$) is found to be 16 hours, and the sample standard deviation (s) is 5 hours.

3. Choose a Significance Level:

- Let's choose a significance level (α) of 0.05.

4. Perform Hypothesis Testing:

- We will use a t-test for the hypothesis testing since the population standard deviation is unknown and the sample size is relatively large.

Calculate the Test Statistic:
The formula for the t-statistic is:
$t=\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$
where:

- $\bar{X}$ = sample mean
- μ = population mean (national average)
- s = sample standard deviation
- n = sample size

Plugging in the values:

t=16−155100=1510=10.5=2t=100516−15=1051=0.51=2

Determine the Degrees of Freedom:

df=n−1=100−1=99df=n−1=100−1=99

Find the Critical Value:

Using a t-distribution table, the critical value for df=99df=99 at α=0.05α=0.05 (two-tailed) is approximately ±1.984.

Compare the Test Statistic to the Critical Value:

Since the calculated t-statistic (2) is greater than the critical value (1.984), we reject the null hypothesis. This suggests that there is a significant difference between the average number of hours spent on social media per week by adults in the city and the national average.

5. Construct a Confidence Interval:

To construct a 95% confidence interval for the population mean, we use the following formula:

Confidence Interval=$\bar{X}$±$t\alpha$/2(sn)Confidence Interval=$\bar{X}$±$t\alpha$/2(ns)

where $t\alpha$/2$t\alpha$/2 is the critical value from the t-distribution for a 95% confidence level.

Plugging in the values:

Confidence Interval=16±1.984(5100)=16±1.984(510)=16±1.984×0.5=16±0.992Confidence Interval=16±1.984(1005)=16±1.984(105)=16±1.984×0.5=16±0.992

So, the 95% confidence interval for the population mean is approximately 15.00815.008 to 16.99216.992 hours.

Interpretation:

- **Hypothesis Testing:** We rejected the null hypothesis, indicating that the average number of hours spent on social media per week by adults in the city is significantly different from the national average of 15 hours.

- **Confidence Interval:** We are 95% confident that the true average number of hours spent on social media per week by adults in the city lies between 15.008 and 16.992 hours.

3.7.2 Population and Sample

The population is the entire group under study, while a sample is a subset of the population used to make inferences.

Hypothesis Testing: This involves testing an assumption regarding a population parameter. Researchers use sample data to test hypotheses.

Confidence Intervals: These provide a range of values within which the population parameter is expected to lie, with a certain level of confidence.

2. Types of Inferential Statistics

Estimation:
Point Estimation: Provides a single value estimate of a population parameter (e.g., mean).

Interval Estimation: Provides a range within which the parameter is expected to lie, expressed as a confidence interval.

Hypothesis Testing:

- Null Hypothesis (H0): Assumes no effect or difference.
- Alternative Hypothesis (H1): Assumes there is an effect or difference.
- Type I Error: Rejecting the null hypothesis when it is true.
- Type II Error: Failing to reject the null hypothesis when it is false.
- P-Value: Probability of obtaining a result at least as extreme as the one observed, assuming the null hypothesis is true.
- Parametric and Non-Parametric Tests:
- Parametric Tests: Assume underlying statistical distributions (e.g., t-test, ANOVA).

Non-Parametric Tests: Do not assume specific distributions (e.g., Chi-square test, Mann-Whitney U test).

3. Examples of Inferential Statistics in Social Research
Survey Analysis: Researchers use inferential statistics to generalize survey results to the broader population. For example, using a sample to estimate the average income of a country's citizens.

Experimental Studies: Inferential statistics help in determining the effectiveness of interventions by comparing treatment and control groups.
Observational Studies: These studies rely on inferential statistics to understand relationships between variables in naturally occurring settings.

4. Use of Statistical Software in Social Research

3.7.3 Introduction to Statistical Software

SPSS: Widely used for its user-friendly interface and robust statistical analysis capabilities.

Stata: Known for its powerful data management and statistical features, popular among social scientists.

R: An open-source programming language offering extensive statistical and graphical techniques.

Benefits of Using Software in Social Research
Efficiency: Software automates complex calculations, saving time and reducing errors.

Accuracy: Ensures precise computations and consistency in analyses.
Advanced Analysis: Facilitates sophisticated statistical techniques that might be cumbersome to perform manually.

- **Basic Operations and Analysis**

Data Entry and Management: Importing, cleaning, and organizing data.
Descriptive Statistics: Summarizing data through measures of central tendency and dispersion.
Inferential Analysis: Conducting hypothesis tests, regression analysis, and other inferential procedures.

3.7.4 Types of Errors

1. Type I Error (α): Occurs when the null hypothesis is rejected when it is actually true. It is a false positive. For example, concluding that a new drug is effective when it is not.
2. Type II Error (β): Occurs when the null hypothesis is not rejected when it is actually false. It is a false negative. For example, concluding that a new drug is not effective when it actually is.

Confidence Intervals

A confidence interval is a range of values that is used to estimate the true value of a population parameter. It provides an interval within which we can be certain the population parameter lies, to a certain level of confidence (usually 95%).

For example, if a 95% confidence interval for the average height of adult men is 5.6 to 5.9 feet, we are 95% confident that the true average height of the population lies within this range.

Application in Generalizing Social Data

Inferential statistics are essential in social research for generalizing findings from a sample to a larger population. For example, a researcher studying the impact of education on income levels may use a sample of individuals from different educational backgrounds and apply inferential statistics to generalize the findings to the entire population.

By using techniques such as hypothesis testing and confidence intervals, researchers can make informed decisions and predictions about social phenomena, ensuring that their conclusions are not merely due to random chance.

Hypothesis Testing (Null and Alternative Hypotheses)
Hypothesis testing is a statistical method used to make decisions about the population based on sample data. It involves two hypotheses:

1. **Null Hypothesis (H_0)**: This is a statement of no effect or no difference. It is the hypothesis that the researcher tries to disprove.
   - Example: H_0: The mean weight of apples is 150 grams.
2. **Alternative Hypothesis (H_1 or Ha)**: This is a statement that indicates the presence of an effect or a difference. It is what the researcher wants to prove.
   - Example: H_1: The mean weight of apples is not 150 grams.

Types of Errors (Type I and Type II)

In hypothesis testing, two types of errors can occur:

1. **Type I Error (α)**: This occurs when the null hypothesis is true, but we incorrectly reject it. It is also known as a "false positive."
   - Example: Concluding that a new drug is effective when it is not.
2. **Type II Error (β)**: This occurs when the null hypothesis is false, but we fail to reject it. It is also known as a "false negative."
   - Example: Concluding that a new drug is not effective when it actually is.

Confidence Intervals

A confidence interval is a range of values, derived from the sample data, that is likely to contain the value of an unknown population parameter. The confidence level (usually expressed as a percentage, such as 95% or 99%) indicates the probability that the interval contains the parameter.

- **Formula for a Confidence Interval for a Population Mean**:

CI=$\bar{x}$±(σn)CI=$\bar{x}$±z($n\sigma$)
where:

- $\bar{x}\bar{x}$ is the sample mean
- zz is the z-score corresponding to the desired confidence level
- $\sigma\sigma$ is the population standard deviation
- nn is the sample size

- **Example**: If you have a sample mean of 100, a population standard deviation of 15, a sample size of 30, and you want a 95% confidence interval, you would use the z-score for 95% confidence (which is approximately 1.96) to calculate the interval.

Self-Assessment Exercises 2

Q1. Techniques of sample Size Determination Q2. sampling methods in public administration research

3.8 Summary

This unit examined in-depth explanation of the concept of sampling, covering its fundamental principles and significance in research practice. Additionally, it explored the essential Steps to Conduct Sampling, emphasizing the meticulous process of selecting representative samples to draw valid and reliable conclusions. The unit also delved into the Techniques of Sample Size Determination, shedding light on the factors influencing the appropriate sample size for research studies.

Furthermore, it discussed various Sampling methods commonly employed in public administration research, elucidating the strengths and limitations of each approach. The unit also addressed the crucial link between Sampling and Hypothesis Testing, demonstrating how sampling methods directly influence the validity and accuracy of hypothesis testing outcomes. Moreover, it provided a comprehensive overview of Hypothesis Testing and Confidence Intervals, outlining their significance in inferential statistics and research interpretation.

In addition, the unit explored the relationship between Population and Sample, highlighting the key distinctions and considerations when generalizing findings from a sample to a larger population. It also introduced students to Statistical Software commonly used for data analysis, emphasizing the importance of leveraging technological tools for effective and efficient statistical computations. Finally, the unit touched upon the Types of Errors that researchers may encounter in statistical analysis, emphasizing the need for rigorous validation and verification processes to mitigate errors and maintain the integrity of study findings.

3.9 References/Further Reading/Web Resources

Agresti, A., & Finlay, B. (2017). Statistical Methods for the Social Sciences. Pearson.

Babbie, E. (2016). The Practice of Social Research. Cengage Learning.

Bryman, A. (2012). Social Research Methods. Oxford University Press.

Creswell, J. W., & Creswell, J. D. (2017). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. SAGE Publications.

Creswell, J. W., & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage publications.

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics. Sage.

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics. SAGE Publications.

Gravetter, F. J., & Wallnau, L. B. (2016). Statistics for the Behavioral Sciences. Cengage Learning.

Gujarati, D. N., & Porter, D. C. (2021). Basic Econometrics. McGraw-Hill Education.

Hinton, P. R. (2014). Statistics Explained: A Guide for Social Science Students. Routledge.

Trochim, W. M. K. (2006). Research Methods Knowledge Base. Atomic Dog Publishing.

Trochim, W. M. K. (2020). Research Methods: The Essential Knowledge Base. Cengage Learning.

1.10 Possible Answers SAEs

Answer to SAEs 1

Q1. Concept of Sampling

Sampling is a crucial aspect of social research, as it allows researchers to draw conclusions about a larger population based on data collected from a subset of that population. Let's go through an example scenario to illustrate the concept of sampling in social research

Q2. Steps to Conduct Sampling

- Define the Population:
- Choose a Sampling Method
- stratified sampling
- Stratify the Population
- Determine Sample Size:
- Randomly Select Samples **from Each Stratum**

Answer to SAEs 2

Q1. Techniques of sample Size Determination

Cochran's Formula: Used for large populations to determine a representative sample size, accounting for desired confidence levels and margin of error.

Krejcie and Morgan's Formula: Provides a pre-calculated table of sample sizes for different population sizes at a 95% confidence level with a 5% margin of error

Yamane's Formula: Another simple method for determining sample size with a known population

9. **Sample Size Tables**: Pre-calculated tables based on different confidence levels and margins of error for quick reference.
10. **Rule of Thumb**: Common heuristics such as using a minimum of 30 participants for each subgroup in the study to ensure adequate statistical power.
11. **Expert Judgment**: Relying on the expertise and experience of researchers to estimate an appropriate sample size based on the specific context and objectives of the study

Q2. Sampling methods in public administration research

In public administration research, sampling is crucial for obtaining representative data from a larger population. Various sampling techniques can be applied depending on the research objectives, resources, and the nature of the study. Commonly used sampling methods in public administration research along with their applications:

1. Probability Samplin
Simple Random Sampling
Systematic Sampling
. Stratified Sampling
Cluster Sampling
Non-Probability Sampling
Convenience Sampling
Purposive (Judgmental) Sampling
Snowball Sampling
Quota Sampling
Application in Public Administration Research
Sampling in public administration research helps in making inferences about a population based on a subset of that population. Here are some specific applications:

1. Policy Evaluation:
Sampling Method: Stratified Sampling
Application: Evaluating the impact of a new education policy across different regions and socio-economic groups.

2. Service Delivery Assessment:
Sampling Method: Cluster Sampling
Application: Assessing the quality of public healthcare services by surveying households in selected clusters across different districts.

3. Employee Surveys:
Sampling Method: Simple Random Sampling
Application: Measuring job satisfaction among employees in a government department to identify areas for improvement.

4. Citizen Satisfaction Surveys:
Sampling Method: Systematic Sampling
Application: Collecting data on citizen satisfaction with municipal services by selecting every nth person from a voter registration list.

5. Case Studies:
Sampling Method: Purposive Sampling

Application: Conducting in-depth case studies on the implementation of e-governance initiatives in specific municipalities known for innovative practices

UNIT4: DESCRIPTIVE STATISTICS

Unit Structure

4.1 Introduction

In our previous unit, we discussed sampling, covering its fundamental principles and significance in research practice, essential steps to conduct sampling, techniques of Sample Size Determination, sampling methods commonly employed in public administration research, sampling and Hypothesis Testing, hypothesis testing and confidence intervals. In this unit, we will be discussing descriptive statistics, measures of Central Tendency, mean, median, and mode for both ungrouped and grouped data

4.2 Learning Outcome

At the end of the unit, you should be able to:

- Define Descriptive statistics
- Calculate the Measures of Central Tendency
- Analyze the Variance
- Calculate the Standard Deviation
- Compute the Interquartile Range (IQR)

4.3 define Descriptive statistics

Descriptive statistics involve summarizing and organizing data so that it can be easily understood. This type of statistical analysis helps in presenting quantitative descriptions in a manageable form

4.3.1 Calculate the Measures of Central Tendency

Calculate the mean, median, and mode for both ungrouped and grouped data.

Ungrouped Data
Let's consider the following dataset of test scores for a class of 15 students:
Scores: 55, 67, 45, 67, 89, 67, 78, 90, 45, 67, 55, 78, 90, 45, 67

Mean
To calculate the mean, sum all the scores and divide by the number of scores.

Mean=∑ScoresNumber of Scores=55+67+45+67+89+67+78+90+45+67+55+78+90+45+6715=100115≈66.73Mean=Number of Scores∑Scores=1555+67+45+67+89+67+78+90+45+67+55+78+90+45+67=151001≈66.73

Median
To find the median, arrange the scores in ascending order and find the middle value.

Sorted Scores: 45, 45, 45, 55, 55, 67, 67, 67, 67, 67, 78, 78, 89, 90, 90
Since there are 15 scores, the median is the 8th value.
Median=67Median=67

Mode

The mode is the value that appears most frequently.

Mode=67(appears 6 times) Mode=67(appears 6 times)
Grouped Data
Now, let's group the same dataset into intervals and calculate the mean, median, and mode.
Grouped Data Intervals

Interval	Frequency (f)	Midpoint (x)
40-49	3	44.5
50-59	2	54.5
60-69	6	64.5
70-79	2	74.5
80-89	1	84.5
90-99	2	94.5

Mean

To calculate the mean for grouped data, use the formula:
Mean=∑(𝑓⋅𝑥)∑𝑓Mean=∑f∑(f⋅x)
Mean=(3⋅44.5)+(2⋅54.5)+(6⋅64.5)+(2⋅74.5)+(1⋅84.5)+(2⋅94.5)15=133.5 +109+387+149+84.5+18915=105215≈70.13Mean=15(3⋅44.5)+(2⋅54.5) +(6⋅64.5)+(2⋅74.5)+(1⋅84.5)+(2⋅94.5) =15133.5+109+387+149+84.5+189=151052≈70.13

Median

To find the median class, locate the class interval that contains the cumulative frequency that is half of the total frequency (N/2 = 15/2 = 7.5). Cumulative frequencies:

Interval	Frequency (f)	Cumulative Frequency (cf)
40-49	3	3
50-59	2	5
60-69	6	11
70-79	2	13
80-89	1	14
90-99	2	15

The median class is 60-69.

Use the formula for the median of grouped data:

Median=𝐿+(N/2−𝐶𝐹𝑓)⋅ℎ
Where:
L = lower boundary of the median class = 59.5
N = total number of frequencies = 15
CF = cumulative frequency before the median class = 5
f = frequency of the median class = 6
h = class interval width = 10

Median=59.5+(7.5−56)⋅10=59.5+(2.56)⋅10=59.5+4.17=63.67Median=59.5+(67.5−5)⋅10=59.5+ (62.5)⋅10=59.5+4.17=63.67

Mode

To find the mode, identify the modal class (the class with the highest frequency), which is 60-69.

Use the formula for the mode of grouped data:
Mode=𝐿+(𝑓1−𝑓02𝑓1−𝑓0−𝑓2)⋅ℎ Mode=L+(2f1−f0−f2f1−f0)⋅h

Where:

L = lower boundary of the modal class = 59.5
f1 = frequency of the modal class = 6
f0 = frequency of the class before the modal class = 2
f2 = frequency of the class after the modal class = 2
h = class interval width = 10
Mode=59.5+(6−22·6−2−2)·10=59.5+(412−4)·10=59.5+(48)·10=59.5+5 =64.5Mode=59.5+(2·6−2−26−2)·10=59.5+(12−44)·10=59.5+(84)·10=59.5+5=64.5.

Measures of dispersion are statistical tools used to describe the spread or variability within a set of data. Here are the key measures of dispersion along with their formulas, examples, and graphs:

Range

Formula: Range = Maximum Value – Minimum Value

Example: Consider the data set: 3,7,8,5,12,14,21,13,183,7,8,5,12,14,21,13,18 Range=21−3=18Range=21−3=18

Self-Assessment Exercise 1

Q1. Define Descriptive statistics
Q2. Consider the following dataset of test scores for a class of 15 students:
Scores: 55, 67, 45, 67, 89, 67, 78, 90, 45, 67, 55, 78, 90, 45, 67
Q3. Consider the following set of data

4.4 Variance

Group Variance
Formula: $s^2=\sum(xi-\bar{x})2$
$n-1$

Formula: For a population: $\sigma 2=\sum(xi-\mu)2N\sigma 2=N\sum(xi-\mu)2$
For a sample: $S^2=\sum(xi-\bar{x})2n-1$ $S^2=n-1\sum(xi-\bar{x})2$
Where:

xi = each value in the data set
μ = population mean
$\bar{x}$ = sample mean
N = number of data points in the population
n = number of data points in the sample

Example: Consider the sample data set: 4,8,6,5,34,8,6,5,3 x^-=4+8+6+5+35=5.2x‾=54+8+6+5+3 =5.2 s2=(4−5.2)2+(8−5.2)2+(6−5.2)2+(5−5.2)2+(3−5.2)25−1=1.44+7.84+0.64+0.04+4.844=3.2s2=5−1(4−5.2)2+(8−5.2)2+(6−5.2)2+(5−5.2)2+(3−5.2)2=41.44+7.84+0.64+0.04+4.84=3.2

4.5 Standard Deviation

Standard deviation is a measure of the amount of variation or dispersion in a set of values. It tells you how many the values in a dataset deviate from the mean (average) of the dataset.

Formula for Standard Deviation

For a sample: s=1n−1∑i=1(xi−x^-)2s=n−11∑i=1n(xi−x^-)2
For a population: σ=1N∑i=1(xi−μ)2σ=N1∑i=1N(xi−μ)2
Where:
is the sample standard deviation
σ is the population standard deviation
n is the number of observations in the sample
N is the number of observations in the population
x is each individual observation
x^- is the sample mean
is the population mean
Example

Let's calculate the sample standard deviation for the dataset: 2, 4, 4, 4, 5, 5, 7, 9.

1. Calculate the mean (x^-x^-): x^-=2+4+4+4+5+5+7+9/8=5
2. Calculate each deviation from the mean, square it, and sum them up: $(2−5)^2+(4−5)^2+(4−5)^2$+(4−5)2+(5−5)2+(5−5)2+(7−5)2+(9−5)2(2−5)+$(4−5)^2$+(4−5)+(4−5)2+(5−5)2+(5−5)2+(7−5)2+(9−5)2 =9+1+1+1+0+0+4+16=32=9+1+1+1+0+0+4+16=32
3. Divide by the number of observations minus one (n-1): 32/7≈ 4.57
4. Take the square root: s=4.57 ≈ 2.14

So, the sample standard deviation is approximately 2.14.

Group Standard Deviation

Group standard deviation is used when you have data that is grouped into different categories or classes. The formula takes into account the frequency of each group.

Formula for Group Standard Deviation

$s=(xi-\bar{x})2$
Where:
Example
Let's say we have the following grouped data:

Group	Midpoint ($xixi$)	Frequency ($fifi$)
1-3	2	5
4-6	5	8
7-9	8	7

Calculate the mean ($\bar{x}$): $\bar{x}$

$$(\bar{\mathbf{x}}):\ \bar{\mathbf{x}} = \sum_{i}^{n} \mathrm{Fi}(\mathrm{xi} - \bar{\mathrm{x}})\,2$$

$\bar{x}$ = (2×5) + (5×8) + (8×7)/20
=10+40+56/20=5.3
N is the total number of observations
k is the number of groups
fi is the frequency of the ii-th group
xi is the midpoint of the ii-th group
$\bar{x}$ is the mean of the grouped data

1. Calculate each deviation from the mean, square it, multiply by the frequency, and sum them up: $\sum i=1kfi(xi-\bar{x})^2$ =5(2−5.3)2+8(5−5.3)2+7(8−5.3)2
2. $\sum i=1kfi(xi-\bar{x})2=$ 5(2−5.3)2+8(5−5.3)2+7(8−5.3)2 =5(10.89)+8(0.09)+7(7.29)=5(10.89)+ 8(0.09)+7(7.29) =54.45+0.72+51.03=106.2=54.45+0.72+51.03=106.2
3. Divide by the total number of observations minus one (N-1): 106.219≈5.5919106.2≈5.59
4. Take the square root: s=5.59≈2.36s=5.59≈2.36
5. So, the group standard deviation is approximately 2.36

4.6 Interquartile Range (IQR)

The Interquartile Range (IQR) is a measure of statistical dispersion, or how spread out the values in a data set is. It is calculated as the difference between the third quartile (Q3) and the first quartile (Q1).

Example: Consider the data set: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19
Arrange the data in ascending order (already done).
Find Q1 (the median of the first half of the data): 5
Find Q3 (the median of the second half of the data): 15
Calculate IQR: $IQR=Q3-Q1$=15−5=10IQR=Q3−Q1=15−5=10

Self-Assessment Exercise 2

Q1. State the Formula for Standard Deviation

Q2. Calculate the sample standard deviation for the dataset: 2, 4, 4, 4, 5, 5, 7, 9. Q3. Calculate the following Group Standard Deviation using the above data

4.7 Summary

In this unit, we delved into Descriptive statistics, which are fundamental in summarizing and analyzing data. We explored Measures of Central Tendency, such as mean, median, and mode, which provide insights into the central or average value of a dataset. Additionally, we covered Standard Deviation, a crucial measure of variability that quantifies the spread of data points around the mean. Understanding how to interpret and analyze these statistical measures is essential for drawing meaningful conclusions from data sets and making informed decisions based on statistical insights. By mastering these concepts, you will be better equipped to analyze data effectively and identify patterns or trends that can inform decision-making processes in various fields, from public administration, business and finance to healthcare and social sciences.

4.8 References/Further Reading/Web Resources

Zakari, M. (2022). National Open University of Nigeria course material for Quantitative Methods for Public Administration. www.nou.edu.ng

4.9 Possible Answers SAEs

Answer to SAEs 1

Q1. Define Descriptive statistics
Descriptive statistics involve summarizing and organizing data so that it can be easily understood. This type of statistical analysis helps in presenting quantitative descriptions in a manageable form
Calculate the mean, median, and mode for both ungrouped and grouped data.

Q2. Consider the following dataset of test scores for a class of 15

students:
Scores: 55, 67, 45, 67, 89, 67, 78, 90, 45, 67, 55, 78, 90, 45, 67

Mean
To calculate the mean, sum all the scores and divide by the number of scores.

Mean=∑ScoresNumber of Scores=55+67+45+67+89+67+78+90+45+67+55+78+90+45+6715=100115≈66.73Mean=Number of Scores∑Scores=1555+67+45+67+89+67+78+90+45+67+55+78+90+45+67=151001≈66.73

Mode
The mode is the value that appears most frequently.
Mode=67(appears 6 times) Mode=67(appears 6 times)
Grouped Data

Now, let's group the same dataset into intervals and calculate the mean, median, and mode.
Grouped Data Intervals

Interval	Frequency (f)	Midpoint (x)
40-49	3	44.5
50-59	2	54.5
60-69	6	64.5
70-79	2	74.5
80-89	1	84.5
90-99	2	94.5

Q3. Consider the following set of data
set: 3,7,8,5,12,14,21,13,183,7,8,5,12,14,21,13,18 Range=21−3=18Range=21−3=18

Standard Deviation
Standard deviation is a measure of the amount of variation or dispersion in a set of values. It tells you how many the values in a dataset deviate from the mean (average) of the dataset.

Answer to SAEs 2

Q1. State the Formula for Standard Deviation
For a sample: s=1n−1∑i=1(xi−x¯)2s=n−11∑i=1n(xi−x¯)2

Q2. Calculate the sample standard deviation for the dataset: 2, 4, 4, 4, 5, 5, 7, 9.
Calculate the mean (x¯x¯): x¯=2+4+4+4+5+5+7+9/8=5
Calculate each deviation from the mean, square it, and sum them up: $(2-5)^2+(4-5)^2+(4-5)^2$+(4−5)2+(5−5)2+(5−5)2+(7−5)2+(9−5)2(2−5)+$(4-5)^2$+(4−5)+(4−5)2+(5−5)2+(5−5)2+(7−5)2+(9−5)2 =9+1+1+1+0+0+4+16=32=9+1+1+1+0+0+4+16=32
Divide by the number of observations minus one (n-1): 32/7≈ 4.57
Take the square root: s=4.57 ≈ 2.14
So, the sample standard deviation is approximately 2.14.

Group Standard Deviation
Group standard deviation is used when you have data that is grouped into different categories or classes. The formula takes into account the frequency of each group.

Formula for Group Standard Deviation
s=(xi−x¯)2
Where:

Q3. Calculate the following Group Standard Deviation using the above data

Group	Midpoint ($xixi$)	Frequency ($fifi$)
1-3	2	5
4-6	5	8
7-9	8	7

Calculate the mean (x¯): x¯

$$(\bar{\mathbf{x}}):\ \bar{\mathbf{x}} = \sum_{i}^{n} \mathrm{Fi}(\mathrm{xi} - \bar{\mathrm{x}})\ 2$$

x¯ = (2×5) + (5×8) + (8×7)/20

=10+40+56/20=5.3
N is the total number of observations
k is the number of groups
fi is the frequency of the ii-th group
xi is the midpoint of the ii-th group
$\bar{x}$ is the mean of the grouped data

6. Calculate each deviation from the mean, square it, multiply by the frequency, and sum them up: $\sum i=1kfi(xi-\bar{x})^2$ =5(2−5.3)2+8(5−5.3)2+7(8−5.3)2
7. $\sum i=1kfi(xi-\bar{x})2=$ 5(2−5.3)2+8(5−5.3)2+7(8−5.3)2 =5(10.89)+8(0.09)+7(7.29)=5(10.89)+8(0.09)+7(7.29) =54.45+0.72+51.03=106.2=54.45+0.72+51.03=106.2
8. Divide by the total number of observations minus one (N-1): 106.219≈5.5919106.2≈5.59
9. Take the square root: s=5.59≈2.36s=5.59≈2.36
10. So, the group standard deviation is approximately 2.36

UNIT 5: FREQUENCY DISTRIBUTION

5.1 Introduction

In our previous unit, we discussed descriptive statistics, measures of Central Tendency, mean, median, and mode for both ungrouped and grouped data. In this unit, we will be discussing various essential topics in the realm of frequency distribution. It dives into the creation and interpretation of frequency tables, which provide a structured way to organize and present data. Additionally, it explores the visualization of data through frequency polygons, offering a graphical representation of the50istributeon. Histograms are also studied to depict the frequency of data values in intervals, allowing for a clear understanding of the data's distribution pattern. The unit delves into cumulative frequency distribution, which showcases the running total of frequencies as data points increase. Students will also learn about Ogives, which present cumulative frequencies in a more visual format. Furthermore, the exploration extends to pie charts, bar charts, and line charts, providing students with a comprehensive understanding of how to effectively represent and analyze data using different graphical techniques. These visual aids help in highlighting patterns, trends, and comparisons within datasets, enabling a deeper insight into the information being presented. Through practical application and theoretical understanding, students will gain a strong foundation in frequency distribution methods and charting techniques.

5.2 Learning Outcome

At the end of the unit, you should be able to:

- **Define** frequency distribution
- Graphical Representation of Histogram, Frequency polygon, Cumulative frequency distribution An Ogive
- Explain the Relative distributions
- Graphical Representation of Pie charts, bar charts and line charts
- Graphical Representation of Bar charts
- Graphical Representation of Simple bar chart
- Graphical Representation of A component or stacked bar chart
- Graphical Representation of Multiple Bar Charts
- Explain the demonstrate the Scatter diagrams
- Explain the demonstrate the Skewness and Kurtosis

5.3 Define frequency distribution

A frequency distribution is a summary of how often different values occur within a data set. It can be presented in a table format.

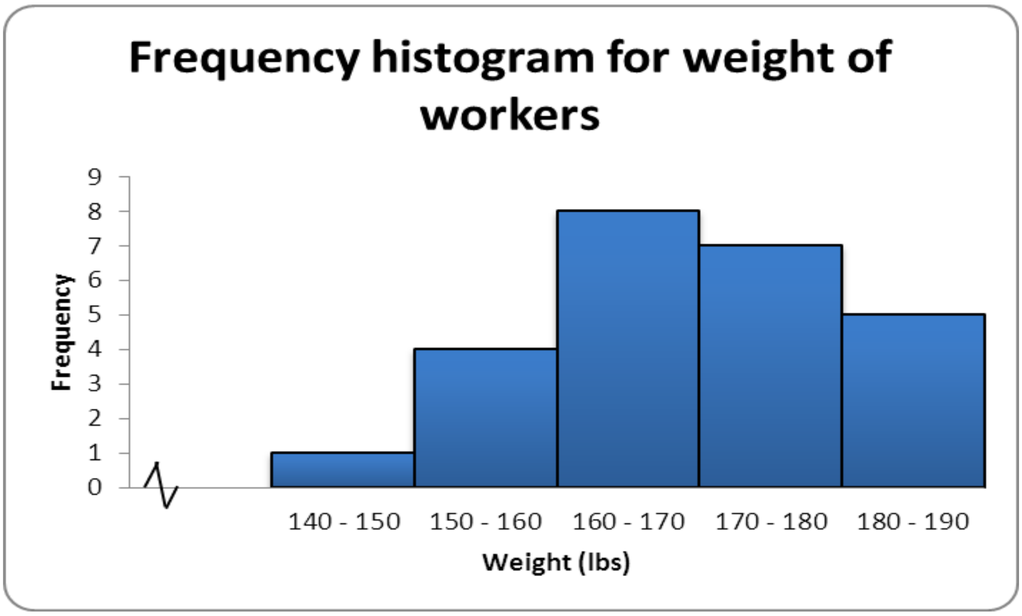
Example: Consider the data set: 1, 2, 2, 3, 3, 3, 4, 4, 4, 4

Value	Frequency
1	1
2	2
3	3
4	4
5	1
7	2
8	3
9	2
10	1

5.4 A Graphical Representation of the Distribution

5.4.1 Histogram

A frequency distribution can be graphically depicted as a histogram. Using the data from the example in section 3.1, the histogram can be depicted as: Remember that histograms are similar to bar charts, but the bars touch each other. If you graph data and part of an axis is not to scale (in example the x-axis from 0 to 140 is not to scale), show a 'broken' axis.



5.4.2 Frequency polygon

A frequency polygon is constructed by plotting the frequency of each interval above the midpoint of that interval and then joining the points with straight lines. The polygon is closed by considering one additional interval (with zero frequency) at each end of the distribution and extending a straight line to the midpoint of each of these intervals.
Before constructing a frequency polygon, calculate the midpoints for each interval.

Using the data from the example in section 1.3.1, the midpoints are calculated as:

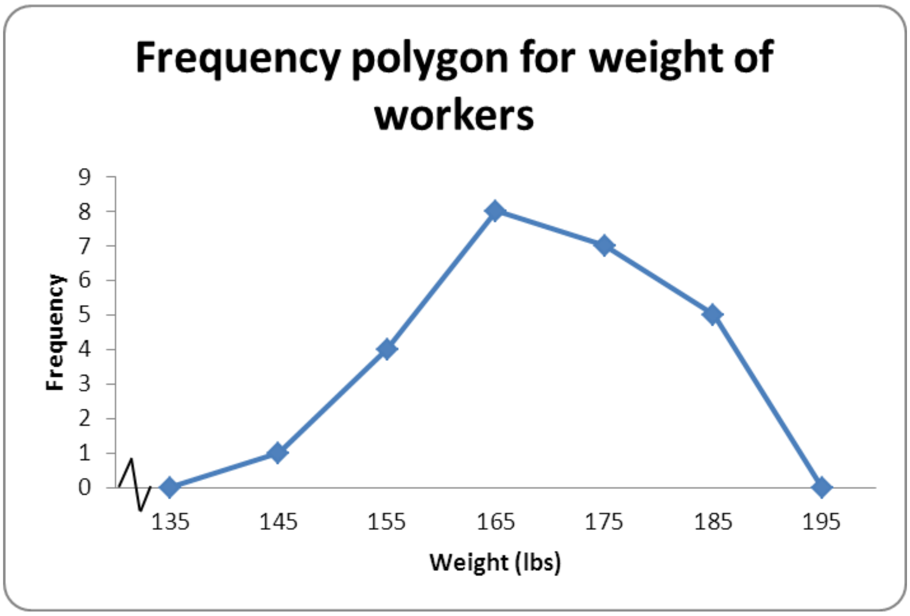
Interval (weight in lbs)	Frequency	Midpoint
140 – 150	1	145
150 – 160	4	155
160 – 170	8	165
170 – 180	7	175
180 – 190	5	185

An easy way to calculate midpoint of an interval is to halve sum of the lower and upper limits. In the case of example, for the first interval:

$$\frac{\text{Lower limit + Upper Limit}}{2} \quad \frac{140+150}{2} = 145$$

$$\frac{lower\ limit + upper\ limit}{2} = \frac{140 + 150}{2} = 145$$

The frequency polygon can be depicted as:



5.4.3 Cumulative frequency distribution

A cumulative frequency distribution summarises the cumulative frequency of a dataset. It results in a 'running total' of frequencies.
EXAMPLE

Using the data from the example in section 1.3.1:

For each interval, calculate the cumulative frequency by adding the frequency count of the interval in question to the cumulative frequency of the interval before.

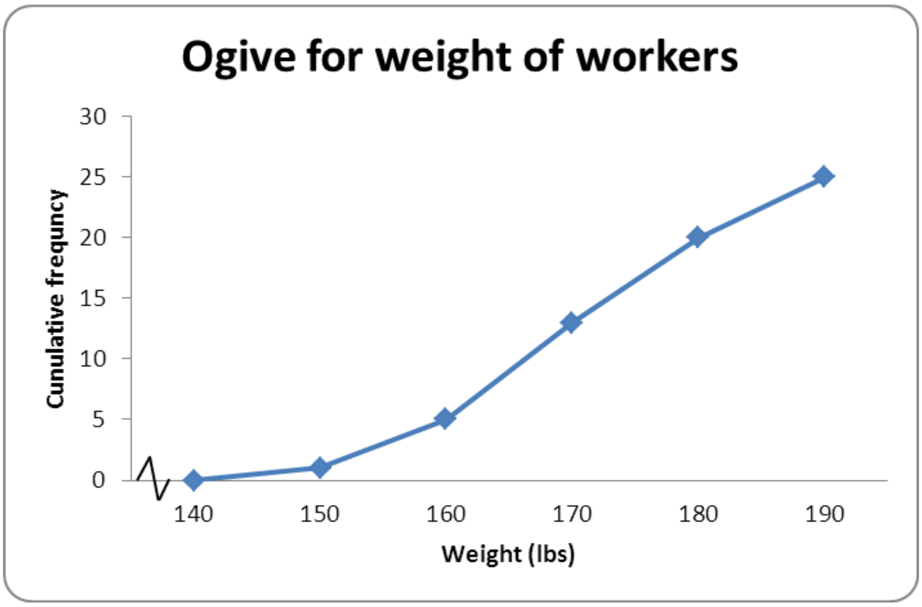
Interval (weight in lbs)	Frequency	Cumulative frequency
140 – 150	1	1
150 – 160	4	5
160 – 170	8	13
170 – 180	7	20
180 – 190	5	25

5.4.4 An Ogive

An ogive is a graph of the cumulative frequency distribution.
To construct the ogive, the cumulative relative frequency of each interval is plotted above the upper limit of that interval and the points representing the cumulative frequencies are then joined by straight lines. The ogive is

closed at the lower end by extending a straight line to the lower limit of the first interval.

Example:
Using the data from the example in section 1.3.1:



Self-Assessment Exercise 1

1.	Define the cumulative frequency distribution
2.	Explain the ogive in the cumulative frequency distribution

5.4.5 Relative distributions

For each of the frequency distribution and the cumulative frequency distribution, relative distributions can be calculated. A relative frequency distribution includes the percentage of

Sample size or relative frequency (frequency relative to the total sample size) for each interval.
Example
Using the data from the example in section 1.3.1:

Interval (weight in lbs)	Frequency	Relative frequency (factor)	Relative frequency (percentage)
140 – 150	1	0.04	4%
150 – 160	4	0.16	16%
160 – 170	8	0.32	32%
170 – 180	7	0.28	28%
180 – 190	5	0.20	20%

A relative cumulative frequency distribution includes the cumulative percentage of sample size or relative cumulative frequency (cumulative frequency relative to the total sample size) for each interval.

Example

Using the data from the example in section 1.3.1:

Interval (weight in lbs)	Frequency	Cumulative frequency	Relative frequency (factor)	Relative frequency (percentage)
140 – 150	1	1	0,04	4%
150 – 160	4	5	0,20	20%
160 – 170	8	13	0,52	52%
170 – 180	7	20	0,80	80%
180 – 190	5	25	1,00	100%

5.5 Pie charts, bar charts and line charts

The methods described in the previous section are appropriate for summarising quantitative data. But we should also be able to describe data that are qualitative or categorical. These data consist of attributes or names of the categories into which the observations are sorted.
A pie chart is a useful method for displaying the percentage of observations that fall into each cate gory of qualitative data.
A pie chart is an effective method of showing the percentage breakdown of a whole entity.

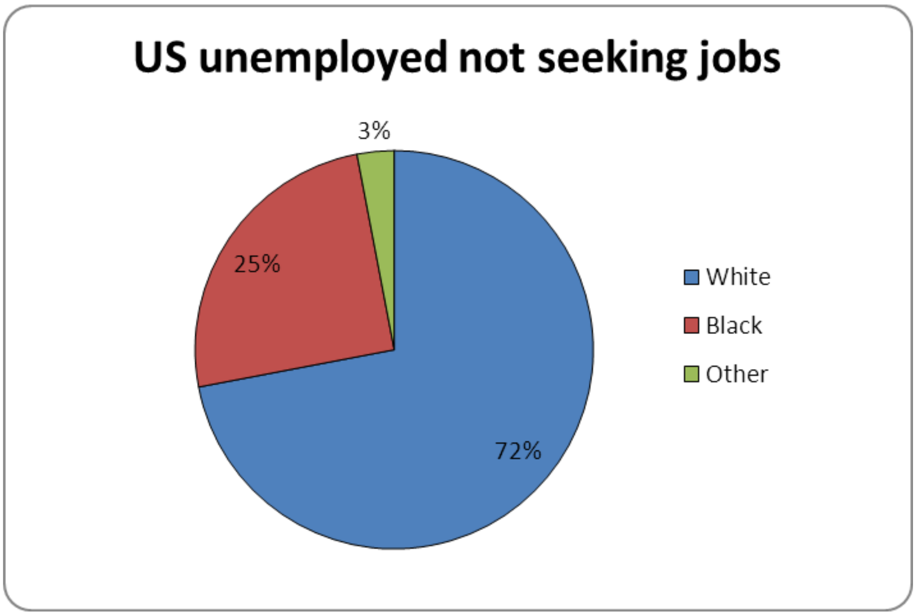
Example

A New York Times article reports that "6 million Americans who say they want work are not even seeking jobs". These 6 million Americans are broken down by race:

Race	Frequency
White	4, 320 000
Black	1, 500 000
Other	180, 000
Total	6,000,000

We need to first determine the percentage of the 6 million Americans belonging to each of the three racial categories: 72% white, 25% black and 3% other. Each category is represented by a slice of the pie (a circle) that is proportional in size to the percentage (or relative frequency) corresponding to that category.

Since the entire circle corresponds to 360°, the angle between the lines demarcating the white sector is therefore $(0{,}72)(360) = 259{,}2°$. In a similar manner, we can determine the angles for the black and other sectors as 90° and 10,8°, respectively.



5.6 Bar charts

Bar charts are a quick and easy way of showing variation in or between variables.

Rectangles of equal width are drawn so that the area enclosed by each rectangle is proportional to the size of the variable it represents. This type of graph not only illustrates a general trend, but also allows a quick and accurate comparison of one period with another or the illustration of a situation a particular time.

When drawing up bar charts take care to:
Make the bars reasonably wide so that they can be clearly seen.
Draw them neatly and professionally.
Ensure that the bars all have the same width.
Ensure that the gaps between the bars have the same width.
We can produce a variety of bar charts to provide an overview of the data.

5.6.1 Simple bar chart

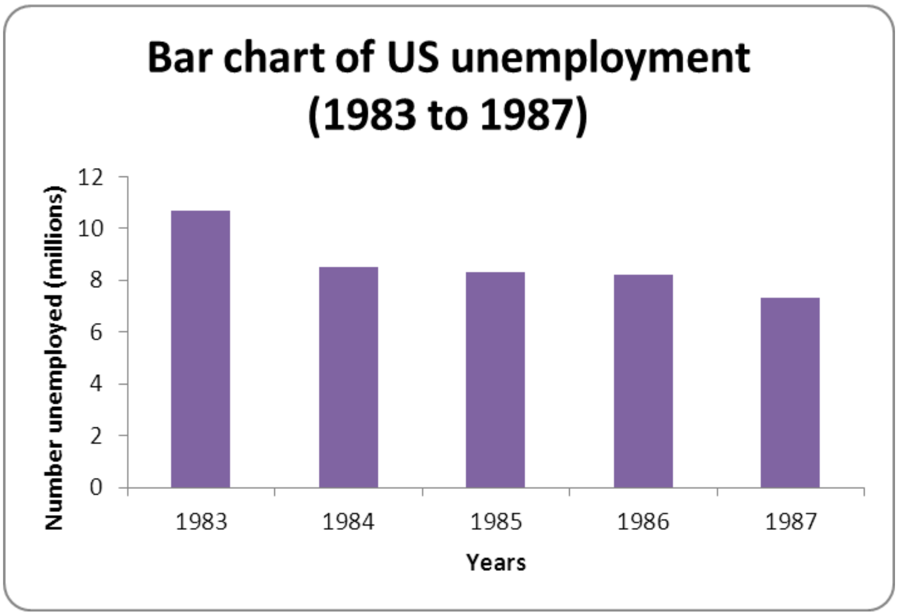
A simple bar chart comprises bars representing each variable drawn either vertically or horizontally. While a bar chart can be used to display the frequency of observations that fall into each category, if the categories consist of points in time and the objective is to focus on the trend in frequencies over time, a line chart is useful.

Example

According to the Nigeria primetimes (27 September 1987), the June levels of unemployment in the Nigeria for five years are:

Year	Unemployed (millions)
1983	10.7
1984	8.5
1985	8.3
1986	8.2
1987	7.3

For the bar chart, the five years or categories are represented by intervals of equal width on the horizontal axis. The height of the vertical bar erected above any year is proportional to the frequency (number of unemployed) corresponding to that year.



A line chart is obtained by plotting the frequency of a category above the point on the horizontal axis representing that category and then joining the points with straight lines.

Self-Assessment Exercises 2

Define Bar Chart

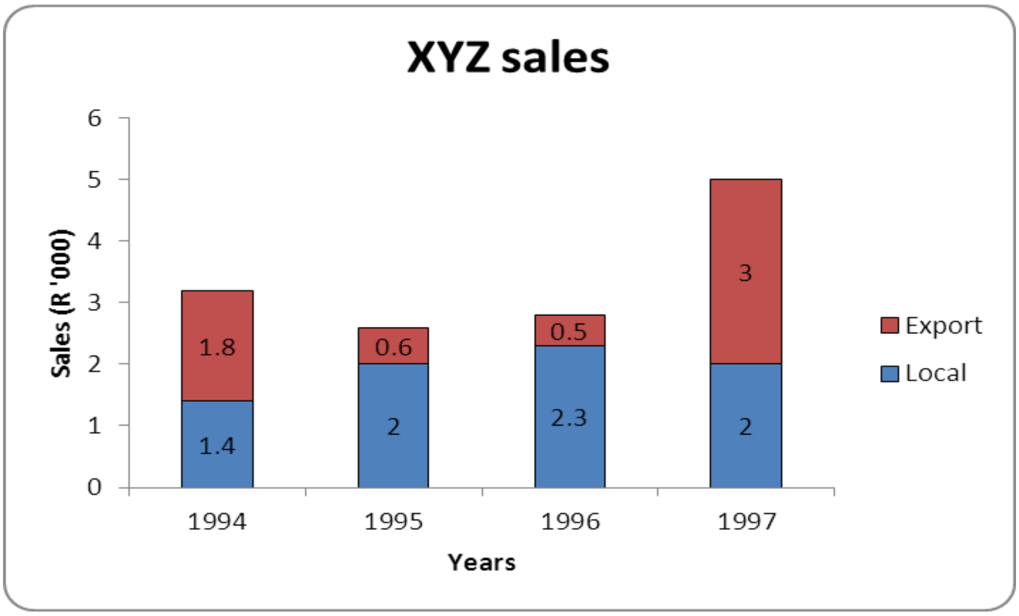
Explain the concept of Simple Bar Chart

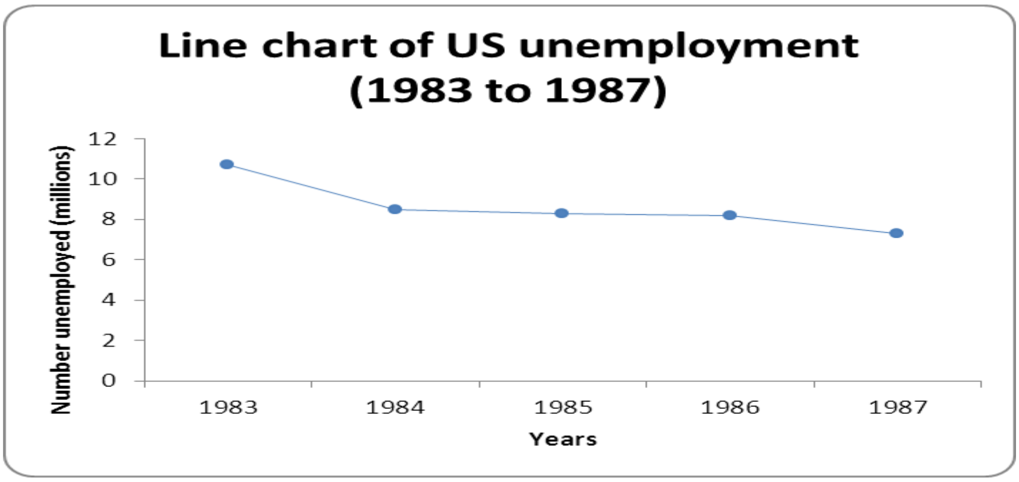
5.6.2 A component or stacked bar chart

In a component or stacked bar chart, a single bar is drawn for each variable, with the heights of the bars representing the totals of the categories. Each bar is then subdivided to show the components that make up the total bar. These components may be identified by colouring or shading, accompanied by an explanatory key to show what each component represents.

Years	XYZ Sales/Categories Local	 Export
1994	1.4	1.8
1995	2.0	0.6
1996	2.3	0.5
1997	2.0	3.0

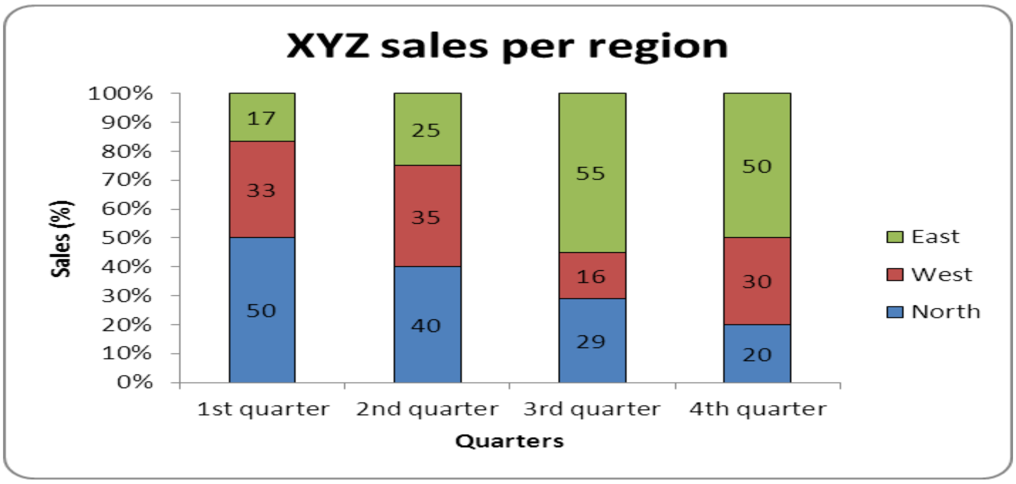
Percentage component bar chart





A percentage component comprises components converted to percentages of the total with the bars divided in proportion to these percentages. The scale is a percentage scale and the height of each bar is therefore 100%.

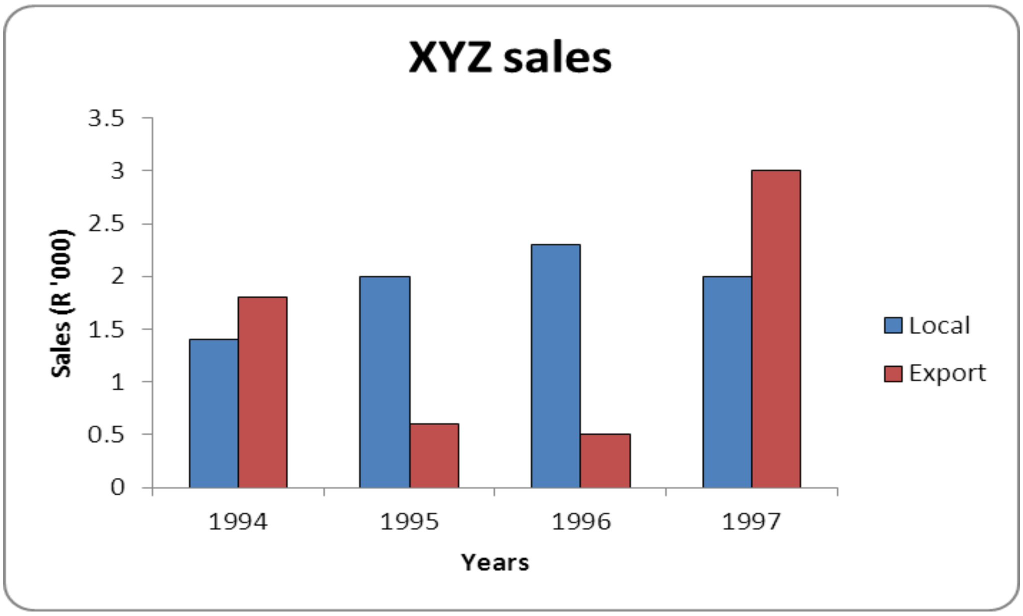
Quarters Sales/Region	North	West	East	Total
1st Quarter	50	33	17	100
2nd Quarter	40	35	25	100
3rd Quarter	29	16	55	100
4th Quarter	20	30	50	100



5.6.3 Multiple Bar Charts

For multiple or cluster bar charts, two or more bars are grouped together in each category. The use of a key helps to distinguish between the categories.

Years	XYZ Sales/Categories Local	Export
1994	1.4	1.8
1995	2.0	0.5
1996	2.3	0.4
1997	1.8	3.0



5.7 Scatter diagrams

The relationship between two quantitative variables can be depicted in a scatter diagram. Economists, for example, are interested in the relationship between inflation rates and unemployment rates. Business owners are interested in many variables, including the relationship between their advertising expenditures and sales levels.

A scatter diagram is a plot of all pairs of values (x, y) for the variables x and y.

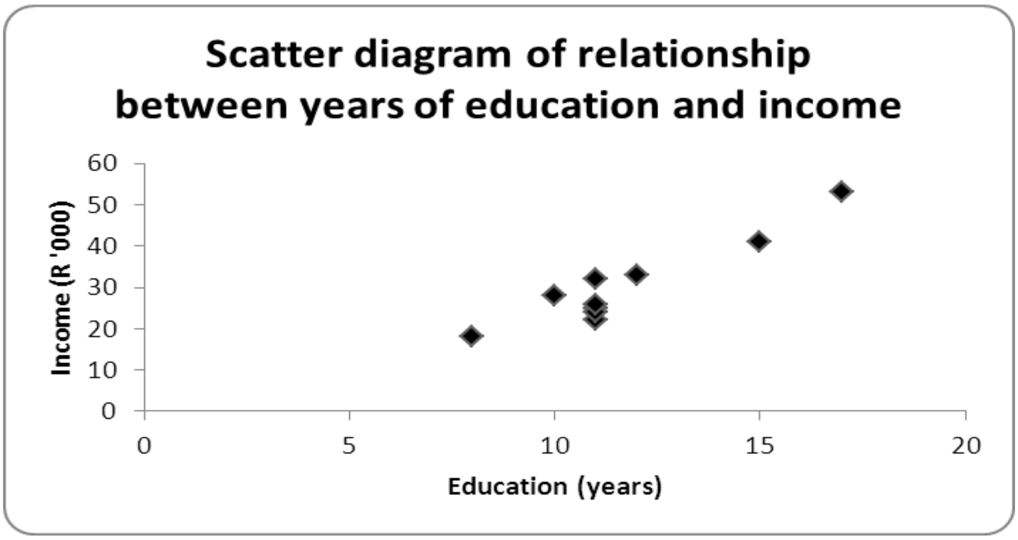
Example

An educational economist wants to establish the relationship between an individual's income and education. She takes a random sample of 10 individuals and asks for their income (in N´000s) and education (in years).

x (years of education)	y (income in N´000)
11	25
12	33
11	22
15	41
8	18
10	28
11	32
11	24
17	53

11	26

If we feel the value of one variable (such as income) depends to some degree on the value of the other variable (such as years of education), the first variable (income) is called the dependent variable and is plotted on the vertical or y-axis. The second variable is the independent variable and is plotted on the x-axis. Think of the independent variable (x-axis) as the 'cause' and the dependent variable (y-axis) as the 'effect'.
The scatter diagram allows us to observe two characteristics about the relationship between education and income (y): Because these two variables move together, ie their values tend to increase together and



decrease together; there is a positive relationship between the two variables. The relationship between income and years of education appears to be linear, since we can imagine drawing a straight line (as opposed to a curved line) through the scatter diagram that approximates the positive relationship between the two variables. The pattern of a scatter diagram provides us with information about the relationship between two variables. Figure 1 depicts a positive linear relationship.
If two variables move in opposite directions and the scatter diagram consists of points that appear to cluster around a straight line, then the variables have a negative linear relationship.

5.8 Skewness and Kurtosis

Skewness and kurtosis are statistical measures that describe the shape of a distribution.
Skewness
Skewness measures the asymmetry of the probability distribution of a real-valued random variable about its mean. The formula for skewness (γ) is:
$\gamma=(n-1)(n-2)$

∑i=1(xi−x‾s)
γ= (n−1)(n−2)/
n
∑i=1 (sxi−x‾)3
S
Where:
n is the number of observations.
x is each individual observation.
$\bar{x}$ is the mean of the observations.
s is the standard deviation of the observations.
Example:
Consider the dataset: [2, 3, 5, 7, 11]
Calculate the mean (x‾x‾): x‾=2+3+5+7+11/5=5.6
Calculate the standard deviation
(S^2)=(2−5.6)2+(3−5.6)2+(5−5.6)2+(7−5.6)2+(11−5.6)25−1=3.5777s=5−1(2−5.6)2+(3−5.6)2+(5−5.6)2+(7−5.6)2+(11−5.6)2=3.5777
Calculate the skewness: γ=5(5−1)(5−2)[(2−5.63.5777)3+(3−5.63.5777)3+(5−5.63.5777)3+(7−5.63.5777)3+(11−5.63.5777)3]γ=(5−1)(5−2)5[(3.57772−5.6)3+(3.57773−5.6)3+(3.57775−5.6)3+(3.57777−5.6)3+(3.577711−5.6)3]
γ≈−0.268γ≈−0.268

Kurtosis
Kurtosis measures the "tailedness" of the probability distribution. The formula for kurtosis (κ) is:

$$\kappa = \frac{n(n-1)}{(n-1)(n-2)(n-3)} \quad \sum i=1 \quad \frac{n(xi-\bar{x})^4}{S} - \frac{(n-1)^2}{(n-2)(n-3)}$$

Where:
n = is the number of observations.
x_i = is each individual observation.
$\bar{x}$ = is the mean of the observations.
s = is the standard deviation of the observations.
Example:
Using the same dataset: [2, 3, 5, 7, 11]
Calculate the mean (x‾x‾) and standard deviation (ss) as before.
Calculate the kurtosis: κ=5(5+1)(5−1)(5−2)(5−3)[(2−5.63.5777)4+(3−5.63.5777)4+(5−5.63.5777)4+(7−5.63.5777)4+(11−5.63.5777)4]−3(5−1)2(5−2)(5−3)κ=(5−1)(5−2)(5−3)5(5+1)[(3.57772−5.6)4+(3.57773−5.6)4+(3.57775−5.6)4+(3.57777−5.6)4+(3.577711−5.6)4]−(5−2)(5−3)3(5−1)2
κ≈−1.36κ≈−1.36

Interpretation:

Skewness: A skewness of -0.268 indicates a slight left skew (negative skew).

Kurtosis: A kurtosis of -1.36 indicates a distribution that is less "tailed" than a normal distribution (platykurtic).

Self-Assessment Exercise 2

Q1. Define the concept of Scatter diagrams with analysis Q2. Explain concept of Skewness with example

5.9 Summary

In this unit, we delved into a multitude of important topics regarding frequency distribution. Firstly, we explored the significance of interpreting frequency tables, which serve as a fundamental tool for organizing and displaying data in a clear and concise manner. These tables play a crucial role in facilitating a structured analysis of numerical information, enabling us to identify patterns and trends within datasets effectively.

Moreover, we examined the concept of data visualization through the utilization of frequency polygons. By utilizing frequency polygons, individuals gain access to a powerful method of visually representing data distributions. These polygons offer a visually appealing and intuitive way of illustrating the frequency of specific data points and their relationships within a dataset. They serve as a bridge between raw data and meaningful insights, allowing for a more in-depth understanding of the underlying patterns and characteristics present in the data.

Furthermore, we explored the significance of histograms and Ogives in the realm of frequency distribution analysis. Histograms provide a visual representation of data distribution through the use of bars, highlighting the frequency of occurrences within specific intervals or categories. On the other hand, Ogives offer a cumulative view of the data distribution, providing valuable insights into the cumulative frequency distribution within a dataset.

In essence, our exploration of these essential topics in frequency distribution within this unit has equipped us with a comprehensive understanding of the various tools and techniques available for effectively analyzing and presenting data. By mastering the interpretation of frequency tables, harnessing the power of frequency polygons, and leveraging histograms and Ogives, we are better prepared to navigate the complexities of data analysis and draw meaningful conclusions from diverse datasets.

5.10 References/Further Reading/Web Resources

5.11 Possible Answers SAEs

Answer to SAEs 1

Q1 Define frequency distribution

A frequency distribution is a summary of how often different values occur within a data set. It can be presented in a table format.

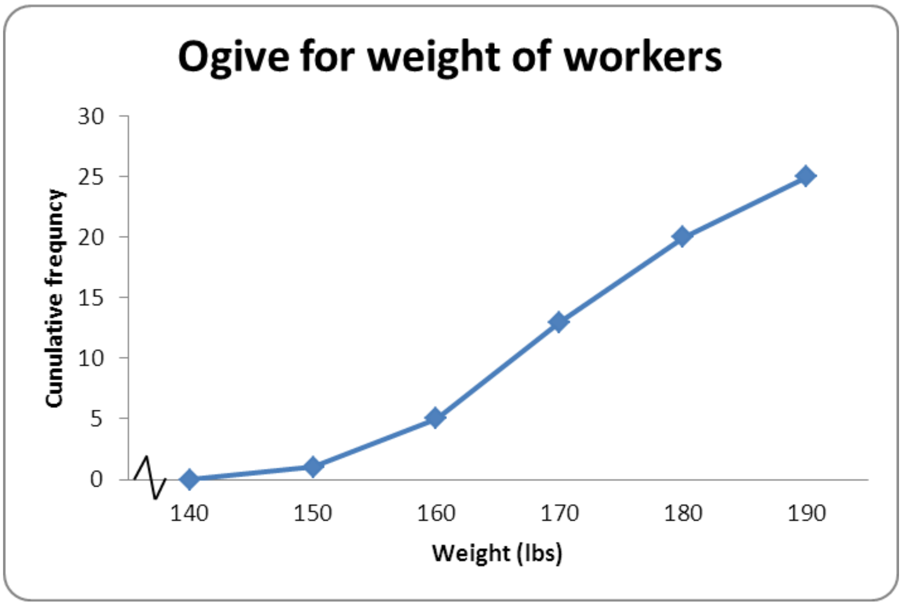
Example: Consider the data set: 1, 2, 2, 3, 3, 3, 4, 4, 4, 4

Value	Frequency
1	1
2	2
3	3
4	4
5	1
7	2
8	3
9	2
10	1

Q2 An Ogive

An ogive is a graph of the cumulative frequency distribution.
To construct the ogive, the cumulative relative frequency of each interval is plotted above the upper limit of that interval and the points representing the cumulative frequencies are then joined by straight lines. The ogive is closed at the lower end by extending a straight line to the lower limit of the first interval.
Example:

Using the data from the example in section 1.3.1:



Answer to SAEs 2

Q1. Scatter diagrams

The relationship between two quantitative variables can be depicted in a scatter diagram. Economists, for example, are interested in the relationship between inflation rates and unemployment rates. Business owners are interested in many variables, including the relationship between their advertising expenditures and sales levels.

A scatter diagram is a plot of all pairs of values (x, y) for the variables x and y.

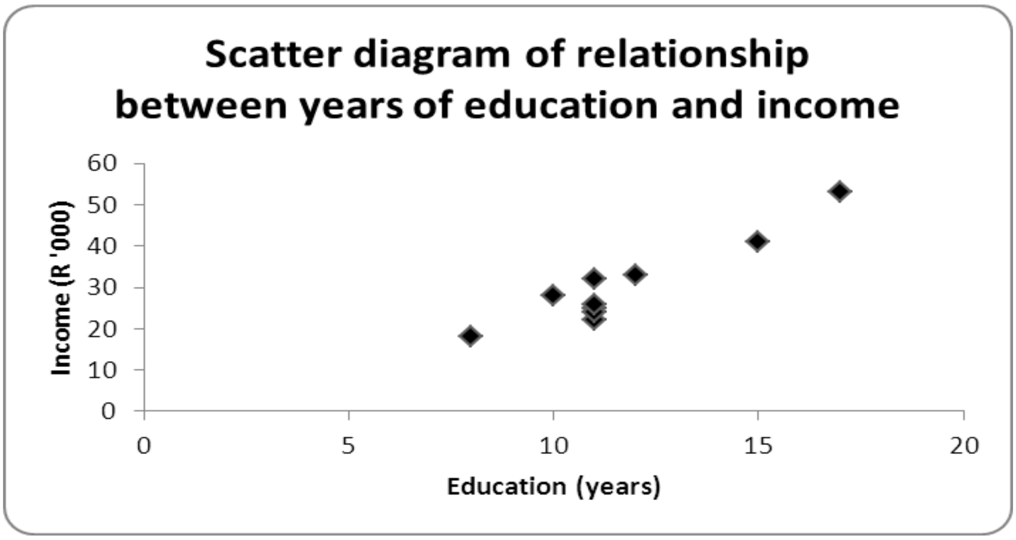
Example

An educational economist wants to establish the relationship between an individual's income and education. She takes a random sample of 10 individuals and asks for their income (in N´000s) and education (in years).

x (years of education)	y (income in N´000)
11	25
12	33
11	22
15	41
8	18
10	28
11	32

11	24
17	53
11	26

If we feel the value of one variable (such as income) depends to some degree on the value of the other variable (such as years of education), the first variable (income) is called the dependent variable and is plotted on the vertical or y-axis. The second variable is the independent variable and is plotted on the x-axis. Think of the independent variable (x-axis) as the 'cause' and the dependent variable (y-axis) as the 'effect'.
The scatter diagram allows us to observe two characteristics about the relationship between education and income (y): Because these two variables move together, ie their values tend to increase together and



decrease together; there is a positive relationship between the two variables. The relationship between income and years of education appears to be linear, since we can imagine drawing a straight line (as opposed to a curved line) through the scatter diagram that approximates the positive relationship between the two variables. The pattern of a scatter diagram provides us with information about the relationship between two variables. Figure 1 depicts a positive linear relationship.
If two variables move in opposite directions and the scatter diagram consists of points that appear to cluster around a straight line, then the variables have a negative linear relationship.

5.8 Skewness

Skewness and kurtosis are statistical measures that describe the shape of a distribution.

Skewness

Skewness measures the asymmetry of the probability distribution of a real-valued random variable about its mean. The formula for skewness (γ) is:

$\gamma=(n-1)(n-2)$

$\sum i=1(xi-\bar{x}s)$

$\gamma= (n-1)(n-2)/$

n

$\sum i=1\ (sxi-\bar{x})^3$

– S

Where:

n is the number of observations.

x is each individual observation.

$\bar{x}$ is the mean of the observations.

s is the standard deviation of the observations.

Example:

Consider the dataset: [2, 3, 5, 7, 11]

Calculate the mean ($\bar{x}\bar{x}$): $\bar{x}$=2+3+5+7+11/5=5.6

Calculate the standard deviation (ss): s=(2−5.6)2+(3−5.6)2+(5−5.6)2+(7−5.6)2+(11−5.6)25−1=3.5777s= 5−1(2−5.6)2+(3−5.6)2+(5−5.6)2+(7−5.6)2+(11−5.6)2=3.5777

Calculate the skewness: γ=5(5−1)(5−2)[(2−5.63.5777)3+(3−5.63.5777)3+(5−5.63.5777)3+(7−5.63.5777)3+(11−5.63.5777)3]γ=(5−1)(5−2)5[(3.57772−5.6)3+(3.57773−5.6)3+(3.57775−5.6)3+(3.57777−5.6)3+(3.577711−5.6)3]

γ≈−0.268γ≈−0.268

MODULE 2

UNIT 1 REGRESSION ANALYSIS

Unit Structure

1.1 Introduction

The previous unit examined Concept of Sampling, essential Steps to Conduct Sampling, Techniques of Sample Size Determination, various Sampling methods commonly employed in public administration research, link between Sampling and Hypothesis Testing, Hypothesis Testing and Confidence Intervals, outlining their significance in inferential statistics and research interpretation, relationship between Population and Sample and the Types of Errors. This unit will be discussion the regression analysis.

1.2 Learning Outcome

At the end of the unit, you should be able to:

- **Define the Regression Analysis**
- **Outline the Linear Regression**
- **Highlight the Multiple Regressions**
- Interpretation the Regression Coefficients
- Explain the Application in Predicting Social Outcomes

- **Outline the assumptions of Multiple Linear Regressions**
- Analyze the Coefficient of Determination
- **State the assumptions of Linear Regression**
- Draw the A linear function
- Define the Non-Parametric Methods
- Analyze the Chi-Square Test
- State and analyze the T-Test

1.3 Regression Analysis

Regression Analysis is a statistical method used to examine the relationship between one dependent variable and one or more independent variables. It helps in understanding how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.

1.3.1 Linear Regression:

- **Definition:** Linear regression is the simplest form of regression analysis where the relationship between the dependent variable and a single independent variable is modeled as a straight line.

- **Equation:** $Y=a+bX+\epsilon$Y=a+bX+ϵ
- Y is the dependent variable.
- X is the independent variable.
- a is the intercept.
- b is the slope of the line.
- ϵ is the error term.

1.3.2 Multiple Regressions:

Definition: Multiple regression involves more than one independent variable. It models the relationship between a dependent variable and multiple independent variables.

Equation: $Y=a+b1X1+b2X2+...+bnXn+\epsilon$Y=a+b1X1+b2X2+...+bnXn +ϵ

YY is the dependent variable.

X1,2,...,XnX1,X2,...,Xn are the independent variables.

aa is the intercept.

b1,2,...,bnb1,b2,...,bn are the coefficients for each independent variable.

ϵϵ is the error term.

1.3.3 Interpretation of Regression Coefficients

Intercept (a): The expected value of the dependent variable when all independent variables are zero.

Slope (b): Represents the change in the dependent variable for a one-unit change in the independent variable, holding all other variables constant.
In multiple regression, each coefficient indicates the change in the dependent variable for a one-unit change in the corresponding independent variable Xi, holding all other variables constant.

Significance: The p-value associated with each coefficient helps determine whether the relationship between the independent variable and the dependent variable is statistically significant.

1.4 Application in Predicting Social Outcomes

Regression analysis is widely used in social research to predict and understand various social outcomes. Here are some applications:

Education: Predicting student performance based on factors like socioeconomic status, parental education, and school resources.

Health: Analyzing the impact of lifestyle choices, socioeconomic factors, and healthcare access on health outcomes.

Economics: Understanding the relationship between employment rates, income levels, and economic policies.

Sociology: Examining the effects of social policies, community programs, and demographic factors on social behaviors and outcomes.

Public Policy: Evaluating the effectiveness of policy interventions by modeling the relationship between policy changes and social indicators.
By using regression analysis, researchers can make informed predictions and decisions, identify key factors influencing social outcomes, and develop strategies to address social issues effectively.

SELF-ASSESSMENT EXERCISE 1

Q1. Define the Regression Analysis
Q2. State the formula for Linear Regression
Q3. State the formula for Regressions

1.5 Assumptions of Multiple Linear Regressions

There are four key assumptions that multiple linear regression makes about the data:

- **Linear relationship:** There exists a linear relationship between the independent variable, x, and the dependent variable, y.

- **Independence:** The residuals are independent. In particular, there is no correlation between consecutive residuals in time series data.

- **Homoscedasticity:** The residuals have constant variance at every level of x.

- **Normality:** The residuals of the model are normally distributed For a complete explanation of how to test these assumptions, check on websites bellows for tutorials on step-by-step examples of how to perform multiple linear regressions using different statistical software:

1.6 Multiple Linear Regressions Using Software for the links:

How to Perform Multiple Linear Regression in SPSS
How to Perform Multiple Linear Regression in R
How to Perform Multiple Linear Regression in Python
How to Perform Multiple Linear Regression in Excel
How to Perform Multiple Linear Regression in Stata
How to Perform Linear Regression in Google Sheets

3.5 The Coefficient of Determination

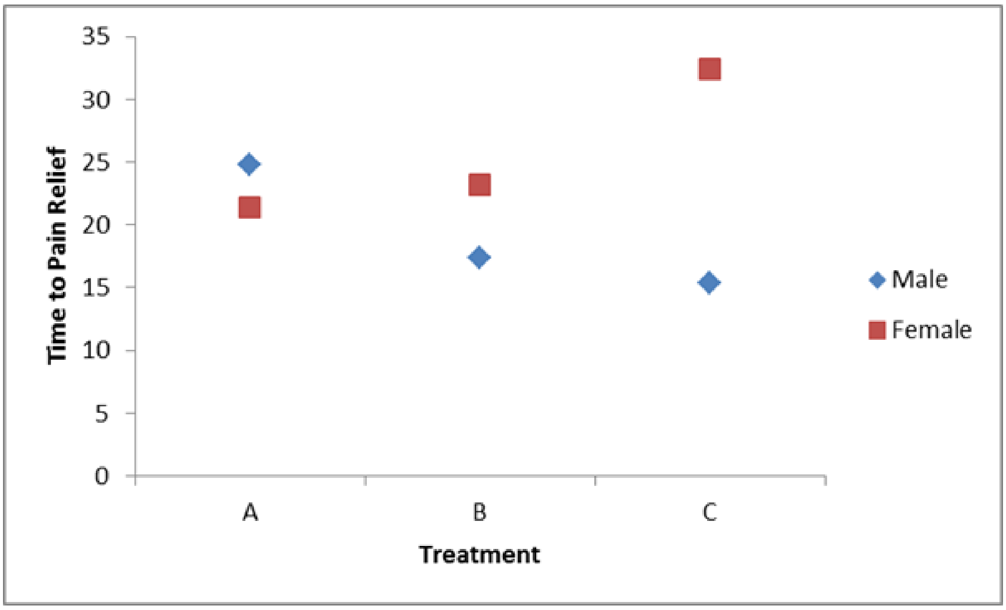
One way to measure how well the least squares regression line "fits" the data is using the coefficient of determination, denoted as R2.

The coefficient of determination is the proportion of the variance in the response variable that can be explained by the predictor variable.

The coefficient of determination can range from 0 to 1. A value of 0 indicates that the response variable cannot be explained by the predictor variable at all. A value of 1 indicates that the response variable can be perfectly explained without error by the predictor variable.

An R2 between 0 and 1 indicates just how well the response variable can be explained by the predictor variable. For example, an R2 of 0.2 indicates that 20% of the variance in the response variable can be explained by the predictor variable; an R2 of 0.77 indicates that 77% of the variance in the response variable can be explained by the predictor variable.

Notice in our output from earlier we got an R2 of 0.9311, which indicates that 93.11% of the variability in height can be explained by the predictor variable of weight:



This tells us that weight is a very good predictor of height.

3.5.1 Assumptions of Linear Regression

For the results of a linear regression model to be valid and reliable, we need to check that the following four assumptions are met:

1. Linear relationship: There exists a linear relationship between the independent variable, x, and the dependent variable, y.

2. Independence: The residuals are independent. In particular, there is no correlation between consecutive residuals in time series data.

3. Homoscedasticity: The residuals have constant variance at every level of x.

4. Normality: The residuals of the model are normally distributed.
If one or more of these assumptions are violated, then the results of our linear regression may be unreliable or even misleading.

1.6 A linear function

A linear function is represented in a data. Here are two choices.

Weight (lbs)	Height (inches)
140	60
155	62
159	67
179	70
192	71
200	72
212	75

1.7 Non-Parametric Methods

Definition

Non-parametric methods are statistical techniques that do not assume a specific distribution for the data. These methods are particularly useful when:

1. Data does not meet the assumptions of parametric tests: For example, when the data is not normally distributed, has outliers, or is ordinal rather than interval or ratio.

2. Sample size is small: Non-parametric methods are less sensitive to small sample sizes.

3. Data is ordinal or nominal: When the data is ranked or categorical, non-parametric methods are more appropriate.

1.7.1 Chi-Square Test

1. Chi-Square Test: Purpose: Used to determine if there is a significant association between two categorical variables. Example: Testing whether there is a relationship between gender (male/female) and voting preference (yes/no).

Application in Social Data with Non-Normal Distribution
Non-parametric methods are particularly useful in social sciences where data often do not meet the assumptions required for parametric tests. For example:

1. Survey Data: Often, survey responses are ordinal (e.g., Likert scale ratings) and may not be normally distributed. Non-parametric tests like

the Mann-Whitney U test can be used to compare responses between different groups.

2. Behavioral Studies: In studies measuring behaviors or attitudes, the data may be skewed or have outliers. Non-parametric methods can provide more reliable results in these cases.
3. Small Sample Sizes: In qualitative research or pilot studies where sample sizes are small, non-parametric methods can be more appropriate and provide more accurate insights.

A Chi-Square Test is a statistical test used to determine if there's a significant association between categorical variables. Let's go through an example step-by-step.

Example Scenario:
Suppose you want to test whether there is an association between gender (male, female) and preference for a new product (like, dislike).

Step 1: Collect Data
You survey 100 people and get the following results:

	Like	Dislike	Total
Male	20	30	50
Female	25	25	50
Total	45	55	100

Step 2: State the Hypotheses

- **Null Hypothesis (H0):** There is no association between gender and product preference.

- **Alternative Hypothesis (H1):** There is an association between gender and product preference.

Step 3: Calculate Expected Frequencies
The expected frequency for each cell is calculated using the formula: $Eij=(Row\ Total \times Column\ Total)Grand\ TotalEij$ $=Grand\ Total(Row\ Total \times Column\ Total)$
For example, the expected frequency for males who like the product: $E11=(50\times45)100=22.5E11=100(50\times45)=22.5$
Similarly, calculate for all cells:

	Like	Dislike
Male (E)	22.5	27.5
Female (E)	22.5	27.5

Step 4: Calculate the Chi-Square Statistic
$\chi2=\sum(Oij-Eij)2/Eij$ $\chi2=\sum Eij(Oij-Eij)2/\ Eij$

Where O_{ij} is the observed frequency and E_{ij} is the expected frequency.

For each cell:

Male, Like: $\frac{(20-22.5)^2}{22.5}=0.278$
Male, Dislike: $\frac{(30-27.5)^2}{27.5}=0.227$
Female, Like: $\frac{(25-22.5)^2}{22.5}=0.278$
Female, Dislike: $\frac{(25-27.5)^2}{27.5}=0.227$
Sum these values to get the Chi-Square statistic: $\chi^2=0.278+0.227+0.278+0.227=1.01$

Step 5: Determine the Degrees of Freedom
$df=(rows-1)\times(columns-1)$ $df=(2-1)\times(2-1)=1$

Step 6: Find the Critical Value and Compare
Using a Chi-Square distribution table, find the critical value for $df=1$ at a significance level (e.g., 0.05). The critical value is approximately 3.841.

Step 7: Make a Decision

If χ^2 > critical value, reject the null hypothesis.
If $\chi^2 \leq$ critical value, fail to reject the null hypothesis. In this case, $\chi^2=1.01$ which is less than 3.841. Therefore, we fail to reject the null hypothesis.

Conclusion
There is no significant association between gender and product preference in this sample.

1.7.2 T-Test

Scenario:
Imagine you are a researcher studying the effect of a new teaching method on student performance. You have two groups of students: one group uses the traditional teaching method (Group A), and the other group uses the new teaching method (Group B). After a semester, you want to compare the test scores of the two groups to see if there is a significant difference.

Steps to Perform a T-Test:

1. **Formulate Hypotheses:**

Null Hypothesis (H0): There is no significant difference in the mean test scores between the two groups.
Alternative Hypothesis (H1): There is a significant difference in the mean test scores between the two groups.

2. **Collect Data:**

Group A (Traditional Method): Test scores = [78, 82, 85, 90, 88, 76, 95, 89]
Group B (New Method): Test scores = [85, 87, 90, 92, 91, 88, 94, 93]

Calculate the Means and Standard Deviations:

Mean of Group A: $\bar{X}_A$
Mean of Group B: $\bar{X}_B$
Standard Deviation of Group A: s_A
Standard Deviation of Group B: s_B

3. **Choose the Type of T-Test:**

Since we have two independent groups, we will use an independent samples t-test.

4. **Calculate the T-Statistic:** The formula for the t-statistic in an independent samples t-test is: $t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$ where n_A and n_B are the sample sizes of Group A and Group B, respectively.

5. **Determine the Degrees of Freedom:** The degrees of freedom (df) for an independent samples t-test can be calculated using: $df = n_A + n_B - 2$

6. **Find the Critical Value and Compare:**

Use a t-distribution table to find the critical value for a chosen significance level (e.g., $\alpha = 0.05$) and the calculated degrees of freedom.
Compare the calculated t-statistic to the critical value to determine whether to reject the null hypothesis.

Example Calculation:

Let's assume the following values for simplicity:

Mean of Group A ($\bar{X}_A$) = 85.375
Mean of Group B ($\bar{X}_B$) = 90
Standard Deviation of Group A (s_A) = 6.5
Standard Deviation of Group B (s_B) = 2.8

Sample size of Group A (*nAnA*) = 8
Sample size of Group B (*nBnB*) = 8
Using the formula:

t=85.375−906.528+2.828=−4.62542.258+7.848=−4.6255.28125+0.98=−4.6256.26125=−4.6252.503≈−1.85t=86.52+82.8285.375−90=842.25+87.84−4.625=5.28125+0.98−4.625=6.26125−4.625=2.503−4.625≈−1.85
Degrees of freedom:
df=8+8−2=14*df*=8+8−2=14

Using a t-distribution table, the critical value for *df*=14*df*=14 at *α*=0.05*α*=0.05 (two-tailed) is approximately ±2.145. Since -1.85 is within the range of -2.145 to 2.145, we fail to reject the null hypothesis. Therefore, there is no significant difference in the mean test scores between the two groups at the 0.05 significance level.

SELF-ASSESSMENT EXERCISE 2

1.	Explain the Coefficient of Determination
2.	**State the Assumptions of Linear Regression**

1.8 Summary

The unit explains that, using linear regression, we can find the line that best "fits" our data. This line is known as the least squares regression line and it can be used to help us understand the relationships between weight and height. Usually you would use software like Microsoft Excel, SPSS, or a graphing calculator to actually find the equation for this line. The formula for the line of best fit is written as:
$\hat{y} = b_0 + b_1x$

Regression line "fits" the data is using the coefficient of determination, denoted as R2.

The coefficient of determination is the proportion of the variance in the response variable that can be explained by the predictor variable.

The coefficient of determination can range from 0 to 1. A value of 0 indicates that the response variable cannot be explained by the predictor variable at all. A value of 1 indicates that the response variable can be perfectly explained without error by the predictor variable.

Non-parametric methods are versatile tools in statistical analysis, especially useful when dealing with non-normal distributions, small sample sizes, or ordinal data. They provide robust alternatives to parametric tests and are widely used in social sciences to analyze data that do not meet the stringent assumptions of parametric methods.

1.9 References/Further Reading/Web Resources

1.10 Possible Answers SAEs

Answer to SAEs 1

Q1. Regression Analysis

Regression Analysis is a statistical method used to examine the relationship between one dependent variable and one or more independent variables. It helps in understanding how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed

Q2. Linear Regression:

Definition: Linear regression is the simplest form of regression analysis where the relationship between the dependent variable and a single independent variable is modeled as a straight line.

Equation: $Y=a+bX+\epsilon$Y=a+bX+ϵ
Y is the dependent variable.
X is the independent variable.
a is the intercept.
b is the slope of the line.
ϵ is the error term.

Q3. Multiple Regressions:

Definition: Multiple regression involves more than one independent variable. It models the relationship between a dependent variable and multiple independent variables.

Equation: $Y=a+b1X1+b2X2+...+bnXn+\epsilon$Y=a+b1X1+b2X2+...+bnXn+ϵ
YY is the dependent variable.
X1,2,...,XnX1,X2,...,Xn are the independent variables.
aa is the intercept.
b1,2,...,bnb1,b2,...,bn are the coefficients for each independent variable.
ϵϵ is the error term.

Answer to SAEs 2

Q1. The Coefficient of Determination

One way to measure how well the least squares regression line "fits" the data is using the coefficient of determination, denoted as R2.

The coefficient of determination is the proportion of the variance in the response variable that can be explained by the predictor variable.

The coefficient of determination can range from 0 to 1. A value of 0 indicates that the response variable cannot be explained by the predictor variable at all. A value of 1 indicates that the response variable can be perfectly explained without error by the predictor variable.

An R2 between 0 and 1 indicates just how well the response variable can be explained by the predictor variable. For example, an R2 of 0.2 indicates that 20% of the variance in the response variable can be explained by the predictor variable; an R2 of 0.77 indicates that 77% of the variance in the response variable can be explained by the predictor variable.

Notice in our output from earlier we got an R2 of 0.9311, which indicates that 93.11% of the variability in height can be explained by the predictor variable of weight:

Q2. Assumptions of Linear Regression

For the results of a linear regression model to be valid and reliable, we need to check that the following four assumptions are met:

1. Linear relationship: There exists a linear relationship between the independent variable, x, and the dependent variable, y.

2. Independence: The residuals are independent. In particular, there is no correlation between consecutive residuals in time series data.

3. Homoscedasticity: The residuals have constant variance at every level of x.

4. Normality: The residuals of the model are normally distributed.
If one or more of these assumptions are violated, then the results of our linear regression may be unreliable or even misleading.

UNIT 2 ANALYSIS OF VARIANCE (ANOVA)

Unit Structure

1.1 Introduction

This unit will discuss analysis of variance (ANOVA) and hypothesis testing, ANOVA in SPSS., One-way ANOVA in SPSS Statistics, Assumptions and Test Procedure in SPSS Statistics The specific test considered here is called analysis of variance (ANOVA) and is a test of hypothesis that is appropriate to compare means of a continuous variable in two or more independent comparison groups. For example, in some clinical trials there are more than two comparison groups. In a clinical trial to evaluate a new medication for asthma, investigators might compare an experimental medication to a placebo and to a standard treatment (i.e., a medication currently being used). In an observational study such as the Framingham Heart Study, it might be of interest to compare mean blood pressure or mean cholesterol levels in persons who are underweight, normal weight, overweight and obese.

2.2 Learning Outcomes

- At the end of this unit, you should be able to:
- Explain the concept of Analysis of Variance (ANOVA)

- Discuss the ANOVA Approach
- Demonstrate ANOVA in SPSS.
- Explain One-way ANOVA in SPSS Statistics
- State the Assumptions
- Highlight the Test Procedure in SPSS Statistics
- Explain the SPSS Statistics
- Demonstrate the Setup step in SPSS Statistics
- Demonstrate descriptive Table
- 10 Explain the reporting output of the one-way ANOVA

2.3 Analysis of Variance (ANOVA)

2.3.1 What is Analysis of Variance (ANOVA)?

Analysis of Variance (ANOVA) is the technique use to test the difference between more than two independent means. It is an extension of the two independent samples procedure which applies when there are exactly two independent comparison groups. The ANOVA technique applies when there are two or more than two independent groups. The ANOVA procedure is used to compare the means of the comparison groups and is conducted using the same five step approach used in the scenarios discussed in previous sections. Because there are more than two groups, however, the computation of the test statistic is more involved. The test statistic must take into account the sample sizes, sample means and sample standard deviations in each of the comparison groups.

If one is examining the means observed among, say three groups, it might be tempting to perform three separate group to group comparisons, but this approach is incorrect because each of these comparisons fails to take into account the total data, and it increases the likelihood of incorrectly concluding that there are statistically significate differences, since each comparison adds to the probability of a type I error. Analysis of variance avoids these problemss by asking a more global question, i.e., whether there are significant differences among the groups, without addressing differences between any two groups in particular (although there are additional tests that can do this if the analysis of variance indicates that there are differences among the groups).

The fundamental strategy of ANOVA is to systematically examine variability within groups being compared and also examine variability among the groups being compared

2.3.2 The ANOVA Approach

Consider an example with four independent groups and a continuous outcome measure. The independent groups might be defined by a particular characteristic of the participants such as BMI (e.g., underweight, normal weight, overweight, obese) or by the investigator (e.g., randomizing participants to one of four competing treatments, call them A, B, C and D). Suppose that the outcome is systolic blood pressure, and we wish to test whether there is a statistically significant difference in mean systolic blood pressures among the four groups

Analysis of Variance, i.e. ANOVA in SPSS, is used for examining the differences in the mean values of the dependent variable associated with the effect of the controlled independent variables, after taking into account the influence of the uncontrolled independent variables. Essentially, ANOVA in SPSS is used as the test of means for two or more populations.

ANOVA in SPSS must have a dependent variable which should be metric (measured using an interval or ratio scale). ANOVA in SPSS must also have one or more independent variables, which should be categorical in nature. In ANOVA in SPSS, categorical independent variables are called factors. A particular combination of factor levels, or categories, is called a treatment.

2.4 ANOVA In SPSS.

In ANOVA in SPSS, there is one way ANOVA which involves only one categorical variable, or a single factor. For example, if a researcher wants to examine whether heavy, medium, light and nonusers of cereals differed in their preference for Total cereal, then the differences can be examined by the one way ANOVA in SPSS. In one way ANOVA in SPSS, a treatment is the same as the factor level.

If two or more factors are involved in ANOVA in SPSS, then it is termed as n way ANOVA. For example, if the researcher also wants to examine the preference for Total cereal by the customers who are loyal to it and those who are not, then we can use n way

In ANOVA in SPSS, from the menu we choose:

“Analyze” then go to “Compare Means” and click on the “One-Way ANOVA.”

Now, let us discuss in detail how the software operates ANOVA:

The first step is to identify the dependent and independent variables. The dependent variable is generally denoted by Y and the independent variable is denoted by X. X is a categorical variable having c categories. The sample size in each category of X is generally denoted as n, and the total sample size N=nXc.

The next step in ANOVA in SPSS is to examine the differences among means. This involves decomposition of the total variation observed in the dependent variable. This variation in ANOVA in SPSS is measured by the sums of the squares of the mean.

The total variation in Y in ANOVA in SPSS is denoted by SSy, which can be decomposed into two components:

SSy=SSbetween+SSwithin
where the subscripts between and within refers to the categories of X in ANOVA in SPSS. SSbetween is the portion of the sum of squares in Y related to the independent variable or factor X. Thus it is generally referred to as the sum of squares of X. SSwithin is the variation in Y related to the variation within each category of X. It is generally referred to as the sum of squares for errors in ANOVA in SPSS.
The logic behind decomposing SSy is to examine the differences in group means.

The next task in ANOVA in SPSS is to measure the effects of X on Y, which is generally done by the sum of squares of X, because it is related to the variation in the means of the categories of X. The relative magnitude of the sum of squares of X in ANOVA in SPSS increases as the differences among the means of Y in categories of X increases. The relative magnitude of the sum of squares of X in ANOVA in SPSS increases as the variation in Y within the categories of X decreases.

The strength of the effects of X on Y is measured with the help of $\eta 2$ in ANOVA in SPSS .The value of $\eta 2$ varies between 0 and 1. It assumes a value 0 in ANOVA in SPSS when all the category means are equal, indicating that X has no effect on Y. The value of $\eta 2$ becomes 1, when there is no variability within each category of X but there is still some variability between the categories.

The final step in ANOVA in SPSS is to calculate the mean square which is obtained by dividing the sum of squares by the corresponding degrees of freedom. The null hypothesis of equal means, which is done by an F statistic, is the ratio between the mean square related to the independent variable and the mean square related to the error.

N way ANOVA in ANOVA in SPSS involves simultaneous examination of two or more categorical independent variables, which is also computed in a similar manner.

A major advantage of ANOVA in SPSS is that the interactions between the independent variables can be examined. Follow this link for details:

2.4 One-way ANOVA in SPSS Statistics

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups (although you tend to only see it used when there are a minimum of three, rather than two groups).

For example, you could use a one-way ANOVA to understand whether exam performance differed based on test anxiety levels amongst students, dividing students into three independent groups (e.g., low, medium and high-stressed students). Also, it is important to realize that the one-way ANOVA is an **omnibus** test statistic and cannot tell you which specific groups were statistically significantly different from each other; it only tells you that at least two groups were different. Since you may have three, four, five or more groups in your study design, determining which of these groups differ from each other is important. You can do this using a post hoc test (N.B., we discuss post hoc tests later in this guide).

Note: If your study design not only involves one dependent variable and one independent variable, but also a third variable (known as a "covariate") that you want to "statistically control", you may need to perform an ANCOVA (analysis of covariance), which can be thought of as an extension of the one-way ANOVA. To learn more, see our SPSS Statistics guide on ANCOVA. Alternatively, if your dependent variable is the time until an event happens, you might need to run a Kaplan-Meier analysis.

This "quick start" guide shows you how to carry out a one-way ANOVA using SPSS Statistics, as well as interpret and report the results from this test. Since the one-way ANOVA is often followed up with a post hoc test, we also show you how to carry out a post hoc test using SPSS Statistics. However, before we introduce you to this procedure, you need to understand the different assumptions that your data must meet in order for a one-way ANOVA to give you a valid result. We discuss these assumptions next.

SELF-ASSESSMENT EXERCISE 1

1. Explain the concept of Analysis of Variance (ANOVA)
2. Discuss the ANOVA Approach
3. Demonstrate ANOVA in SPSS.
4. Explain One-way ANOVA in SPSS Statistics

2.4.1 Assumptions

When you choose to analyse your data using a one-way ANOVA, part of the process involves checking to make sure that the data you want to analyse can actually be analysed using a one-way ANOVA. You need to do this because it is only appropriate to use a one-way ANOVA if your data "passes" six assumptions that are required for a one-way ANOVA to give you a valid result. In practice, checking for these six assumptions just adds a little bit more time to your analysis, requiring you to click a few more buttons in SPSS Statistics when performing your analysis, as well as think a little bit more about your data, but it is not a difficult task.

Before we introduce you to these six assumptions, do not be surprised if, when analysing your own data using SPSS Statistics, one or more of these assumptions is violated (i.e., is not met). This is not uncommon when working with real-world data rather than textbook examples, which often only show you how to carry out a one-way ANOVA when everything goes well! However, don’t worry. Even when your data fails certain assumptions, there is often a solution to overcome this. First, let’s take a look at these six assumptions:

- Assumption #1: Your dependent variable should be measured at the interval or ratio level (i.e., they are continuous). Examples of variables that meet this criterion include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth. You can learn more about interval and ratio variables in our article: Types of Variable.

- Assumption #2: Your independent variable should consist of two or more categorical, independent groups. Typically, a one-way ANOVA is used when you have three or more categorical, independent groups, but it can be used for just two groups (but an independent-samples t-test is more commonly used for two groups). Example independent variables that meet this criterion include ethnicity (e.g., 3 groups: Caucasian, African American and Hispanic), physical activity level (e.g., 4 groups: sedentary, low, moderate and high), profession (e.g., 5 groups: surgeon, doctor, nurse, dentist, therapist), and so forth.

- Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves. For example, there must be different participants in each group with no participant being in more than one group. This is more of a study design issue than something you can test for, but it is an important assumption of the one-way ANOVA. If your study fails this assumption, you will need to use another statistical test instead of the one-way ANOVA (e.g., a repeated measures design). If you are unsure whether your study meets this assumption, you can use our Statistical Test Selector, which is part of our enhanced guides.
- Assumption #4: There should be no significant outliers. Outliers are simply single data points within your data that do not follow the usual pattern (e.g., in a study of 100 students' IQ scores, where the mean score was 108 with only a small variation between students, one student had a score of 156, which is very unusual, and may even put her in the top 1% of IQ scores globally). The problem with outliers is that they can have a negative effect on the one-way ANOVA, reducing the validity of your results. Fortunately, when using SPSS Statistics to run a one-way ANOVA on your data, you can easily detect possible outliers. In our enhanced one-way ANOVA guide, we: (a) show you how to detect outliers using SPSS Statistics; and (b) discuss some of the options you have in order to deal with outliers. You can learn more about our enhanced one-way ANOVA guide on our Features: One-way ANOVA page.
- Assumption #5: Your dependent variable should be approximately normally distributed for each category of the independent variable. We talk about the one-way ANOVA only requiring approximately normal data because it is quite "robust" to violations of normality, meaning that assumption can be a little violated and still provide valid results. You can test for normality using the Shapiro-Wilk test of normality, which is easily tested for using SPSS Statistics. In addition to showing you how to do this in our enhanced one-way ANOVA guide, we also explain what you can do if your data fails this assumption (i.e., if it fails it more than a little bit). Again, you can learn more on our Features: One-way ANOVA page.
- Assumption #6: There needs to be homogeneity of variances. You can test this assumption in SPSS Statistics using Levene's test for homogeneity of variances. If your data fails this assumption, you will need to not only carry out a Welch ANOVA instead of a one-way ANOVA, which you can do using SPSS Statistics, but also use a different post hoc test. In our enhanced one-way ANOVA guide, we (a) show you how to perform Levene’s test for homogeneity of variances in SPSS Statistics, (b) explain some of

the things you will need to consider when interpreting your data, and (c) present possible ways to continue with your analysis if your data fails to meet this assumption, including running a Welch ANOVA in SPSS Statistics instead of a one-way ANOVA, and a Games-Howell test instead of a Tukey post hoc test (learn more on our Features: One-way ANOVA page).

You can check assumptions #4, #5 and #6 using SPSS Statistics. Before doing this, you should make sure that your data meets assumptions #1, #2 and #3, although you don't need SPSS Statistics to do this. Remember that if you do not run the statistical tests on these assumptions correctly, the results you get when running a one-way ANOVA might not be valid. This is why we dedicate a number of sections of our enhanced one-way ANOVA guide to help you get this right. You can find out about our enhanced one-way ANOVA guide on our Features: One-way ANOVA page, or more generally, our enhanced content as a whole on our Features: Overview page.

Test Procedure in SPSS Statistics

Follow this link https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php for Test Procedure in SPSS Statistics, we illustrate the SPSS Statistics procedure to perform a one-way ANOVA assuming that no assumptions have been violated. First, we set out the example we use to explain the one-way ANOVA procedure in SPSS Statistics.

2.5 SPSS Statistics

Example
A manager wants to raise the productivity at his company by increasing the speed at which his employees can use a particular spreadsheet program. As he does not have the skills in-house, he employs an external agency which provides training in this spreadsheet program. They offer 3 courses: a beginner, intermediate and advanced course. He is unsure which course is needed for the type of work they do at his company, so he sends 10 employees on the beginner course, 10 on the intermediate and 10 on the advanced course. When they all return from the training, he gives them a problem to solve using the spreadsheet program, and times how long it takes them to complete the problem. He then compares the three courses (beginner, intermediate, advanced) to see if there are any differences in the average time it took to complete the problem.

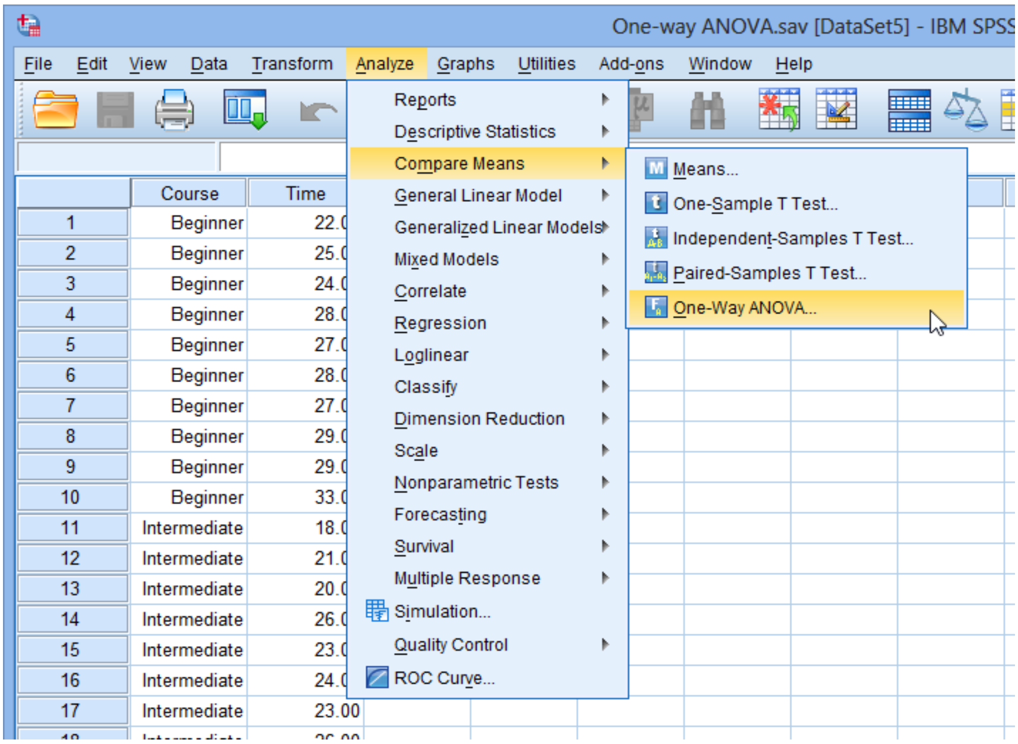
SPSS Statistics

1.5.1 Setup in SPSS Statistics

In SPSS Statistics, we separated the groups for analysis by creating a grouping variable called Course (i.e., the independent variable), and gave the beginners course a value of "1", the intermediate course a value of "2" and the advanced course a value of "3". Time to complete the set problem was entered under the variable name Time (i.e., the dependent variable). In our enhanced one-way ANOVA guide, we show you how to correctly enter data in SPSS Statistics to run a one-way ANOVA (see on our Features: One-way ANOVA page). You can learn about our enhanced data setup content in general on our Features: Data Setup. Alternately, see our generic, "quick start" guide: Entering Data in SPSS Statistics.

The eight steps below show you how to analyse your data using a one-way ANOVA in SPSS Statistics when the six assumptions in the previous section, Assumptions, have not been violated. At the end of these eight steps, we show you how to interpret the results from this test. If you are looking for help to make sure your data meets assumptions #4, #5 and #6, which are required when using a one-way ANOVA, and can be tested using SPSS Statistics, you can learn more on our Features: One-way ANOVA page

Click Analyze > Compare Means > One-Way ANOVA... on the top menu, as shown below.



Published with written permission from SPSS Statistics, IBM Corporation.

2.6 Descriptive Table

The descriptive table (see below) provides some very useful descriptive statistics, including the mean, standard deviation and 95% confidence intervals for the dependent variable (Time) for each separate group (Beginners, Intermediate and Advanced), as well as when all groups are combined (Total). These figures are useful when you need to describe your data.

ANOVA Table

This is the table that shows the output of the ANOVA analysis and whether there is a statistically significant difference between our group means. We can see that the significance value is 0.021 (i.e., $p = .021$), which is below 0.05. and, therefore, there is a statistically significant difference in the mean length of time to complete the spreadsheet problem between the different courses taken. This is great to know, but we do not know which of the specific groups differed. Luckily, we can find this out in the **Multiple Comparisons** table which contains the results of the Tukey post hoc test.

Multiple Comparisons Table

From the results so far, we know that there are statistically significant differences between the groups as a whole. The table below, **Multiple Comparisons**, shows which groups differed from each other. The Tukey post hoc test is generally the preferred test for conducting post hoc tests on a one-way ANOVA, but there are many others. We can see from the table below that there is a statistically significant difference in time to complete the problem between the group that took the beginner course and the intermediate course ($p = 0.046$), as well as between the beginner course and advanced course ($p = 0.034$). However, there were no differences between the groups that took the intermediate and advanced course ($p = 0.989$).

It is also possible to run comparisons between specific groups that you decided were of interest before you looked at your results. For example, you might have expressed an interest in knowing the difference in the completion time between the beginner and intermediate course groups.

This type of comparison is often called a planned contrast or a simple custom contrast. However, you do not have to confine yourself to the comparison between two time points only. You might have had an interest in understanding the difference in completion time between the beginner course group and the average of the intermediate and advanced course groups. This is called a complex contrast. All these types of custom contrast are available in SPSS Statistics. In our enhanced guide we show you how to run custom contrasts in SPSS Statistics using syntax (or sometimes a combination of the graphical user interface and syntax) and how to interpret and report the results. In addition, we also show you how to "trick" SPSS Statistics into applying a Bonferroni adjustment for multiple comparisons which it would otherwise not do.

2.6.1 Reporting the output of the one-way ANOVA

Based on the results above, you could report the results of the study as follows (N.B., this does not include the results from your assumptions tests or effect size calculations):

There was a statistically significant difference between groups as determined by one-way ANOVA ($F(2,27) = 4.467$, $p = .021$). A Tukey post hoc test revealed that the time to complete the problem was statistically significantly lower after taking the intermediate (23.6 ± 3.3 min, $p = .046$) and advanced (23.4 ± 3.2 min, $p = .034$) course compared to the beginners course (27.2 ± 3.0 min). There was no statistically significant difference between the intermediate and advanced groups ($p = .989$).

In our enhanced one-way ANOVA guide, we show you how to write up the results from your assumptions tests, one-way ANOVA and Tukey post hoc results if you need to report this in a dissertation, thesis, assignment or research report. We do this using the Harvard and APA styles (see our Features: One-way ANOVA page to learn more). It is also worth noting that in addition to reporting the results from your assumptions, one-way ANOVA and Tukey post hoc test, you are increasingly expected to report an effect size. Whilst there are many different ways you can do this, we show you how to calculate an effect size from your SPSS Statistics results in our enhanced one-way ANOVA guide. Effect sizes are important because whilst the one-way ANOVA tells you whether differences between group means are "real" (i.e., different in the population), it does not tell you the "size" of the difference. Providing an effect size in your results helps to overcome this limitation. You can learn more about our enhanced one-way ANOVA guide on our Features: One-way ANOVA page, or our enhanced content in general on our Features: Overview page.

SELF-ASSESSMENT EXERCISE 2

1. State the Assumptions
2. Highlight the Test Procedure in SPSS Statistics

3. Explain the SPSS Statistics 4. Demonstrate the Setup step in SPSS Statistics

2.6 Summary

This unit discussed the analysis of variance (ANOVA), ANOVA in SPSS, One-way ANOVA in SPSS Statistics, Assumptions and Test Procedure in SPSS Statistics. The specific test considered here is called analysis of variance (ANOVA) and is a test of hypothesis that is appropriate to compare means of a continuous variable in two or more independent comparison groups. For example, in some clinical trials there are more than two comparison groups. In a clinical trial to evaluate a new medication for asthma, investigators might compare an experimental medication to a placebo and to a standard treatment (i.e., a medication currently being used). In an observational study such as the Framingham Heart Study, it might be of interest to compare mean blood pressure or mean cholesterol levels in persons who are underweight, normal weight, overweight and obese.

2.7 References/Further Reading

statistics.laerd.com (2022). One-way ANOVA in SPSS Statistics. Retrieved from https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php

statistics.laerd.com (2022). ANOVA in SPSS Statistics. Retrieved from https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php

2.8 Possible Answers to SAEs

Answers to SAEs 1

1. Analysis of Variance (ANOVA) is the technique use to test the difference between more than two independent means. It is an extension of the two independent samples procedure which applies when there are exactly two independent comparison groups

2. The ANOVA Approach

Consider an example with four independent groups and a continuous outcome measure. The independent groups might be defined by a particular characteristic of the participants such as BMI (e.g., underweight, normal weight, overweight, obese) or by the investigator (e.g., randomizing participants to one of four competing treatments, call them A, B, C and D). Suppose that the outcome is systolic blood pressure, and we wish to test whether there is a statistically significant difference in mean systolic blood pressures among the four groups

3. Analysis of Variance, i.e. ANOVA in SPSS, is used for examining the differences in the mean values of the dependent variable associated with the effect of the controlled independent variables, after taking into account the influence of the uncontrolled independent variables. Essentially, ANOVA in SPSS is used as the test of means for two or more populations

4. ANOVA in SPSS.

In ANOVA in SPSS, there is one way ANOVA which involves only one categorical variable, or a single factor. For example, if a researcher wants to examine whether heavy, medium, light and nonusers of cereals differed in their preference for Total cereal, then the differences can be examined by the one way ANOVA in SPSS. In one way ANOVA in SPSS, a treatment is the same as the factor level

Answers to SAEs 2

1. Assumptions

When you choose to analyse your data using a one-way ANOVA, part of the process involves checking to make sure that the data you want to analyse can actually be analysed using a one-way ANOVA. You need to do this because it is only appropriate to use a one-way ANOVA if your data "passes" six assumptions that are required for a one-way ANOVA to give you a valid result. In practice, checking for these six assumptions just adds a little bit more time to your analysis, requiring you to click a few more buttons in SPSS Statistics when performing your analysis, as well as think a little bit more about your data, but it is not a difficult task

Assumption #1: Your dependent variable should be measured at the interval or ratio level (i.e., they are continuous).

Assumption #2: Your independent variable should consist of two or more categorical, independent groups

Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves.

Assumption #4: There should be no significant outliers. Outliers are simply single data points within your data that do not follow the usual pattern

Assumption #5: Your dependent variable should be approximately normally distributed for each category of the independent variable
Assumption #6: There needs to be homogeneity of variances. You can test this assumption in SPSS Statistics using Levene's test for homogeneity of variances

2. **Test Procedure in SPSS Statistics**

Follow this link https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php for Test Procedure in SPSS Statistics, we illustrate the SPSS Statistics procedure to perform a one-way ANOVA assuming that no assumptions have been violated. First, we set out the example we use to explain the one-way ANOVA procedure in SPSS Statistics.

3. SPSS Statistics

A manager wants to raise the productivity at his company by increasing the speed at which his employees can use a particular spreadsheet program. As he does not have the skills in-house, he employs an external agency which provides training in this spreadsheet program. They offer 3 courses: a beginner, intermediate and advanced course. He is unsure which course is needed for the type of work they do at his company, so he sends 10 employees on the beginner course, 10 on the intermediate and 10 on the advanced course. When they all return from the training, he gives them a problem to solve using the spreadsheet program, and times how long it takes them to complete the problem. He then compares the

three courses (beginner, intermediate, advanced) to see if there are any differences in the average time it took to complete the problem.
SPSS Statistics

4. Setup in SPSS Statistics

In SPSS Statistics, we separated the groups for analysis by creating a grouping variable called Course (i.e., the independent variable), and gave the beginners course a value of "1", the intermediate course a value of "2" and the advanced course a value of "3". Time to complete the set problem was entered under the variable name Time (i.e., the dependent variable). In our enhanced one-way ANOVA guide, we show you how to correctly enter data in SPSS Statistics to run a one-way ANOVA (see on our Features: One-way ANOVA page). You can learn about our enhanced data setup content in general on our Features: Data Setup. Alternately, see our generic, "quick start" guide: Entering Data in SPSS Statistics

UNIT 3 STATISTICAL PACKAGE FOR SOCIAL SCIENCE (SPSS) FOR ANALYSIS

Unit Structure

3.1 Introduction

This unit will examine the procedure for Statistical Package for the Social Sciences in Analyzing data.

3.2 Learning Outcomes

At the end of this unit, should be able to:

- Explain the meaning of Statistical Package for the Social Sciences (SPSS?)
- State the Common Uses of SPSS
- Mention the Data Requirements
- Explain the Hypotheses
- Demonstrate the Data Set-Up
- Explain the problem Statement
- Demonstrate the Running the Test
- Explain the decision and conclusions

3.3 Statistical Package for the Social Sciences (SPSS)

3.3.1 What is Statistical Package for the Social Sciences (SPSS?)

Statistical Package for the Social Sciences (SPSS) is a software program used by researchers in various disciplines for quantitative analysis of complex data. This introductory level SPSS unit introduces SPSS environment, basic data preparation and management, descriptive statistics, and common statistical analysis (T-test, ANOVA, correlation, regression). Hands-on-activity with example data is provided to learn these basics. This unit is a good start for people who are new to SPSS and help you harness this useful analysis tool. For a complete SPSS package follow this link: https://researchcommons.library.ubc.ca/introduction-to-spss-for-statistical-analysis/. The bivariate Pearson Correlation produces a sample correlation coefficient, *r*, which measures the strength and direction of linear relationships between pairs of continuous variables. By extension, the Pearson Correlation evaluates whether there is statistical evidence for a linear relationship among the same pairs of variables in the population, represented by a population correlation coefficient, ρ (“rho”).

- The Pearson Correlation is a parametric measure.
- This measure is also known as:
- Pearson’s correlation
- Pearson product-momen
- t correlation (PPMC)

3.3.2 Common Uses of SPSS

The bivariate Pearson Correlation is commonly used to measure the following:

- Correlations among pairs of variables
- Correlations within and between sets of variables
- The bivariate Pearson correlation indicates the following
- Whether a statistically significant linear relationship exists between two continuous variables
- The strength of a linear relationship (i.e., how close the relationship is to being a perfectly straight line)
- The direction of a linear relationship (increasing or decreasing)

Note: The bivariate Pearson Correlation cannot address non-linear relationships or relationships among categorical variables. If you wish to understand relationships that involve categorical variables and/or non-linear relationships, you will need to choose another measure of association.

Note: The bivariate Pearson Correlation only reveals *associations* among continuous variables. The bivariate Pearson Correlation does not provide any inferences about causation, no matter how large the correlation coefficient is.

3.3.3 Data Requirements

To use Pearson correlation, your data must meet the following requirements:

Two or more continuous variables (i.e., interval or ratio level)
Cases must have non-missing values on both variables
Linear relationship between the variables
Independent cases (i.e., independence of observations)
There is no relationship between the values of variables between cases.

This means that:
the values for all variables across cases are unrelated
for any case, the value for any variable cannot influence the value of any variable for other cases no case can influence another case on any variable
The biviariate Pearson correlation coefficient and corresponding significance test are not robust when independence is violated.

- Bivariate normality
- Each pair of variables is bivariately normally distributed
- Each pair of variables is bivariately normally distributed at all levels of the other variable(s)

This assumption ensures that the variables are linearly related; violations of this assumption may indicate that non-linear relationships among variables exist. Linearity can be assessed visually using a scatterplot of the data.

Random sample of data from the population
No outliers

3.4 Hypotheses

The null hypothesis (H_0) and alternative hypothesis (H_1) of the significance test for correlation can be expressed in the following ways, depending on whether a one-tailed or two-tailed test is requested:

Two-tailed significance test:

H_0: $\rho = 0$ ("the population correlation coefficient is 0; there is no association")
H_1: $\rho \neq 0$ ("the population correlation coefficient is not 0; a nonzero correlation could exist")
One-tailed significance test:
H_0: $\rho = 0$ ("the population correlation coefficient is 0; there is no association")
H_1: $\rho > 0$ ("the population correlation coefficient is greater than 0; a positive correlation could exist")
OR

H_1: $\rho < 0$ ("the population correlation coefficient is less than 0; a negative correlation could exist") where ρ is the population correlation coefficient.
Test Statistic

The sample correlation coefficient between two variables *x* and *y* is denoted *r* or r_{xy}, and can be computed as:

rxy=cov(x,y)var(x)−−−−−√⋅var(y)−−−−−√rxy=cov(x,y)var(x)⋅var(y)
where cov(*x*, *y*) is the sample covariance of *x* and *y*; var(*x*) is the sample variance of *x*; and var(*y*) is the sample variance of *y*.
Correlation can take on any value in the range [-1, 1]. The sign of the correlation coefficient indicates the direction of the relationship, while the magnitude of the correlation (how close it is to -1 or +1) indicates the strength of the relationship.

-1 : perfectly negative linear relationship
0 : no relationship
+1 : perfectly positive linear relationship
The strength can be assessed by these general guidelines (which may vary by discipline):
$.1 < | r | < .3$ … small / weak correlation
$.3 < | r | < .5$ … medium / moderate correlation
$.5 < | r |$ ……… large / strong correlation

Note: The direction and strength of a correlation are two distinct properties. The scatterplots below show correlations that are $r = +0.90$, $r = 0.00$, and $r = -0.90$, respectively. The strength of the nonzero correlations are the same: 0.90. But the direction of the correlations is different: a negative correlation corresponds to a decreasing relationship, while and a positive correlation corresponds to an increasing relationship.
$r = -0.90$

$r = 0.00$

$r = 0.90$

Note that the *r* = 0.00 correlation has no discernable increasing or decreasing linear pattern in this particular graph. However, keep in mind that Pearson correlation is only capable of detecting *linear* associations, so it is possible to have a pair of variables with a strong nonlinear relationship and a small Pearson correlation coefficient. It is good practice to create scatterplots of your variables to corroborate your correlation coefficients.

SELF-ASSESSMENT EXERCISE 1

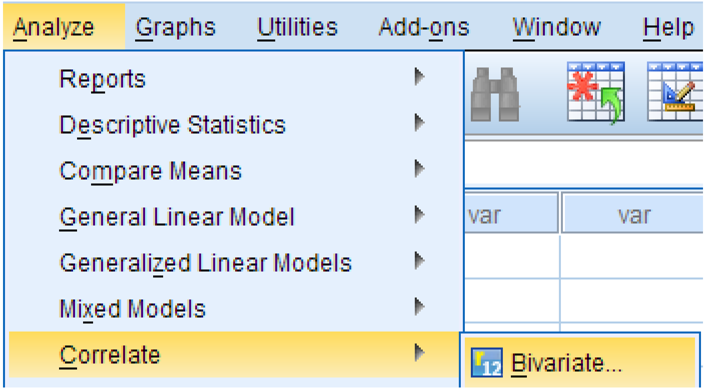
1. Explain the meaning of Statistical Package for the Social Sciences (SPSS?)
2. State the Common Uses of SPSS
3. Mention the Data Requirements
4. Explain the Hypotheses

3.4.1 Data Set-Up

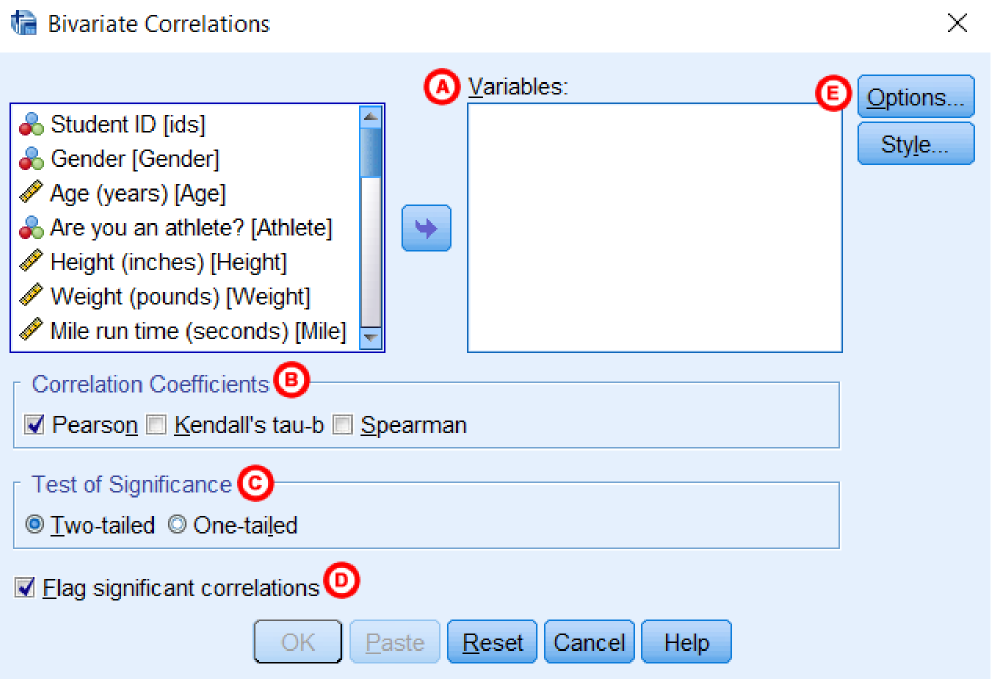
Your dataset should include two or more continuous numeric variables, each defined as scale, which will be used in the analysis.
Each row in the dataset should represent one unique subject, person, or unit. All of the measurements taken on that person or unit should appear in that row. If measurements for one subject appear on multiple rows -- for example, if you have measurements from different time points on separate rows -- you should reshape your data to "wide" format before you compute the correlations.

1. Run a Bivariate Pearson Correlation

To run a bivariate Pearson Correlation in SPSS, click **Analyze** > **Correlate** > **Bivariate**.



The Bivariate Correlations window opens, where you will specify the variables to be used in the analysis. All of the variables in your dataset appear in the list on the left side. To select variables for the analysis, select the variables in the list on the left and click the blue arrow button to move them to the right, in the **Variables** field.



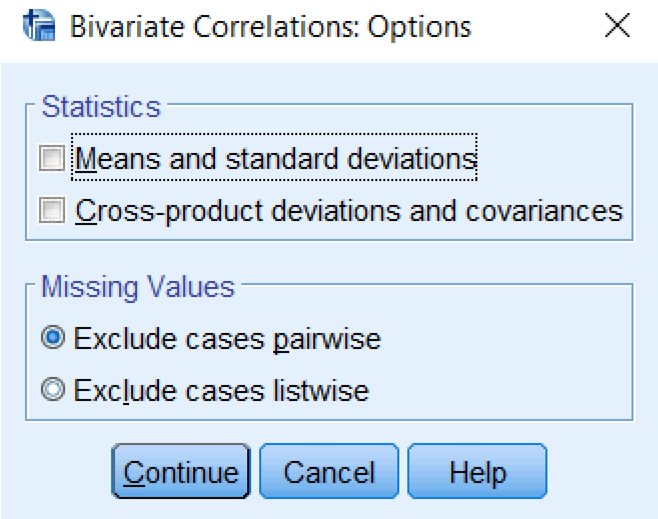
A Variables: The variables to be used in the bivariate Pearson Correlation. You must select at least two continuous variables, but may select more than two. The test will produce correlation coefficients for each pair of variables in this list.

B Correlation Coefficients: There are multiple types of correlation coefficients. By default, **Pearson** is selected. Selecting Pearson will produce the test statistics for a bivariate Pearson Correlation.

C Test of Significance: Click **Two-tailed** or **One-tailed**, depending on your desired significance test. SPSS uses a two-tailed test by default.

D Flag significant correlations: Checking this option will include asterisks (**) next to statistically significant correlations in the output. By default, SPSS marks statistical significance at the alpha = 0.05 and alpha = 0.01 levels, but not at the alpha = 0.001 level (which is treated as alpha = 0.01)

E Options: Clicking **Options** will open a window where you can specify which **Statistics** to include (i.e., **Means and standard deviations**, **Cross-product deviations and covariances**) and how to address **Missing Values** (i.e., **Exclude cases pairwise or Exclude cases listwise**). Note that the pairwise/listwise setting does not affect your computations if you are only entering two variable, but can make a very large difference if you are entering three or more variables into the correlation procedure.



Example: Understanding the linear association between weight and height

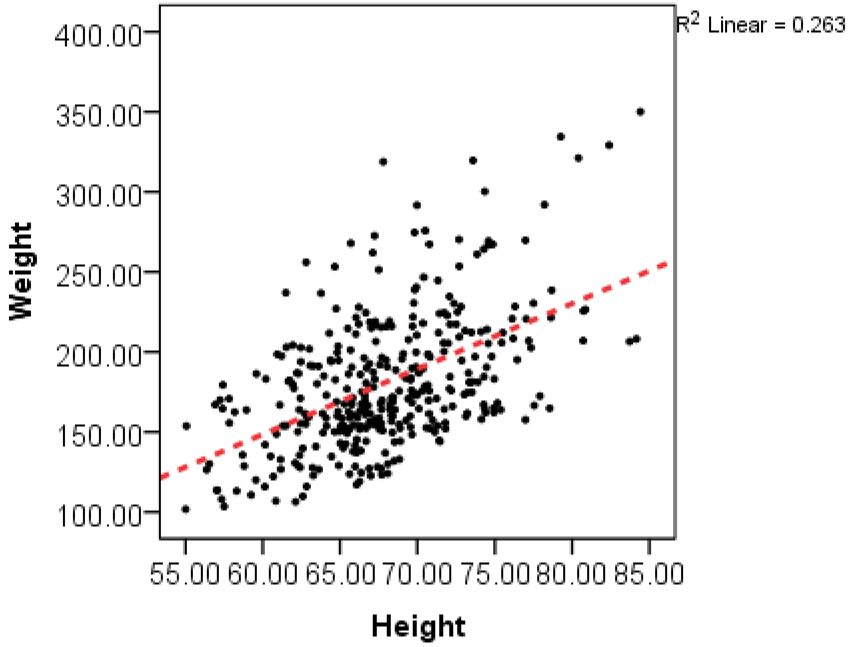
3.4.2 Problem Statement

Perhaps you would like to test whether there is a statistically significant linear relationship between two continuous variables, weight and height (and by extension, infer whether the association is significant in the population). You can use a bivariate Pearson Correlation to test whether there is a statistically significant linear relationship between height and weight, and to determine the strength and direction of the association.

Before the Test

In the sample data, we will use two variables: "Height" and "Weight." The variable "Height" is a continuous measure of height in inches and exhibits a range of values from 55.00 to 84.41 (**Analyze** > **Descriptive Statistics** > **Descriptives**). The variable "Weight" is a continuous measure of weight in pounds and exhibits a range of values from 101.71 to 350.07.

Before we look at the Pearson correlations, we should look at the scatterplots of our variables to get an idea of what to expect. In particular, we need to determine if it's reasonable to assume that our variables have linear relationships. Click **Graphs > Legacy Dialogs > Scatter/Dot**. In the Scatter/Dot window, click **Simple Scatter**, then click **Define**. Move variable Height to the X Axis box, and move variable Weight to the Y Axis box. When finished, click **OK**.



To add a linear fit like the one depicted, double-click on the plot in the Output Viewer to open the Chart Editor. Click **Elements > Fit Line at Total**. In the Properties window, make sure the Fit Method is set to **Linear**, then click **Apply**. (Notice that adding the linear regression trend line will also add the R-squared value in the margin of the plot. If we take the square root of this number, it should match the value of the Pearson correlation we obtain.)

From the scatterplot, we can see that as height increases, weight also tends to increase. There does appear to be some linear relationship.

3.5 Running the Test

To run the bivariate Pearson Correlation, click **Analyze** > **Correlate** > **Bivariate**. Select the variables Height and Weight and move them to the Variables box. In the **Correlation Coefficients** area, select **Pearson**. In the **Test of Significance** area, select your desired significance test, two-tailed or one-tailed. We will select a two-tailed significance test in this example. Check the box next to **Flag significant correlations**.

Click **OK** to run the bivariate Pearson Correlation. Output for the analysis will display in the Output Viewer.
Syntax

CORRELATIONS
/VARIABLES=Weight Height
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.

OUTPUT
Tables
The results will display the correlations in a table, labeled **Correlations**.

Correlations

		Height	Weight
Height	Pearson Correlation	1	.513**
	Sig. (2-tailed)		.000
	N	(A) 408	(B) 354
Weight	Pearson Correlation	.513**	1
	Sig. (2-tailed)	.000	
	N	(C) 354	(D) 376

**. Correlation is significant at the 0.01 level (2-tailed).

A Correlation of Height with itself (r=1), and the number of nonmissing observations for height (n=408).

B Correlation of height and weight (r=0.513), based on n=354 observations with pairwise nonmissing values.

C Correlation of height and weight (r=0.513), based on n=354 observations with pairwise nonmissing values.

D Correlation of weight with itself (r=1), and the number of nonmissing observations for weight (n=376).

The important cells we want to look at are either B or C. (Cells B and C are identical, because they include information about the same pair of variables.) Cells B and C contain the correlation coefficient for the correlation between height and weight, its p-value, and the number of complete pairwise observations that the calculation was based on.

The correlations in the *main diagonal* (cells A and D) are all equal to 1.

This is because a variable is always perfectly correlated with itself. Notice, however, that the sample sizes are different in cell A (*n*=408) versus cell D (*n*=376). This is because of missing data -- there are more missing observations for variable Weight than there are for variable Height.

If you have opted to flag significant correlations, SPSS will mark a 0.05 significance level with one asterisk (*) and a 0.01 significance level with two asterisks (0.01). In cell B (repeated in cell C), we can see that the Pearson correlation coefficient for height and weight is .513, which is significant ($p < .001$ for a two-tailed test), based on 354 complete observations (i.e., cases with nonmissing values for both height and weight).

1.5.2 Decision and Conclusions

Based on the results, we can state the following:

- Weight and height have a statistically significant linear relationship (r=.513, $p < .001$).
- The direction of the relationship is positive (i.e., height and weight are positively correlated), meaning that these variables tend to increase together (i.e., greater height is associated with greater weight).
- The magnitude, or strength, of the association is approximately moderate ($.3 < | r | < .5$).

SELF-ASSESSMENT EXERCISE 2

1. Demonstrate the Data Set-Up
2. Explain the problem Statement
3. Demonstrate the Running the Test
4. Explain the decision and conclusions

1.6 Summary

The unit explained Statistical Package for the Social Sciences (SPSS) is a software program used by researchers in various disciplines for quantitative analysis of complex data. This introductory level SPSS unit introduces SPSS environment, basic data preparation and management, descriptive statistics, and common statistical analysis (T-test, ANOVA, correlation, regression).

The bivariate Pearson Correlation is commonly used to measure the following:

a. Correlations among pairs of variables
b. Correlations within and between sets of variables
c. The bivariate Pearson correlation indicates the following
a. Whether a statistically significant linear relationship exists between two continuous variables
b. The strength of a linear relationship (i.e., how close the relationship is to being a perfectly straight line)
c. The direction of a linear relationship (increasing or decreasing)

To use Pearson correlation, your data must meet the following requirements:

i. Two or more continuous variables (i.e., interval or ratio level)
ii Cases must have non-missing values on both variables
i. Linear relationship between the variables
ii. Independent cases (i.e., independence of observations)

The null hypothesis (H_0) and alternative hypothesis (H_1) of the significance test for correlation can be expressed in the following ways, depending on whether a one-tailed or two-tailed test is requested:
Two-tailed significance test:

H_0: $\rho = 0$ ("the population correlation coefficient is 0; there is no association")

H_1: $\rho \neq 0$ ("the population correlation coefficient is not 0; a nonzero correlation could exist")
One-tailed significance test:

Your dataset should include two or more continuous numeric variables, each defined as scale, which will be used in the analysis.

Each row in the dataset should represent one unique subject, person, or unit. All of the measurements taken on that person or unit should appear in that row. If measurements for one subject appear on multiple rows
Perhaps you would like to test whether there is a statistically significant linear relationship between two continuous variables, weight and height (and by extension, infer whether the association is significant in the population). You can use a bivariate Pearson Correlation to test whether there is a statistically significant linear relationship between height and weight, and to determine the strength and direction of the association

To run the bivariate Pearson Correlation, click **Analyze** > **Correlate** > **Bivariate**. Select the variables Height and Weight and move them to the Variables box. In the **Correlation Coefficients** area, select **Pearson**. In the **Test of Significance** area, select your desired significance test, two-tailed or one-tailed

Based on the results, we can state the following:

Weight and height have a statistically significant linear relationship (r=.513, p < .001).

The direction of the relationship is positive (i.e., height and weight are positively correlated), meaning that these variables tend to increase together (i.e., greater height is associated with greater weight).
The magnitude, or strength, of the association is approximately moderate (.3 < | *r* | < .5).

3.7 References/Further Readings/Web Resources

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

UBC Library Research Commons (2022).Introduction to SPSS for statistical analysis. Retrieved from **https://researchcommons.library.ubc.ca/** introduction-to-spss-for-statistical-analysis/

Perry H. (2014). SPSS Explained. Retrieved from **https://researchcommons.library.ubc.ca/** introduction-to-spss-for-statistical-analysis/

Alan, B. (2012). Social Research Methods. Retrieved from **https://researchcommons.library.ubc.ca/** introduction-to-spss-for-statistical-analysis/

Ton J. C. A., Zwinderman, H. (2010). SPSS for Starters. Retrieved from **https://researchcommons.library.ubc.ca/** introduction-to-spss-for-statistical-analysis/

Rachad, A. (2012) Interpreting Quantitative Data with IBM SPSS Statistics. Retrieved from **https://researchcommons.library.ubc.ca/** introduction-to-spss-for-statistical-analysis/

3.8 Possible Answers to SAEs

Answers to SAEs 1

1 Statistical Package for the Social Sciences (SPSS) is a software program used by researchers in various disciplines for quantitative analysis of complex data. This introductory level SPSS unit introduces SPSS environment, basic data preparation and management, descriptive statistics, and common statistical analysis (T-test, ANOVA, correlation, regression).

2 Common Uses of SPSS
The bivariate Pearson Correlation is commonly used to measure the following:

i. Correlations among pairs of variables
ii. Correlations within and between sets of variables

The bivariate Pearson correlation indicates the following
Whether a statistically significant linear relationship exists between two continuous variables

The strength of a linear relationship (i.e., how close the relationship is to being a perfectly straight line)

The direction of a linear relationship (increasing or decreasing)
3 Data Requirements

To use Pearson correlation, your data must meet the following requirements:

Two or more continuous variables (i.e., interval or ratio level)
Cases must have non-missing values on both variables
Linear relationship between the variables
Independent cases (i.e., independence of observations)

4 Hypotheses

The null hypothesis (H_0) and alternative hypothesis (H_1) of the significance test for correlation can be expressed in the following ways, depending on whether a one-tailed or two-tailed test is requested:
Two-tailed significance test:

H_0: $\rho = 0$ ("the population correlation coefficient is 0; there is no association")

H_1: $\rho \neq 0$ ("the population correlation coefficient is not 0; a nonzero correlation could exist")
One-tailed significance test:

Answers to SAEs 2

1 Data Set-Up

Your dataset should include two or more continuous numeric variables, each defined as scale, which will be used in the analysis.

Each row in the dataset should represent one unique subject, person, or unit. All of the measurements taken on that person or unit should appear in that row. If measurements for one subject appear on multiple rows

2 Problem Statement
Perhaps you would like to test whether there is a statistically significant linear relationship between two continuous variables, weight and height (and by extension, infer whether the association is significant in the population). You can use a bivariate Pearson Correlation to test whether there is a statistically significant linear relationship between height and weight, and to determine the strength and direction of the association

3 Running the Test
To run the bivariate Pearson Correlation, click Analyze > Correlate > Bivariate. Select the variables Height and Weight and move them to the Variables box. In the Correlation Coefficients area, select Pearson. In the Test of Significance area, select your desired significance test, two-tailed or one-tailed

4 Decision and Conclusions
Based on the results, we can state the following:
Weight and height have a statistically significant linear relationship (r=.513, $p < .001$).

The direction of the relationship is positive (i.e., height and weight are positively correlated), meaning that these variables tend to increase together (i.e., greater height is associated with greater weight).
The magnitude, or strength, of the association is approximately moderate ($.3 < | r | < .5$).

UNIT 4 TIME SERIES ANALYSIS

Unit Structure

4.1 Introduction

This unit will be discussing Time Series Analysis. It will look at why organizations use time series data analysis, the types of Time Series Analysis and Models of time series analysis are discussed. Basic Objectives of the Analysis and Time Series forecasting will be discussed**.**

4.2 Learning Outcomes

At the end of this unit, should be able to:

1. Explain the concept of Time Series Analysis.
2. Explain why organizations use time series data analysis,
3. State the models of Time Series Analysis
4. Demonstrate the techniques of time series analysis
5. State the Basic Objectives of the Analysis
6. Explain the Time Series forecasting

4.3 Time Series Analysis

4.3.1 Concept of Time series analysis

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. However, this type of analysis is not merely the act of collecting data over time.

What sets time series data apart from other data is that the analysis can show how variables change over time. In other words, time is a crucial variable because it shows how the data adjusts over the course of the data points as well as the final results. It provides an additional source of information and a set order of dependencies between the data.

Time series analysis typically requires a large number of data points to ensure consistency and reliability. An extensive data set ensures you have a representative sample size and that analysis can cut through noisy data. It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance. Additionally, time series data can be used for forecasting—predicting future data based on historical data.

4.3.2 Why organizations use time series data analysis

Time series analysis helps organizations understand the underlying causes of trends or systemic patterns over time. Using data visualizations, business users can see seasonal trends and dig deeper into why these trends occur. With modern analytics platforms, these visualizations can go far beyond line graphs.

When organizations analyze data over consistent intervals, they can also use time series forecasting to predict the likelihood of future events. Time series forecasting is part of predictive analytics. It can show likely changes in the data, like seasonality or cyclic behavior, which provides a better understanding of data variables and helps forecast better.

For example, Des Moines Public Schools analyzed five years of student achievement data to identify at-risk students and track progress over time. Today's technology allows us to collect massive amounts of data every day and it's easier than ever to gather enough consistent data for comprehensive analysis

Time series analysis examples

Time series analysis is used for non-stationary data—things that are constantly fluctuating over time or are affected by time. Industries like finance, retail, and economics frequently use time series analysis because currency and sales are always changing. Stock market analysis is an excellent example of time series analysis in action, especially with automated trading algorithms. Likewise, time series analysis is ideal for forecasting weather changes, helping meteorologists predict everything from tomorrow's weather report to future years of climate change.

Examples of time series analysis in action include:

- Weather data
- Rainfall measurements
- Temperature readings
- Heart rate monitoring (EKG)
- Brain monitoring (EEG)
- Quarterly sales
- Stock prices
- Automated stock trading
- Industry forecasts
- Interest rates

4.4 Model of Time Series Analysis

Because time series analysis includes many categories or variations of data, analysts sometimes must make complex models. However, analysts can't account for all variances, and they can't generalize a specific model to every sample. Models that are too complex or that try to do too many things can lead to a lack of fit. Lack of fit or over fitting models lead to those models not distinguishing between random error and true relationships, leaving analysis skewed and forecasts incorrect.
Models of time series analysis include:

Classification: Identifies and assigns categories to the data.

Curve fitting: Plots the data along a curve to study the relationships of variables within the data.

Descriptive analysis: Identifies patterns in time series data, like trends, cycles, or seasonal variation.

Explanative analysis: Attempts to understand the data and the relationships within it, as well as cause and effect.

Exploratory analysis: Highlights the main characteristics of the time series data, usually in a visual format.

Forecasting: Predicts future data. This type is based on historical trends. It uses the historical data as a model for future data, predicting scenarios that could happen along future plot points.

Intervention analysis: Studies how an event can change the data.
Segmentation: Splits the data into segments to show the underlying properties of the source information.

Data classification
Further, time series data can be classified into two main categories:

Stock time series data means measuring attributes at a certain point in time, like a static snapshot of the information as it was.

Flow time series data means measuring the activity of the attributes over a certain period, which is generally part of the total whole and makes up a portion of the results.

Data variations
In time series data, variations can occur sporadically throughout the data:

Functional analysis can pick out the patterns and relationships within the data to identify notable events.

Trend analysis means determining consistent movement in a certain direction. There are two types of trends: deterministic, where we can find the underlying cause, and stochastic, which is random and unexplainable.

Seasonal variation describes events that occur at specific and regular intervals during the course of a year. Serial dependence occurs when data points close together in time tend to be related.
Time series analysis and forecasting models must define the types of data relevant to answering the business question. Once analysts have chosen the relevant data they want to analyze, they choose what types of analysis and techniques are the best fit.

4.4.1 Important Considerations for Time Series Analysis

While time series data is data collected over time, there are different types of data that describe how and when that time data was recorded. For example:

Time series data is data that is recorded over consistent intervals of time.

Cross-sectional data consists of several variables recorded at the same time.

Pooled data is a combination of both time series data and cross-sectional data

SELF-ASSESSMENT EXERCISE 1

1. Explain the concept of Time Series Analysis.
2. Explain why organizations use time series data analysis,
3. State the models of Time Series Analysis

4.5 Time Series Analysis Techniques

Just as there are many types and models, there are also a variety of methods to study data. Here are the three most common.

4.5.1 Box-Jenkins ARIMA models: These univariate models are used to better understand a single time-dependent variable, such as temperature over time, and to predict future data points of variables. These models work on the assumption that the data is stationary. Analysts have to account for and remove as many differences and seasonalities in past data points as they can. Thankfully, the ARIMA model includes terms to account for moving averages, seasonal difference operators, and autoregressive terms within the model.

4.5.2 Box-Jenkins Multivariate Models: Multivariate models are used to analyze more than one time-dependent variable, such as temperature and humidity, over time.

Holt-Winters Method: The Holt-Winters method is an exponential smoothing technique. It is designed to predict outcomes, provided that the data points include seasonality.

In this lesson, we'll describe some important features that we must consider when describing and modeling a time series. This is meant to be an introductory overview, illustrated by example, and not a complete look at how we model a univariate time series. Here, we'll only consider univariate time series. We'll examine relationships between two or more time series later on.

4.5.3 Univariate Time Series

A univariate time series is a sequence of measurements of the same variable collected over time. Most often, the measurements are made at regular time intervals.

One difference from standard linear regression is that the data are not necessarily independent and not necessarily identically distributed. One defining characteristic of a time series is that it is a list of observations where the ordering matters. Ordering is very important because there is dependency and changing the order could change the meaning of the data.

4.6 Basic Objectives of the Analysis

The basic objective usually is to determine a model that describes the pattern of the time series. Uses for such a model are:

- To describe the important features of the time series pattern.
- To explain how the past affects the future or how two time series can "interact".
- To forecast future values of the series.
- To possibly serve as a control standard for a variable that measures the quality of product in some manufacturing situations.

4.7 Types of Models

There are two basic types of "time domain" models.

1. Autoregressive Integrated Moving Average

Models that relate the present value of a series to past values and past prediction errors - these are called ARIMA models (for Autoregressive Integrated Moving Average). We'll spend substantial time on these.

Ordinary regression models that use time indices as x-variables. These can be helpful for an initial description of the data and form the basis of several simple forecasting methods.

Important Characteristics to Consider First
Some important questions to first consider when first looking at a time series are:

Is there a **trend**, meaning that, on average, the measurements tend to increase (or decrease) over time?

Is there seasonality, meaning that there is a regularly repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, and so on?

Are there outliers? In regression, outliers are far away from your line. With time series data, your outliers are far away from your other data.

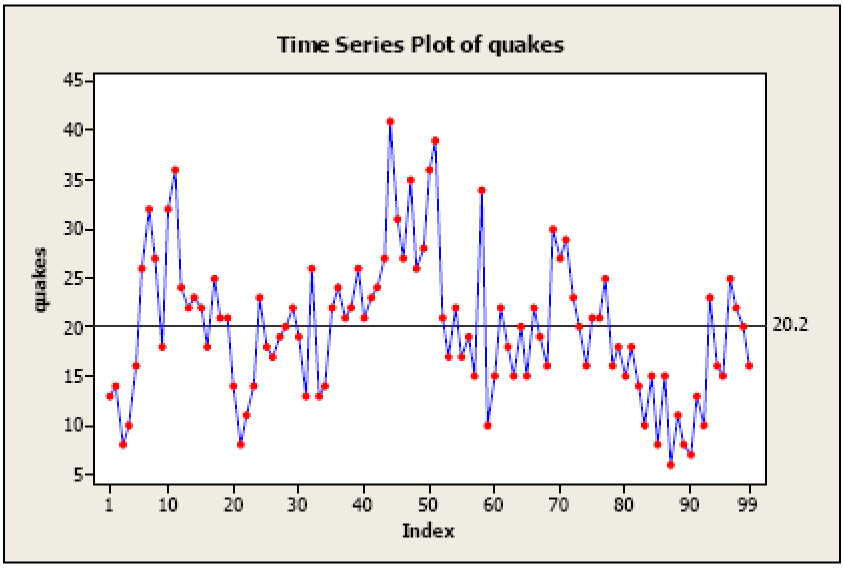
Is there a long-run cycle or period unrelated to seasonality factors?

Is there constant variance over time, or is the variance non-constant?

Are there any abrupt changes to either the level of the series or the variance?

Example 1-1

The following plot is a **time series plot** of the annual number of earthquakes in the world with seismic magnitude over 7.0, for 99 consecutive years. By a time series plot, we simply mean that the variable is plotted against time.



Some features of the plot:

There is no consistent trend (upward or downward) over the entire time span. The series appears to slowly wander up and down. The horizontal line drawn at quakes = 20.2 indicates the mean of the series. Notice that the series tends to stay on the same side of the mean (above or below) for a while and then wanders to the other side.

Almost by definition, there is no seasonality as the data are annual data. There are no obvious outliers.

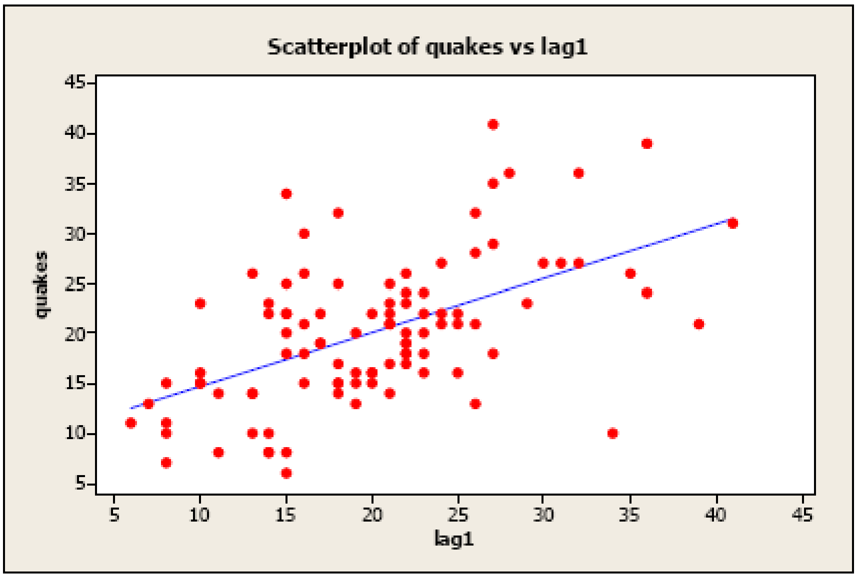
It's difficult to judge whether the variance is constant or not.

One of the simplest ARIMA type models is a model in which we use a linear model to predict the value at the present time using the value at the previous time. This is called an AR(1) model, standing for autoregressive model of order 1. The order of the model indicates how many previous times we use to predict the present time.

A start in evaluating whether an AR(1) might work is to plot values of the series against **lag 1 values** of the series. Let x_t denote the value of the series at any particular time t, so x_{t-1} denotes the value of the series one time before time t. That is, x_{t-1} is the lag 1 value of x_t. As a short example, here are the first five values in the earthquake series along with their lag 1 values:

T	**xt**	**xt−1 (lag 1 value)**
1	13	*
2	14	13
3	8	14
4	10	8
5	16	10

For the complete earthquake dataset, here's a plot of x_t versus x_{t-1} :



Although it's only a moderately strong relationship, there is a positive linear association so an AR(1) model might be a useful model.

The AR(1) model

Theoretically, the AR(1) model is written

$x_t=\delta+\phi_1 x_{t-1}+w_t$

Assumptions:

$w_t \sim iid N(0,\sigma_w^2)$, meaning that the errors are independently distributed with a normal distribution that has mean 0 and constant variance.

Properties of the errors w_t are independent of x.

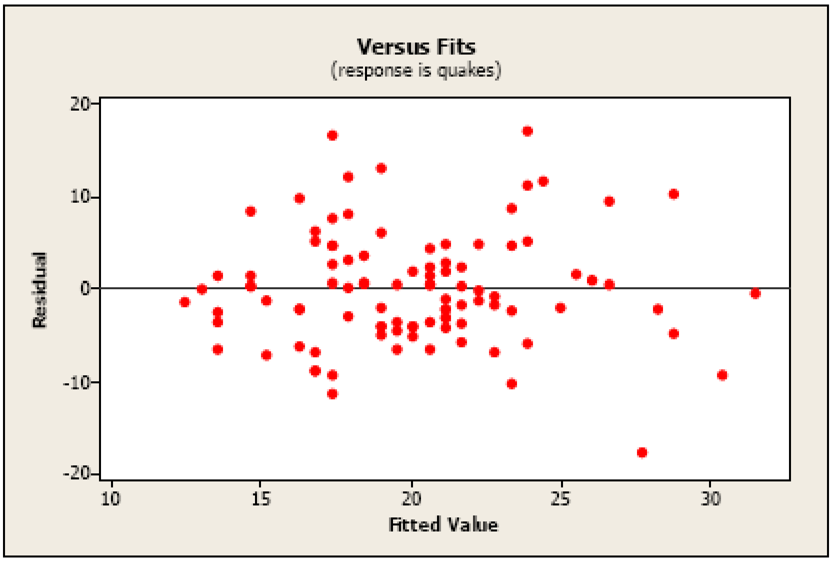
This is essentially the ordinary simple linear regression equation, but there is one difference. Although it's not usually true, in ordinary least squares regression we assume that the x-variable is not random but instead is something we can control. That's not the case here, but in our first

encounter with time series we'll overlook that and use ordinary regression methods. We'll do things the "right" way later in the course.
Following is Minitab output for the AR(1) regression in this example:
quakes = 9.19 + 0.543 lag1
98 cases used, 1 cases contain missing values

Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	9.191	1.819	5.05	0.000
lag1	0.54339	0.08528	6.37	0.000

S = 6.12239 R-Sq = 29.7% R-Sq(adj) = 29.0%
We see that the slope coefficient is significantly different from 0, so the lag 1 variable is a helpful predictor. The R2 value is relatively weak at 29.7%, though, so the model won't give us great predictions.
Residual Analysis

In traditional regression, a plot of residuals versus fits is a useful diagnostic tool. The ideal for this plot is a horizontal band of points. Following is a plot of residuals versus predicted values for our estimated model. It doesn't show any serious problems. There might be one possible outlier at a fitted value of about 28.

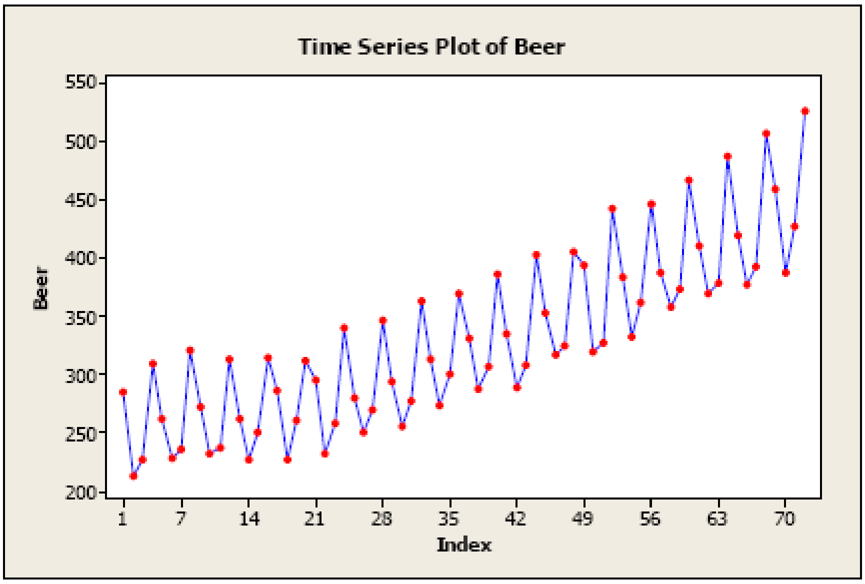


Example 1-2

The following plot shows a time series of quarterly production of beer in Australia for 18 years.

- Some important features are:
- There is an upward trend, possibly a curved one.
- There is seasonality – a regularly repeating pattern of highs and lows related to quarters of the year.

There are no obvious outliers.
There might be increasing variation as we move across time, although that's uncertain.



There are ARIMA methods for dealing with series that exhibit both trend and seasonality, but for this example, we'll use ordinary regression methods.

Classical regression methods for trend and seasonal effects

To use traditional regression methods, we might model the pattern in the beer production data as a combination of the trend over time and quarterly effect variables.

Suppose that the observed series is x_t, for $t=1,2,\ldots,n$.
For a linear trend, use t (the time index) as a predictor variable in a regression.

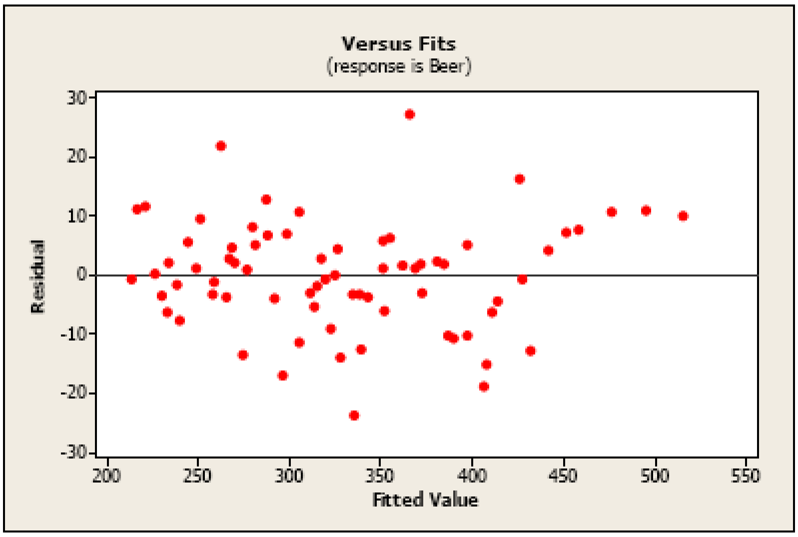
For a quadratic trend, we might consider using both t and t^2.

For quarterly data, with possible seasonal (quarterly) effects, we can define indicator variables such as $S_j=1$ if the observation is in quarter j of a year and 0 otherwise. There are 4 such indicators.

Let $\epsilon_t \sim iid N(0,\sigma^2)$. A model with additive components for linear trend and seasonal (quarterly) effects might be written
$x_t=\beta_1 t+\alpha_1 S_1+\alpha_2 S_2+\alpha_3 S_3+\alpha_4 S_4+\epsilon_t$

To add a quadratic trend, which may be the case in our example, the model is
xt=β1t+β2t2+α1S1+α2S2+α3S3+α4S4+ϵt

Note!
We've deleted the "intercept" from the model. This isn't necessary, but if we include it we'll have to drop one of the seasonal effect variables from the model to avoid collinearity issues.

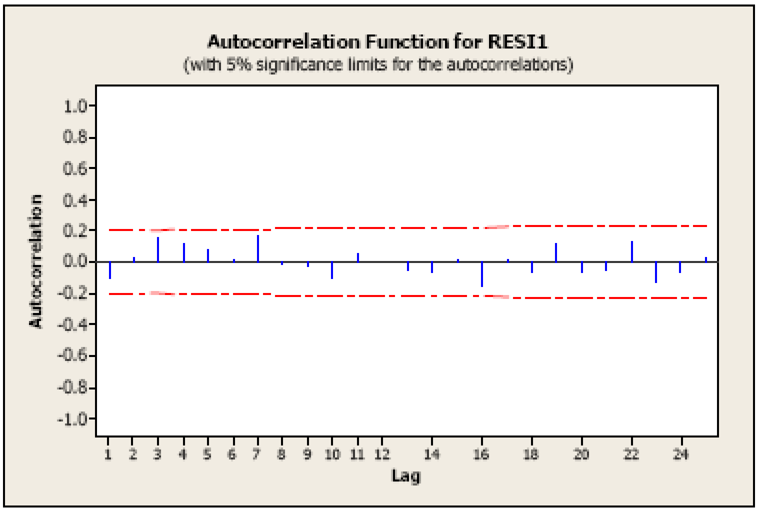


When data are gathered over time, we typically are concerned with whether a value at the present time can be predicted from values at past times. We saw this in the earthquake data of example 1 when we used an AR(1) structure to model the data. For residuals, however, the desirable result is that the correlation is 0 between residuals separated by any given time span. In other words, residuals should be unrelated to each other.
Sample Autocorrelation Function (ACF)

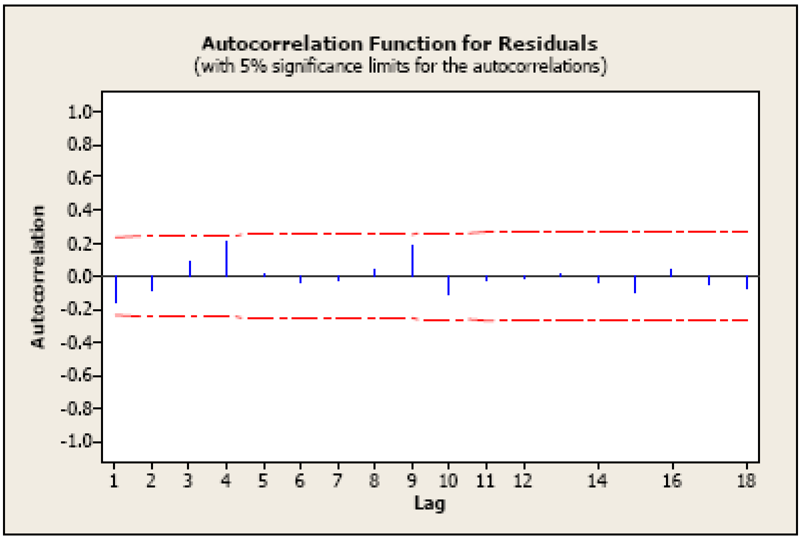
The sample autocorrelation function (ACF) for a series gives correlations between the series xt and lagged values of the series for lags of 1, 2, 3, and so on. The lagged values can be written as xt−1,xt−2,xt−3, and so on. The ACF gives correlations between xt and xt−1, xt and xt−2, and so on. The ACF can be used to identify the possible structure of time series data. That can be tricky going as there often isn't a single clear-cut interpretation of a sample autocorrelation function. We'll get started on that in Lesson 1.2 this week. The ACF of the residuals for a model is also useful. The ideal for a sample ACF of residuals is that there aren't any significant correlations for any lag.

Following is the ACF of the residuals for Example 1, the earthquake example, where we used an AR(1) model. The "lag" (time span between observations) is shown along the horizontal, and the autocorrelation is on the vertical. The red lines indicated bounds for statistical significance.

This is a good ACF for residuals. Nothing is significant; that's what we want for residuals.



The ACF of the residuals for the quadratic trend plus seasonality model we used for Example 2 looks good too. Again, there appears to be no significant autocorrelation in the residuals. The ACF of the residual follows:



4.8 Time Series forecasting

Time series forecasting is a technique for the prediction of events through a sequence of time. The technique is used across many fields of study, from the geology to behavior to economics. The techniques predict future events by analysing the trends of the past, on the assumption that future trends will hold similar to historical trends.

Consider the following periodic data:

Year	Y

1990	**50**
1991	**80**
1992	**90**
1993	**49**
1994	**75**
1995	**58**
1996	**82**
1997	**73**
1998	**95**

From the above data using the equation y = ax + b;

1. Calculate Slope a?
2. Calculate Intercept b
3. Assemble the equation of a line
4. Use Ŷ = â + bt to forecast the value of output, Y, for year 2003

SOLUTION

Let t representing period (years) and Y representing output
ty = 30708
y = 625
t = 45
t^2 = 285
$(t)^2$ = 2025

a) $b = \frac{N\,\Sigma(ty) - \Sigma t \Sigma y}{N\,\Sigma(t^2) - (\Sigma t)^2}$ $\quad \frac{9(3412)-45(625)}{9(285) - 2025}$ $\quad \frac{30708-29340}{2562 - 2025}$

$b = \frac{2583}{540} = b = 5$

b) $a = \frac{\Sigma y - b(\Sigma t)}{N}$

$\frac{625 - 5(45)}{9}$ $\quad \frac{625- 229.6}{9}$ $= 395.4/9 = 44$

c) Assemble of Regression equation line=

y = 44x + 5

d) Forecast the value of output, Y, for year 2003. Following the systematic process, the year 2003 is associated with the numerical value, t = 14, so that for t = 14,

To forecast value for output in year 2003, you have to use the

Ŷ = â + bt

Find the value for Ŷ, â and t =

Ŷ = 625/9 = 69.4

t = 45/9 = 5

â = Ŷ – bt

$\hat{a} = 69.4 - 5(5) = 44.4$

Fix the least – squares line: $\hat{Y} = \hat{a} + bt$

$\hat{Y} = 44.4 + 5t$

the forecast value for output in year 2003 is:

$Y = 44.4 + 5(14)$

Therefore, the forecast value for output in year 2003 is114.4

Self-Assessment Exercise 2

1. Demonstrate the techniques of time series analysis
2. State the Basic Objectives of the Analysis
3. Explain the Time Series forecasting

4.9 References/Further Reading

tableau.com(2022). Time Series Analysis Retrieved from https://www.tableau.com/learn/articles/time-series-analysis#definition.

4.10 Possible Answers to SAEs

Answers to SAEs 1

1. Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. However, this type of analysis is not merely the act of collecting data over time.
What sets time series data apart from other data is that the analysis can show how variables change over time

2. Why organizations use time series data analysis

Time series analysis helps organizations understand the underlying causes of trends or systemic patterns over time. Using data visualizations, business users can see seasonal trends and dig deeper into why these trends occur. With modern analytics platforms, these visualizations can go far beyond line graphs

Examples of time series analysis in action include:

- Weather data
- Rainfall measurements
- Temperature readings
- Heart rate monitoring (EKG)
- Brain monitoring (EEG)
- Quarterly sales
- Stock prices
- Automated stock trading
- Industry forecasts
- Interest rates

3. Model of Time Series Analysis

Because time series analysis includes many categories or variations of data, analysts sometimes must make complex models. However, analysts can't account for all variances, and they can't generalize a specific model to every sample. Models that are too complex or that try to do too many things can lead to a lack of fit. Lack of fit or over fitting models lead to those models not distinguishing between random error and true relationships, leaving analysis skewed and forecasts incorrect
Models of time series analysis include:

Classification: Identifies and assigns categories to the data.

Curve fitting: Plots the data along a curve to study the relationships of variables within the data.

Descriptive analysis: Identifies patterns in time series data, like trends, cycles, or seasonal variation.

Explanative analysis: Attempts to understand the data and the relationships within it, as well as cause and effect.

Exploratory analysis: Highlights the main characteristics of the time series data, usually in a visual format.

Forecasting: Predicts future data. This type is based on historical trends. It uses the historical data as a model for future data, predicting scenarios that could happen along future plot points.

Intervention analysis: Studies how an event can change the data.
Segmentation: Splits the data into segments to show the underlying properties of the source information

Answers to SAEs 2

1. Time Series Analysis Techniques

Box-Jenkins ARIMA models: These univariate models are used to better understand a single time-dependent variable, such as temperature over time, and to predict future data points of variable

Box-Jenkins Multivariate Models: Multivariate models are used to analyze more than one time-dependent variable, such as temperature and humidity, over time

Holt-Winters Method: The Holt-Winters method is an exponential smoothing technique. It is designed to predict outcomes, provided that the data points include seasonality

2. Basic Objectives of the Analysis

The basic objective usually is to determine a model that describes the pattern of the time series. Uses for such a model are:

- To describe the important features of the time series pattern.
- To explain how the past affects the future or how two time series can "interact".
- To forecast future values of the series.

- To possibly serve as a control standard for a variable that measures the quality of product in some manufacturing situations

3. Types of Models
There are two basic types of "time domain" models.

Autoregressive Integrated Moving Average
Models that relate the present value of a series to past values and past prediction errors - these are called ARIMA models (for Autoregressive Integrated Moving Average). We'll spend substantial time on these

3. Time Series forecasting
Time series forecasting is a technique for the prediction of events through a sequence of time. The technique is used across many fields of study, from the geology to behavior to economics. The techniques predict future events by analysing the trends of the past, on the assumption that future trends will hold similar to historical trends

UNIT 5: FORECASTING

Unit Structure

5.1 Introduction

This unit will be discussing forecasting. Forecasts are based on past performances. In other words, future values are predicted from past values. This assumes that the future will be basically the same as the past and present, implying that the relationships underlying the phenomenon of interest are stable overtime. Forecasting can be performed at different levels, depending on the use to which it will be put. Simple guessing, based on previous figures, is occasionally adequate. However, where there is a large investment at stake, structured forecasting is essential.
Any forecasts made, however technical or structured should be treated with caution, since the analysis is based on past data and there could be unknown factors present in the future. However it is often reasonable to assume that patterns that have been identified in the analysis of past data will be broadly continued, at least into the short-term future
In preparing plans for the future, the management authority has to make some predictions about what is likely to happen in the future. It shows that the managers know something of future happenings even before things actually happen. Forecasting provides them this knowledge.

5.2 Learning Outcomes

At the end of this unit, should be able to:

- Explain the concept of Forecasting?
- Outline the Steps in Forecasting
- Itemize the types of forecasting
- State the Forecasting Techniques
- Explain the Role of Forecasting
- Explain the Process of Forecasting:
- State and explain the Techniques of Forecasting:

5.3 Forecasting

5.3.1 What is Forecasting?

Forecasting is the process of estimating the relevant events of future, based on the analysis of their past and present behaviour.

Forecasts are based on past performances. In other words, future values are predicted from past values. This assumes that the future will be basically the same as the past and present, implying that the relationships underlying the phenomenon of interest are stable overtime**.**
The future cannot be probed unless one knows how the events have occurred in the past and how they are occurring presently. The past and present analysis of events provides the base helpful for collecting information about their future occurrence.
Thus, forecasting may be defined as the process of assessing the future normally using calculations and projections that take account of the past performance, current trends, and anticipated changes in the foreseeable period ahead.

Whenever the managers plan business operations and organisational set-up for the years ahead, they have to take into account the past, the present and the prevailing economic, political and social conditions. Forecasting provides a logical basis for determining in advance the nature of future business operations and the basis for managerial decisions about the material, personnel and other requirements.

It is, thus, the basis of planning, when a business enterprise makes an attempt to look into the future in a systematic and concentrated way, it may discover certain aspects of its operations requiring special attention. However, it must be recognised that the process of forecasting involves an element of guesswork and the managers cannot stay satisfied and relaxed after having prepared a forecast.

The forecast will have to be constantly monitored and revised—particularly when it relates to a long**-** term period. The managers should try to reduce the element of guesswork in preparing forecasts by collecting the relevant data using the scientific techniques of analysis and inference.

5.3.2 Steps in Forecasting

We outline the basic steps in forecasting as follows:
Step 1: Gather past data: daily, weekly, monthly and yearly.
Step 2: Adjust or clean up the raw data against inflationary factors. Index numbers can be used in deflating inflationary factors.
Step 3: Make forecasts from the "refined" data
Step 4: When the future data (which is been forecast) becomes available, compare forecasts with actual values, by so doing, you will be able to establish the error due to forecasting.

5.3.3 Types of forecasting

1. Short-term forecasts: these are forecasts concerning the near future. They, are characterized by few uncertainties and therefore more accurate then distant future forecasts
2. Long – term forecasts: these concern the distant future. They are characterized by more uncertainties than short – term forecasts
3. Extrapolation: these are forecasts based solely on past and present values of the variable to be forecast.
4. Forecasts based on established relationships between the variable to be forecast and other variables.

5.4 Forecasting Techniques

The two generally used methods of forecasting include:

i. Moving averages
ii. Trend lines or least squares.

On the basis of the definition, the following features of forecasting can be identified:

- Forecasting relates to future events.
- Forecasting is needed for planning process because it devises the future course of action.
- It defines the probability of happening of future events. Therefore, the happening of future events can be precise only to a certain extent.
- Forecasting is made by analysing the past and present factors which are relevant for the functioning of an organisation.

- The analysis of various factors may require the use of statistical and mathematical tools and techniques.

It helps the managers in planning; Forecasting is the key to planning. It generates the planning process. Planning decides the future course of action which is expected to take place in certain circumstances and conditions. Unless the managers know these conditions, they cannot go for effective planning.
Forecasting provides the knowledge of planning premises within which the managers can analyse their strengths and weaknesses and can take appropriate actions in advance before actually they are put out of market. Forecasting provides the knowledge about the nature of future conditions.

SELF-ASSESSMENT EXERCISE 1

1. Explain the concept of Forecasting?
2. Itemize the types of forecasting
3. State the Forecasting Techniques

5.4.1 Role of Forecasting

Since planning involves the future, no usable plan can be made unless the manager is able to take all possible future events into account. This explains why forecasting is a critical element in the planning process. In fact, every decision in the organisation is based on some sort of forecasting
Though forecasting cannot check the future happenings, it provides clues about those and indicates when the alternative actions should be taken. Managers can save their business and face the unfortunate happenings if they know in advance what is going to happen.

5.4.2 Process of Forecasting:

The process of forecasting generally involves the following steps:

1. Developing the Basis:
The future estimates of various business operations will have to be based on the results obtainable through systematic investigation of the economy, products and industry.

2. Estimation of Future Operations:
On the basis of the data collected through systematic investigation into the economy and industry situation, the manager has to prepare quantitative estimates of the future scale of business operations. Here the managers will have to take into account the planning premises.

3. Regulation of Forecasts:

It has already been indicated that the managers cannot take it easy after they have formulated a business forecast. They have to constantly compare the actual operations with the forecasts prepared in order to find out the reasons for any deviations from forecasts. This helps in making more realistic forecasts for future.

4. Review of the Forecasting Process:
Having determined the deviations of the actual performances from the positions forecast by the managers, it will be necessary to examine the procedures adopted for the purpose so that improvements can be made in the method of forecasting.

5.5 Techniques of Forecasting

There are various methods of forecasting. However, no method can be suggested as universally applicable. In fact, most of the forecasts are done by combining various methods.

A brief discussion of the major forecasting methods is given below:

1. Historical Analogy Method:
Under this method, forecast in regard to a particular situation is based on some analogous conditions elsewhere in the past. The economic situation of a country can be predicted by making comparison with the advanced countries at a particular stage through which the country is presently passing.

Similarly, it has been observed that if anything is invented in some part of the world, this is adopted in other countries after a gap of a certain time. Thus, based on analogy, a general forecast can be made about the nature of events in the economic system of the country. It is often suggested that social analogies have helped in indicating the trends of changes in the norms of business behaviour in terms of life.

Likewise, changes in the norms of business behaviour in terms of attitude of the workers against inequality, find similarities in various countries at various stages of the history of industrial growth. Thus, this method gives a broad indication about the future events of general nature.

2. Survey Method:
Surveys can be conducted to gather information on the intentions of the concerned people. For example, information may be collected through surveys about the probable expenditure of consumers on various items.

Both quantitative and qualitative information may be collected by this method.

On the basis of such surveys, demand for various products can be projected. Survey method is suitable for forecasting demand—both of existing and new products. To limit the cost and time, the survey may be restricted to a sample from the prospective consumers.

3. Opinion Poll:
Opinion poll is conducted to assess the opinion of the experienced persons and experts in the particular field whose views carry a lot of weight. For example, opinion polls are very popular to predict the outcome of elections in many countries including India. Similarly, an opinion poll of the sales representatives, wholesalers or marketing experts may be helpful in formulating demand projections (yourarticlelibrary.com, 2022).

If opinion polls give widely divergent views, the experts may be called for discussion and explanation of why they are holding a particular view. They may be asked to comment on the views of the others, to revise their views in the context of the opposite views, and consensus may emerge. Then, it becomes the estimate of future events.

4. Business Barometers:
A barometer is used to measure the atmospheric pressure. In the same way, index numbers are used to measure the state of an economy between two or more periods. These index numbers are the device to study the trends, seasonal fluctuations, cyclical movements, and irregular fluctuations.

These index numbers, when used in combination with one another, provide indications as to the direction in which the economy is proceeding. Thus, with the business activity index numbers, it becomes easy to forecast the future course of action.

However, it should be kept in mind that business barometers have their own limitations and they are not sure road to success. All types of business do not follow the general trend but different index numbers have to be prepared for different activities, etc.

5. Time Series Analysis:
Time series analysis involves decomposition of historical series into its various components, viz. trend, seasonal variances, cyclical variations, and random variances. When the various components of a time series are separated, the variation of a particular situation, the subject under study, can be known over the period of time and projection can be made about the future.

A trend can be known over the period of time which may be true for the future also. However, time series analysis should be used as a basis for forecasting when data are available for a long period of time and tendencies disclosed by the trend and seasonal factors are fairly clear and stable.

6. Regression Analysis:
Regression analysis is meant to disclose the relative movements of two or more inter-related series. It is used to estimate the changes in one variable as a result of specified changes in other variable or variables. In economic and business situations, a number of factors affect a business activity simultaneously.

Regression analysis helps in isolating the effects of such factors to a great extent. For example, if we know that there is a positive relationship between advertising expenditure and volume of sales or between sales and profit, it is possible to have estimate of the sales on the basis of advertising, or of the profit on the basis of projected sales, provided other things remain the same.

7. Input-Output Analysis:
According to this method, a forecast of output is based on given input if relationship between input and output is known. Similarly, input requirement can be forecast on the basis of final output with a given input-output relationship. The basis of this technique is that the various sectors of economy are interrelated and such inter-relationships are well-established.

For example, coal requirement of the country can be predicted on the basis of its usage rate in various sectors like industry, transport, household, etc. and how the various sectors behave in future. This technique yields sector-wise forecasts and is extensively used in forecasting business events as the data required for its application are easily obtained

Example

A monthly sales of ABC Company is given as:

Months	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**
Sale(s)	**50**	**55**	**70**	**50**	**45**	**90**

Using a 3 – period moving averages, forecast the sale for the three (3) months.

Solution

$\frac{50+55+70}{3}$ Jan = 175/3 = 58.3

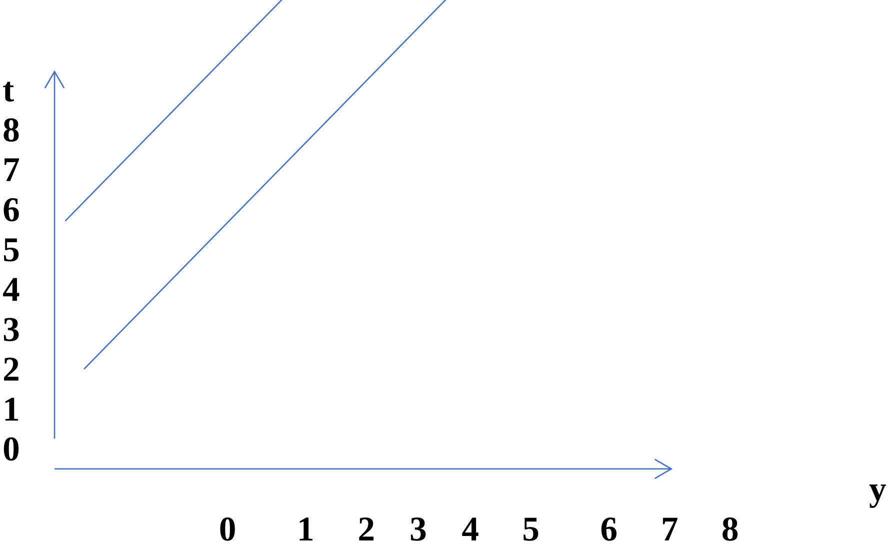
55+70+50 Feb = 175/3 = 58.3

3

$\frac{50+70+45}{3}$ Mar = 165/3 = 55.0

Suppose the line AB in the following straight line reasonably approximates a set of data for 1995 – 2000. Represent the above data in a diagram

Solution



SELF-ASSESSMENT EXERCISE 2

A monthly sales of ABC Company is given as:

Months	Jan	Feb	Mar	Apr	May	Jun
Sale(s)	60	65	70	75	80	90

1. Using a 3 – period moving averages, calculate the forecast sale for three months
2. State the Basic step in forecasting

1.6 Summary

Forecasting is the process of estimating the relevant events of future, based on the analysis of their past and present behaviour.

Forecasts are based on past performances. In other words, future values are predicted from past values. This assumes that the future will be basically the same as the past and present, implying that the relationships underlying the phenomenon of interest are stable overtime**.**

The unit itemized the types of forecasting to include; Short-term forecasts: these are forecasts concerning the near future. They, are characterized by few uncertainties and therefore more accurate then distant future forecasts. Long – term forecasts: these concern the distant future. They are characterized by more uncertainties than short – term forecasts. Extrapolation: these are forecasts based solely on past and present values of the variable to be forecast. Forecasts based on established relationships between the variable to be forecast and other variables.

The two generally used methods of forecasting include: Moving averages and Trend lines or least squares.

1.7 References/Further Readings/Web Resources

yourarticlelibrary.com (2022). forecasting/forecasting. Retrieved from 2022https://www.yourarticlelibrary.com/management/forecasting/forecasting-roles-steps-and-techniques-management-function/70032.

5.8 Possible Answers to SAEs

Answers to SAEs 1

1. What is Forecasting?

Forecasting is the process of estimating the relevant events of future, based on the analysis of their past and present behaviour.
Forecasts are based on past performances. In other words, future values are predicted from past values. This assumes that the future will be basically the same as the past and present, implying that the relationships underlying the phenomenon of interest are stable overtime**.**

2. Types of forecasting

1. Short-term forecasts: these are forecasts concerning the near future. They, are characterized by few uncertainties and therefore more accurate then distant future forecasts

2. Long – term forecasts: these concern the distant future. They are characterized by more uncertainties than short – term forecasts
5. Extrapolation: these are forecasts based solely on past and present values of the variable to be forecast.
6. Forecasts based on established relationships between the variable to be forecast and other variables.

3. Forecasting Techniques

The two generally used methods of forecasting include:

iii. Moving averages
iv. Trend lines or least squares.

Answers to SAEs 2
1.
A monthly sales of ABC Company is given as:

Months	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**
Sale(s)	**60**	**65**	**70**	**75**	**80**	**90**

Using a 3 – period moving averages, forecast the sale for three months
Solution

$\frac{65+70+75}{3}$ Jan = 210/3 = 70.0

$\underline{70+75+80}$ Feb = 225/3 = 75.0

$$\frac{75+80+90}{3} \quad \text{Mar} = 245/3 = 81.7$$

3

2. Basic step in forecasting

- Step 1: Gather past data: daily, weekly, monthly and yearly.
- Step 2: Adjust or clean up the raw data against inflationary factors.
- Step 3: Index numbers can be used in deflating inflationary factors.
- Step 4: Make forecasts from the "refined" data. When the future data (which is been forecast) becomes available, compare forecasts with actual values, by so doing, you will be able to establish the error due to forecasting.

MODULE 3

UNIT 1: INDEX NUMBERS

Unit Structure

1.1 Introduction

This unit will discuss the Index numbers. Index numbers are often used in stabilising the time value of money and in deflating nominal values. An index number measures the percentage change in the value of some economic commodity over a period of time. It will outline the types and Formula for Number Index.

1.2 Learning Outcomes

At the end of this unit, should be able to:

- Explain the Index numbers
- Outline the types and
- State and Demonstrate the Formula for Number Index

1.3 Index Number

1.3.1 What is Index Number?

An index number is a method of evaluating variations in a variable or group of variables in regards to geographical location, time, and other features. The base value of the index number is usually 100, which indicates price, date, level of production, and more.

There are various kinds of index numbers. However, at present, the most relatable is the price index number that particularly indicates the changes in the overall price level (or in the value of money) for a particular time.

According to Croxton and Cowden, index numbers are devices for measuring differences in the magnitude of a group of related variables.

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 locations, or other characteristics

Here, the value of money is not constant, even if it falls or rises it will affect and change the price level. An increase in the price level determines a decline in the value of money. A decrease in the price level means an increase in the value of money.

Therefore, the differences in the value of money are indicated by the differences in the overall price level for a particular time. Therefore, the changes in the overall prices can be evaluated by a statistical device known as 'index number (Byjus.com, 2022).'

1.4 Types of Index Numbers

Price index number: It evaluates the relative differences in costs between two particular points in time.

Quantity index number: It measures the differences in the physical quantity of the product's manufacturing, buying, or selling of one item or a group of items.
Index Numbers: meaning and characteristics

1.4.1 Important characteristics of index numbers

Following are the important characteristics of index numbers.	
(1) Expressed in Percentage	● A change in terms of the absolute values may not be comparable. ● Index numbers are expressed in percentage, so they remove this barrier. Although, we do not use the percentage sign. ● It is possible to compare the agricultural production and industrial production and at the same time being expressed in percentage, we can also compare the change in prices of different commodities.
(2) Relative measures or	● Index numbers measure a net or relative change in a variable or a group of variables.

measures of net changes	● For example, if the price of a certain commodity rises from ₹10 in the year 2007 to ₹15 in the year 2017, the price index number will be 150 showing that there is a 50% increase in the prices over this period.
(3) Measure change over a period of time or in two or more places	● Index numbers measure the net change among the related variables over a period of time or at two or more places. ● For example, change in prices, production, and more, over the two periods or at two places.
(4) Specialised average	● Simple averages like, mean, median, mode, and more can be used to compare the variables having similar units. ● Index numbers are specialised average, expressed in percentage, and help in measuring and comparing the change in those variables that are expressed in different units. ● For example, we can compare the change in the production of industrial goods and agricultural goods.
(5) Measuring changes that are not directly measurable	● Cost of living, business activity, and more are complex things that are not directly measurable. ● With the help of index numbers, it is possible to study the relative changes in such phenomena.

1.4.2 Advantages of index numbers

What are the advantages of index numbers?	
Index numbers are one of the most widely used statistical tools. Some of the advantages or uses of index numbers are as follows:	
(1) Help in formulating policies	● Most of the economic and business decisions and policies are guided by the index numbers. ● Example: ● To increase DA, the government refers to the cost-of-living index. ● To make any policy related to the industrial or agricultural production, the government refers to their respective index numbers.
(2) Help in study of trends	● Index numbers help in the study of trends in variables like, export-import, industrial and

	agricultural production, share prices, and more.
(3) Helpful in forecasting	● Index numbers not only help in the study of past and present behaviour, they are also used for forecasting economic and business activities.
(4) Facilitates comparative study	● To make comparisons with respect to time and place especially where units are different, index numbers prove to be very useful. ● For example, change in 'industrial production' can be compared with change in 'agricultural production' with the help of index numbers.
(5) Measurement of purchasing power of money to maintain standard of living	● Index numbers, such as cost inflation index help in measuring the purchasing power of money at different times between different regions. ● Such analysis helps the government to frame suitable policies for maintaining or raising the standard of living of the people.
(6) Act as economic barometer	● Index numbers are very useful in knowing the level of economic and business activities of a country. So, these are rightly known as economic barometers.

Problems involved in the construction of index numbers

Following are some of the problems involved in the construction of index numbers:	
(1) Purpose of index numbers	● Many different types of index numbers are constructed with different objectives. ● Example: Price index, quantity index, consumer price index, wholesale price index, and more ● So, the first important issue/problem is to define the objective for which the index number is to be constructed.
(2) Selection of base period	● Base period is the period against which the comparisons are made. ● Selection of a suitable base period is a very crucial step. ● It should be of reasonable length and normal one, i.e., it should not be affected by any abnormalities like,

	natural calamities, war, extreme business cycle situations. ● It should neither be too close nor too far.
(3) Selection of commodities	● All the items cannot be included in the construction of an index number. ● Nature and number of items to be included in an index number depends upon the type of index to be constructed. ● For example, to construct a 'consumer price index' those commodities should be considered that are generally consumed and the number should be neither too small nor too big.
(4) Selection of sources of data	● Depending upon the type of index numbers, the correct source should be selected for data. ● Like, to construct CPI, we need retail prices and to construct the wholesale price index, we need wholesale prices. Accordingly, the right and reliable source should be selected.
(5) Selection of weights	● The term 'weight' refers to the relative importance of different items in the construction of index numbers. ● All the items do not have the same importance. ● So, it is necessary to adopt some suitable measures to assign weight.
(6) Selection of an appropriate formula	● There are various formulas for construction of index numbers like Laspeyres' method, Paasche's method, Fisher's method, and more. ● No single formula is appropriate for all types of index numbers. ● The choice of formula depends upon the purpose of the available data.

(A) Consumer price index number	● Consumer price index (CPI) measures changes in the cost of living due to changes in the retail prices of a basket of goods over a period of time. ● Separate cost of living index is prepared for different classes of people. ● It is also known as the cost of living index numbers or retail price index number.
The uses of consumer price index number:	
(1) Helpful in measuring the	● Consumer price index has an inverse relation with the purchasing power of money.

purchasing power of money	● Purchasing power of money = 1/Consumer price index ● As CPI increases, the purchasing power of money decreases.
(2) Helpful in wage negotiations	● CPI helps in determining wages for a particular class. ● It provides the basis for wage negotiations between the workers and employers.
(3) Help government in framing policies	● These index numbers provide guidelines for the formulation of wage policy, price policy, taxation policy, and other general economic policies.
(4) Market analysis	● CPI also helps a market analyst to determine the demand for different goods and services.
(5) Help businessmen in forecasting	● On the basis of CPI of different classes of people, a businessman can make predictions about the demand for his products.

The wholesale price index numbers. What is the utility of wholesale price index numbers?	
(A) Meaning of wholesale price index (WPI)	● Wholesale price index measures the changes in the wholesale prices of the commodities. ● It indicates the change in general price level in the economy. ● In India, it is prepared on a weekly basis. ● These days 2004-05 is considered as the base year
Utility of wholesale price index (WPI)	
(1) Indicator of inflation	● Inflation is a persistent rise in the general price level. ● WPI is used to determine the rate of inflation in an economy.
(2) Forecasting demand and supply	● It is often used to forecast demand and supply situations in the economy. ● An increase in WPI indicates a situation of excess demand over supply of goods. ● A decrease in WPI indicates a situation of excess supply of goods.

(3) Helps in determining real changes in aggregates	● WPI helps us to find out the real changes in aggregates like national income, national expenditure, and more. ● Using WPI of the current year and the base year, we can convert national income at current prices into national income at constant prices.
(4) Cost of projects	● It determines the future cost of long-run[1] projects. ● If WPI has an increasing trend, it will result in an increase in prices of various goods used in the projects. ● As a result the cost of such a project will go up.

Following are the major limitations of index numbers:	
(1) Difficulty in construction of index numbers	● The decision of objective, selection of base period, selection of commodities, selection of sources of data, selection of 'weights', selection of formula, and more are the several difficulties in the construction of index numbers.
(2) Based on sample items, so only approximate indicators	● Index numbers are generally based on a few sample items. So, the results derived are approximate and not perfect.
(3) Ignores quality of commodities	● These days the quality changes occur very fast and the index numbers ignore this aspect. ● So, the results shown by these may not be appropriate.
(4) Limited use	● There is no 'master index number' or 'all in one index number'. ● Use of each index number is restricted to its specific object.
(5) Useful only for short-term comparison	● Over a period of time, rapid changes occur in habits, tastes, preferences, and more. ● So,the index number constructed in the present may not be comparable with the one constructed a few years back.

Self-Assessment Exercise 1

Multiple Choice Questions

Q.1 According to ______________, the index numbers are devices for measuring differences in the magnitude of a group of related variables.

Q.2 Which of the following is the characteristic of an index number?
Q.3 Index numbers measure a net or relative change in a variable or a group of variables

1.5 Formula for Number Index

Index numbers are expressed in terms of a base of 100.
INDEX NUMBER could be Simple Average or Price Relatives Method or Aggregate Average. In this method, we find out the price relative of individual items and average out the individual values. Price relative refers to the percentage ratio of the value of a variable in the current year to its value in the year chosen as the base.

Price relative Index (R) = (P1÷P2) × 100
Example

Calculate the Relative Price Index Time Series for Monthly Salaries of workers

Years	Average Monthly Salaries (N)
1985	1,200
1990	2,000
1995	1,800
2000	3,600
2005	5000
2010	6500

Solution

Years	Average Monthly Salaries (N)	Salary Index (1985 =100)
1985	1,200	100%
1990	2,000	200/1200 = 1.66x100 = 167%
1995	1,800	1800/2000 = 0.9x100 = 90%
2000	3,600	3600/1800 = 2%x100 = 200%
2005	5000	5000/3600 = 1.39x100 = 139%
2010	6500	6500/5000 = 1.3x100 = 130%

Calculate the Simple Aggregate for Average Consumer Prices for Selected Staple Food

Items	**2000**	**2006**
Sugar	10	40
Wheat Flour	11	20
Butter	71	99
Ground beef	91	186
Frying Chicken	39	88

Solution

For Year 2006
40+20+99+186+88
10+11+71+91+39
=195.06 This implies that the prices of the five items are higher by 95.06 percent in 2006 than in 2000.

Self-Assessment Exercise 2
The weighted average method of forecasting to calculate the Share of the following companies.

	No of Share	**No of Share**	**Price Per Share (N)**	**Price Per Share (N)**
Company	2004	2006	2004	2006
A	350	400	0.50	1.25
B	200	180	1.25	3.75
C	140	200	6.25	12.50
D	130	150	12.50	18.75

1.6 Summary

INDEX NUMBER could be Simple Average or Price Relatives Method or Aggregate Average. In this method, we find out the price relative of individual items and average out the individual values. Price relative refers to the percentage ratio of the value of a variable in the current year to its value in the year chosen as the base.

1.7 References/Further Readings/Web Resources

Byjus.com (2022) meaning and characteristics of index numbers https://byjus.com/commerce/meaning-and-characteristics-of-index-numbers/

1.8 Possible Answers to SAEs

Answers to SAEs 1

Q.1 According to _______________, the index numbers are devices for measuring differences in the magnitude of a group of related variables.

a. Spiegel
b. Croxton and Cowden
c. Both (a) and (b)
d. None of the above

Q.2 Which of the following is the characteristic of an index number?

a. Measure change over a period of time or in two or more places
b. Specialised average
c. Expressed in percentage
d. All of the above

Q.3 Index numbers measure a net or relative change in a variable or a group of variables.

a. Absolute
b. Relative
c. Both (a) and (b)
d. None of the above

Answers to SAEs 2

1 Solution

1 .25(400)+3.75(180)+12 .50(200)+18.75(150)
0.50(350)+ 1.25(200)+ 6.25(140)+12.5(130)
500+675+2500+281.5
175+250+875+1625X 100
2.218X100
221.79

UNIT 2: INVENTORY CONTROL

Unit Structure

2.1 Introduction

This unit will be discussing inventory control, **Why Do We Need Inventory Control, and** Types of Inventory Control Systems.

2.2 Learning Outcomes

At the end of this unit, should be able to:

- Explain the meaning of inventory control
- Find out w**hy Do We Need Inventory Control**
- **Outline the t**ypes of Inventory Control Systems

2.3 inventory control

2.3.1 What do mean by inventory control

Inventory control, also known as stock control, refers to the process of managing a company's warehouse inventory levels. The inventory control process involves managing items from the moment they're ordered; throughout their storage, movement, and usage; and to their final destination or disposal.
An inventory control system is a technology solution that manages and tracks a company's goods through the supply chain. This technology will integrate and manage purchasing, shipping, receiving, warehousing, and returns into a single

system. The best inventory control system will automate a lot of manual processes
The four types of inventory most commonly used are Raw Materials, Work-In-Process (WIP), Finished Goods, and Maintenance, Repair, and Overhaul (MRO). You can practice better inventory control and smarter inventory management when you know the type of inventory you have

1.3.2 Why Do We Need Inventory Control?

Inventory control is a vital part of any business's operations and significantly impacts its ability to make a profit.

Here are four major reasons you need inventory control**:**

- Boost inventory efficiency. Your products may be all over the warehouse, making picking and packing take much longer. Inventory control lets you discover and optimize this and many other issues like an incorrect number to increase efficiency and streamline your processes.
- Ensure accurate inventory counts. Inaccurate inventory causes headaches and increases costs. Inventory control gives your insight into inventory numbers and helps determine the reorder point and sales trends.
- Increase sales and mitigate losses. Fill rate is essential for businesses to keep up with customer demand and avoid stockouts. Having a higher fill rate will help increase sales and mitigate losses. Safety stock can ensure that you always have the product available to meet customer demand.
- Ensure customer satisfaction. Customers hate when products are out of stock or on backorder. Controlling your inventory lets you avoid these issues and keep your customers happy.

2.3.3 Relevance of Inventory

Storing inventory has many associated costs that a manager needs to control. Without inventory control, a business's warehouse can quickly become a problem.

By allowing inventory to move about with no interference, a manager risks running into skyrocketing costs and plummeting profits. This, in turn, will lead to the loss of their job and possibly the closure of the business. That's why it's necessary to monitor cycle inventory to ensure standing inventory is at a sufficient level.

2.4 Inventory Solution

When attempting to track and manage inventory, there are a few ways to track your data. First, you can take the old-school approach and manually track inventory using a pen and paper. While this is quick and easy to do, it’s also easy to make mistakes. We don't recommend this approach, regardless of the size of your inventory.

Second, you can use an Excel spreadsheet. We offer a free inventory tracking spreadsheet on our blog, and this is a viable option for a business with a limited inventory or just starting. You'll still need to perform a regular inventory audit or cycle inventory count to ensure accuracy, but it's a step above using a pen and paper.

Finally, the best option is to use an inventory control system and other related inventory tracking software.

SELF-ASSESSMENT EXERCISE 1

1. What is Inventory control
2. Outline the Two major reasons you need inventory control

2.5 Types of Inventory Control Systems

Businesses use inventory control systems to measure the number of goods on hand. Large organizations frequently track inventory at several locations, including shops, warehouses, and websites. **There are two main types of inventory control systems: periodic and perpetual inventory systems.**
Now let's take a closer look at these two:

2.5.1 The Periodic Inventory System

The periodic inventory system is one of small businesses' most commonly used inventory systems. Under this system, businesses keep track of their inventory levels at set intervals, typically once a month. This system is simple to use and understand, making it a popular choice for business owners.
At the end of each interval, businesses will count their inventory levels and update their records accordingly. This system can be used with either physical inventories or virtual inventories.

2.5.2 Benefits to using the periodic inventory System

There are four benefits to using the periodic inventory system**:**
It is easy to set up and maintain.
It provides accurate information on inventory levels.
It allows businesses to manage their inventories more efficiently.
It helps businesses avoid stock-outs and overstocking.
The Perpetual Inventory System

The perpetual inventory system is crucial to most businesses' inventory management systems. This system provides near-real-time data on inventory levels, allowing businesses to manage their stock more effectively.
There are two benefits to using a perpetual inventory system**:**

It allows businesses to avoid the costly and time-consuming process of physically counting their inventory. This saves businesses significant time and money on the labor cost.
This information can be used to decide production and stocking levels. and stocking levels. It helps businesses to predict future demand more accurately. With accurate data, businesses can avoid overstocking or understocking their products, leading to lost sales or missed opportunities.

2.6 Inventory Control System and Software

One of the best ways to take control of your business's inventory is to invest in inventory management software. This software tracks inventory levels, sales trends, and inventory cycles. There are many options on the market with a variety of capabilities and additional tools.

These programs can be tied to your POS system to provide a perpetual inventory count. This updates your inventory levels each time a sale is made. This lets you make the most informed decisions, calculate optimal reorder points, plan for product lead time, and stock more A-level products.
Automated Inventory Control

Automated inventory control takes a perpetual inventory count to the next level and makes decisions using predetermined rules. This feature is built into some of the best inventory control software platforms and allows you to take a more hands-off role.

Here's an example of how this automatic inventory control can work. You may set a minimum supply threshold for reorder if you have a particular product that you always want to keep in stock. Once you hit this number, the system automatically sends a new purchase order to your manufacturer. You can optimize this process with the optimal economic order quantity

SELF-ASSESSMENT EXERCISE 2

1.	Relevance of Inventory
2.	Inventory Solution
3.	Types of Inventory Control Systems

2.7 Summary

Inventory control, also known as stock control, refers to the process of managing a company's warehouse inventory levels. The inventory control process involves managing items from the moment they're ordered; throughout their storage, movement, and usage; and to their final destination or disposal
Boost inventory efficiency. Your products may be all over the warehouse, making picking and packing take much longer. Inventory control lets you discover and optimize this and many other issues like an incorrect number to increase efficiency and streamline your processes. Ensure accurate inventory counts. Inaccurate inventory causes headaches and increases costs. Inventory

control gives your insight into inventory numbers and helps determine the reorder point and sales trends.

2.7 References/Further Readings/Web Resources

Byjus.com (2022) meaning and characteristics of index numbers https://byjus.com/commerce/meaning-and-characteristics-of-index-numbers/

2.8 Possible Answers to SAEs

Answers to SAEs 1

1 Inventory control
Inventory control, also known as stock control, refers to the process of managing a company's warehouse inventory levels. The inventory control process involves managing items from the moment they're ordered; throughout their storage, movement, and usage; and to their final destination or disposal

2 Two major reasons you need inventory control**:**

a. Boost inventory efficiency. Your products may be all over the warehouse, making picking and packing take much longer. Inventory control lets you discover and optimize this and many other issues like an incorrect number to increase efficiency and streamline your processes.

b. Ensure accurate inventory counts. Inaccurate inventory causes headaches and increases costs. Inventory control gives your insight into inventory numbers and helps determine the reorder point and sales trends

Answers to SAEs 2

1 Relevance of Inventory
Storing inventory has many associated costs that a manager needs to control. Without inventory control, a business's warehouse can quickly become a problem.

By allowing inventory to move about with no interference, a manager risks running into skyrocketing costs and plummeting profits

2 Inventory Solution
When attempting to track and manage inventory, there are a few ways to track your data. First, you can take the old-school approach and manually track inventory using a pen and paper. While this is quick and easy to do, it’s also easy to make mistakes. We don’t recommend this approach, regardless of the size of your inventory

3 Types of Inventory Control Systems
Businesses use inventory control systems to measure the number of goods on hand. Large organizations frequently track inventory at several locations, including shops, warehouses, and websites. **There are two main types of inventory control systems: periodic and perpetual inventory systems**

UNIT 3: ECONOMIC ORDER QUANTITY (EOQ)

Unit Structure

3.1 Introduction

This unit discussed the **Economic Order Quantity** and the **Importance of Economic Order Quantity**. The optimal quantity is the exact amount of inventory you should order and keep on hand to meet demand. Finding your optimal order quantity for a product is the goal of calculating its EOQ. However, this number is very difficult to achieve as any slight variance in demand, cost, or price will throw your numbers off.

3.2 Learning Outcomes

At the end of this unit, should be able to:

- Explain the Economic order quantity (EOQ)
- Outline the **Importance of Economic Order Quantity**
- Explain the **Economic Order Quantity Problems**
- State the **Advantages of Economic Order Quantity**
- Demonstrate the **EOQ Formula**

3.3 Economic order quantity (EOQ)

3.3.1 What is Economic order quantity (EOQ?)

Economic order quantity (EOQ) is a production-scheduling method of inventory control that has been used since the early 1900s. This method is built around finding a balance between the amount you sell and the amount you spend on inventory management process.

Economic order quantity is the ideal amount of product a company should purchase to minimize inventory costs. Essentially, it is the amount of product you should order to meet demand without having to store any excess inventory.

3.3.1 Importance of Economic Order Quantity

Managing economic order quantity is an important skill to have when someone is vying for an inventory manager salary. It can help avoid issues like excess stock or dead stock (see what is dead stock) and keep avoidable losses to a minimum. It also helps you establish goals for your inventory KPIs, informs inventory forecasting decisions, and helps increase the company's sales and revenue.

Advantages of Economic Order Quantity
Utilizing EOQ for your business can provide many benefits. Here are just a few.

Minimize costs: All warehouse inventory managers know that storage costs can quickly rise if inventory isn't controlled. By only ordering the amount needed to fulfill customer demand, these costs can be kept very low.
Adapts to your business. Many inventory methods are only viable for certain types of business. EOQ utilizes only your own numbers, so it can benefit any business that uses it.

1.3.1 Economic Order Quantity Problems

Though there are definitely positive aspects of calculating EOQ, there are also a few drawbacks that you need to be aware of.

The math is complicated. You'll see the formula used for EOQ calculations below, and it's safe to say it isn't the easiest to use. Luckily, there are many ways to automate the process and tools to help.

It's based on assumptions. There are a number of assumptions that are required to calculate EOQ. This means any aspect of your business that doesn't match will throw off the numbers and you won't get the optimal quantity. Still, the numbers you find are very helpful for inventory planning

3.4 Advantages of Economic Order Quantity

Utilizing EOQ for your business can provide many benefits. Here are just a few.

Minimize costs. All warehouse inventory managers know that storage costs can quickly rise if inventory isn't controlled. By only ordering the amount needed to fulfill customer demand, these costs can be kept very low. This can be tough if your supplier requires an MOQ (what does MOQ mean?).

Adapts to your business. Many inventory methods are only viable for certain types of business. EOQ utilizes only your own numbers, so it can benefit any business that uses it.

SELF-ASSESSMENT EXERCISE 1

What is Economic order quantity (EOQ?)
Importance of Economic Order Quantity

3.5 EOQ Formula: Economic Order Quantity Formula

Calculating economic order quantity requires you first find a few metrics regarding demand and costs. These are the annual demand for the product in units, the cost per order, and the annual holding cost per unit. Once you've collected this data, it's as easy as plugging them into the formula below.

How to Calculate EOQ

Uncovering the economic order quantity for a product can be done using a slightly complicated formula.
Here's that formula:
EOQ = √ (2 x Demand x Order Cost / Holding Cost)

Economic Order Quantity Formula and Example

If any of that seems confusing to you, let's clear it up a bit with an example.
Let's say you are a wholesale supplier for the food industry. You have a particular product you're looking to optimize, in this case cans of creamed corn. The first thing you do is look at your historical data regarding creamed corn (Bluecart.com, 2022).

After poring through your data, you calculate that you normally sell an average of 2,500 cans each year. You also look through your purchase orders and inventory costs to calculate that each shipment of 100 cans of corn costs $75. And you find that storage of each can costs you $20 per year.

With these variables in hand, you can now calculate your optimal EOQ for cans of creamed corn. Let's plug them in.

EOQ = √ (2 x Demand x Order Cost / Holding Cost)
EOQ = √ (2 x 2500 x 75 / 20)
EOQ = 136.9 or 137 cans

We discover that the optimal order size is 137 cans of creamed corn. Pair this with calculating the optimal reorder point, and you can maximize the profit you make from cans of corn.

Ordering (Replacement) Costs

These are such costs as transportation costs, clerical and administrative costs associated with the physical movement of the purchased external goods. Where the goods are manufactured internally, there are alternative initial costs to be borne with each production run referred to as set-up costs

2. Holding (Carrying) Costs These are:

- storage costs in terms of staffing, equipment maintenance, and handling;
- storage overheads (heat, light, rent, and the like);
- cost of capital tied up in inventory;
- insurance, security and pilferage;
- deterioration or breakages.

3. Stock out Costs

These are costs associated with running out of stock. These include penalty payments, loss of goodwill, idle manpower and machine, and the like.

Economic Ordering Quantity (EOQ): This refers to the external order quantity that minimises total inventory costs.

Economic Batch Quantity (EBQ): This refers to the size of the internal production run that minimises total inventory costs.

Safety Stock: This is a term used to describe the stock held to cover possible deviations in demand or supply during the lead time. It is sometimes referred to as buffer or minimum stock.

Maximum Stock: This is a level used as an indicator above which stocks are deemed to be too high.

Reorder Level: This is the level of stock, which when reached, signals replenishment order.

Reorder Quantity: This is the level of replenishment order.

SELF-ASSESSMENT EXERCISE 2

1. Discuss the Re-order, Lro, Lmin and the Level System in the inventory control system
2. A commodity has a steady rate of demand of 2,000 units per year. Placing an order costs N200 and it costs N50 to hold a unit for a year:

a. Estimate the Economic Order Quantity (EOQ)
b. Find the number of orders placed per year
c. What is the length of the inventory circle?

3.6 Summary

This unit explain that, Economic order quantity (EOQ) is a production-scheduling method of inventory control that has been used since the early 1900s. This method is built around finding a balance between the amount you sell and the amount you spend on inventory management process.

Economic order quantity is the ideal amount of product a company should purchase to minimize inventory costs. Essentially, it is the amount of product you should order to meet demand without having to store any excess inventory

Managing economic order quantity is an important skill to have when someone is vying for an inventory manager salary. It can help avoid issues like excess stock or dead stock (see what is dead stock) and keep avoidable losses to a minimum. It also helps you establish goals for your inventory KPIs, informs inventory forecasting decisions, and helps increase the company's sales and revenue

3.7 References/Further Readings/Web Resources

Bluecart.com (2022). Inventory control and economic-order-quantity .Retreived from https://www.bluecart.com/blog/economic-order-quantity#toc-eoq-formula-economic-order-quantity-formula

3.8 Possible Answers to SAEs

Answers to SAEs 1

1. This is the most commonly used control system. It generally results in lower stocks. The system also enables items to be ordered in more economic quantities and is more responsive to fluctuations in demand than the second system discussed below. The system sets the value of three important levels of stock as warning t or action triggers for management:

Re-order Level: This is an action level of stock which leads to the replenishment order, normally the Economic Order Quantity (EOQ). For a particular time period, the re-order level is computed as follows:

Lro = maximum usage per period x maximum lead time (in periods)
Minimum Level: This is a warning level set such that only in extreme cases (above average demand or late replenishment) should it be breached. It is computed as follows:

Lmin = Re-Order Level – (normal Usage x Average lead time)
2. Given information, D = 2,000; Co = 200; and, Ch = 50
The Economic Order Quantity is determined by:
EOQ = {2DCoCh} {2(2000)(200) 50}= 16000= 126.491
Square root of 16000 is 126.491
Thus, the economic order quantity is about 127 units.
(b) Number of orders per year = Yearly Demand
EOQ= 2000/126.491 = 15.81
What is average inventory level = 63.5

Answers to SAEs 2

1. What is Economic order quantity (EOQ?)
Economic order quantity (EOQ) is a production-scheduling method of inventory control that has been used since the early 1900s. This method is built around finding a balance between the amount you sell and the amount you spend on inventory management process.

Economic order quantity is the ideal amount of product a company should purchase to minimize inventory costs. Essentially, it is the amount of product you should order to meet demand without having to store any excess inventory
2. **Importance of Economic Order Quantity**

Managing economic order quantity is an important skill to have when someone is vying for an inventory manager salary.

UNIT 4 DECISION ANALYSIS

Unit Structure

4.1 Introduction

This unit will be discussing Decision analysis, Decision trees and backward induction

Typically, more than one decision is involved in decision making, in which case it is best to use a tree instead of a matrix.

4.2 Learning Outcomes

At the end of this unit, should be able to:

- Explain the concept of Decision Analysis
- Outline the Basic Elements decision theory
- Discuss the Decision Criteria
- Explain the Decision trees and backward induction
- Itemize the Steps in Decision Analysis

4.3 Decision Analysis

4.3.1 What is Decision Analysis?

Decision analysis is a formalized approach to making optimal choices under conditions of uncertainty. It allows the user to enter costs, probabilities, and health-related quality of life values among other inputs of interest, and then calculates probabilistically weighted means of these outcome measures. In public health, these outcome measures usually include costs. Typically, therefore, decision analysis is the heart of cost-effectiveness analyses in public health and medicine (Gold et al., 1998).

However, just about any outcome measure can be modeled, including vaccine-preventable illnesses averted, deaths avoided, and so forth. Therefore, local health departments, pharmaceutical companies, or other agencies can use decision analysis for internal decision-making processes. Decision analysis is often used by non-health businesses interested in deciding whether they should release a product, perform internal restructuring, and so forth.

One great strength of decision analysis modeling is that it allows for the calculation of a range of possible values around a given mean. This approach, called 'sensitivity analysis,' allows the user to better understand the chances that he or she will make a bad decision if a given strategy is taken.

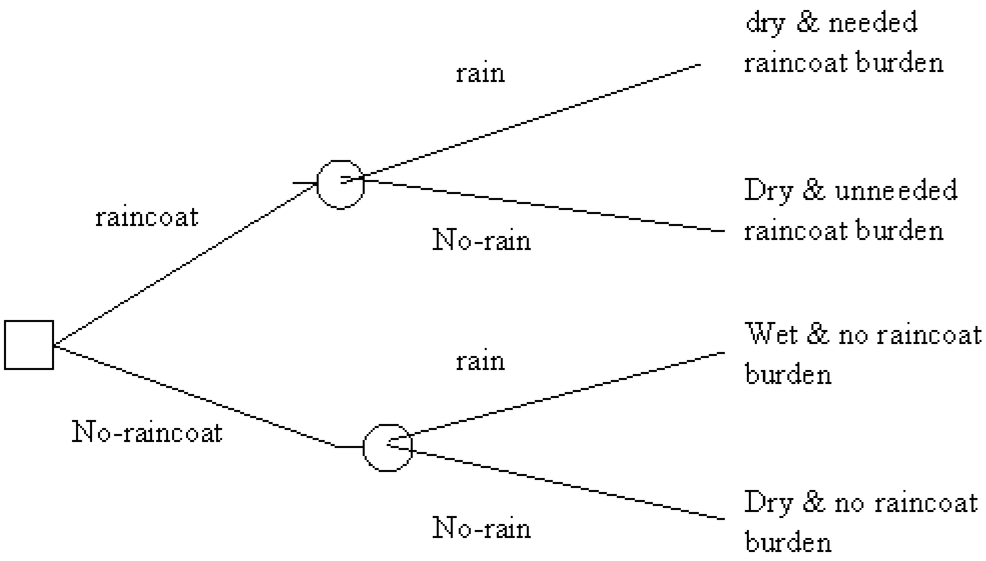
Decision analysis, like cost-effectiveness analysis, is highly dependent on the accuracy and completeness of model inputs, as well as the assumptions that the analysts make. Drugs can have unforeseen side effects, or interventions can have long-term costs that may not be apparent to the analysts. Any of these effects can lead to suboptimal outcomes.

For instance, the optimal treatment strategy for tuberculosis in most instances is a low-cost combination of medications that can be effectively delivered in developing countries. By using the most cost-effective medications, it is possible to maximize the number of lives saved within a given budget. However, as Farmer points out, these medications will be wasted if delivered to populations with a high percentage of drug-resistant tuberculosis (Farmer, 2004). Therefore, decision analysis and cost-effectiveness analysis must be viewed as an adjunct to optimal decision making rather than the final word in health policy.

Decision is sets and recommended rules on how to evaluate the overall environmental, technical, and socioeconomic performance of a set of alternatives and how to choose among them. The steps for decision process are to: define the problem and formulate objectives, find decision alternatives and predict the impacts of each, rank the alternatives and decide. The implementation of the decision is accompanied by regulatory, legal and economic processes, by stakeholders involvement, and by evaluation of the correctness of the decision by following the real responses after the implementation of the project.

4.3.2 Basic Elements decision theory

There are 4 basic elements in decision theory: *acts, events, outcomes,* and *payoffs*. Acts are the actions being considered by the agent -in the example elow, taking the raincoat or not; events are occurrences taking place outside the control of the agent (rain or lack thereof); outcomes are the result of the occurrence (or lack of it) of acts and events (staying dry or not; being burdened by the raincoat or not); payoffs are the values the decision maker is placing on the occurrences (for example, how much being free of the nuisance of carrying an raincoat is worth to one). Payoffs can be positive (staying dry) or negative (the raincoat nuisance). It is often useful to represent a decision problem by a tree.



Here a square indicates a node in the tree where a decision is made and a circle where events take place. The tree does not contain payoffs yet, but they can easily be placed by the outcomes.

In general, we can note two things. First, the nature of the payoffs depends on one's *objectives*. If one is interested only in making money, then payoffs are best accounted for in terms of money. However, if one is interested in, say, safety, then the payoffs are best accounted for in terms of risk of accident, for example. If any numerical approach is possible when disparate objectives are involved, there must be some universal measurable quantity making them comparable. (In fact, utility, of which more later, is such a quantity). Second, decision making trees can become unmanageable very fast if one tries to account for *too many possibilities*. For example, it would be physically impossible to account for all the possibilities involved in the decision of which 50 out of 200 gadgets should be intensively marketed, as the number of possible combinations, $200!/(50! \cdot 150!)$ is simply astronomical. Hence one must use good judgment in limiting the options considered; this is potentially problematic, as one may unwittingly fail to consider a possible action which would produce very good outcomes.

4.3.3 Decision Criteria

How one uses a decision tree or a decision matrix depends on the decision criteria one adopts. Consider the following *payoff matrix* where acts are rows, events columns, and the resulting squares contain the payoffs (outcomes are not represented to avoid clutter). So, suppose that we are considering which widget out of 3 to produce and our goal is making money.

	EVENTS	
	Good sales	Bad sales
ACTS		
Produce A	+$5000	-$1000
Produce B	+$10000	-$3000
Produce C	+$3000	-$500

Here producing B is obviously the best option if things go well, while producing C is the best option if things go badly, as losing $500 is the best of the worst payoffs. The decision criterion telling us to choose C is called "Maximin". Obviously maximin is a rather pessimistic strategy, and for this reason it is controversial. However, if the stakes are very high (for example, suppose that if I lose more than $500 I will be forever ruined), maximin seems a reasonable option. The application of maximin in the original position has played an important role in Rawls' *A Theory of Justice*, the most important work in political philosophy in the last decades. Other decision criteria in cases of uncertainty are *maximax*, *minimax of regret*, and the appeal to subjective probabilities through the *Principle of Indifference*. Unfortunately, none of these principles is always viable.

However, when the probabilities of events are available (that is, in decision under risk) and the agent is indifferent to risk, as when the payoffs involved are *significant but not too significan*t, the criterion usually put forth in decision theory is that of the *expected maximum payoff* (EMP), the counterpart of the principle in gambling enjoining us to choose the bet with the greatest expected value. So, suppose that we could provide the relevant probabilities, as in the following matrix:

	EVENTS			
	Good sales	Bad sales	Payoff	Expected payoff
ACTS				
Produce A Pr(good sales)= 80% Pr(bad sales) =20%	+$5000 x 80% = 4000	-$1000 x 20% = -200	+$4000	+$3800
Produce B Pr(good sales)= 60% Pr(bad sales) =40%	+$10000 x 60% = 6000	-$3000 x 40% = -1200	+$7000	+$4800
Produce C Pr(good sales)= 50% Pr(bad sales) =50%	+$3000 x 50% = $1500	-$500 x 50% = -$250	+$2500	+$1250

Then, EMP would tell us to produce B, as the expected payoff is the greatest. Most business decisions fall into this category. For example, if a company makes dozens of decisions with comparable payoffs every day, then EMP is the best business strategy, as it is for a casino.

SELF-ASSESSMENT EXERCISE 1

1. Explain the concept of Decision Analysis
2. Outline the Basic Elements decision theory
3. Discuss the Decision Criteria

4.4 Decision trees and backward induction

Typically, more than one decision is involved in decision making, in which case it is best to use a tree instead of a matrix. For example, consider the following situation, in which no probabilities are involved.

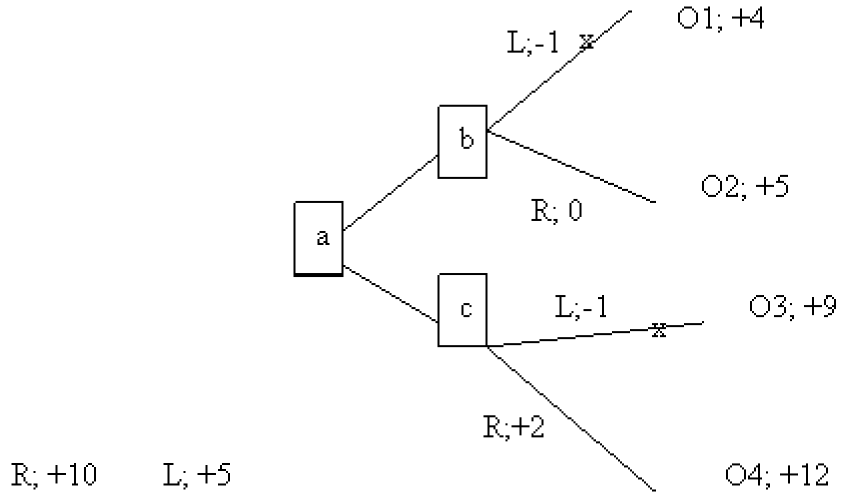
You have arrived at a fork in the road on you way home.

If you go left, you'll have to go through a very sunny patch in the mid of a very hot day. However, this will also allow you to admire a beautiful bloom of wildflowers growing by the side of the path. You shall then arrive at another fork. If you take another left, you will have to go by a neighbor's house with a very unpleasant guard dog that will growl at you from the other side of the fence. By contrast, if you go right at the second fork, you'll go by a very noisy and dusty part of the road. Whichever of the two you take, you shall get home quickly.

If you go right at the first fork, you'll go through the woods, which are very cool this time of the year. However, there will be little to admire until you get to another fork in the road. If you go left at this fork, you will see some beautiful meadows; unfortunately, it will take you longer to reach your home, a bad thing since you are in a bit of a hurry. If you go right, you shall get home in good time.

Suppose that you assign the following utilities:

Getting hot: -10; seeing the wildflowers: +15; being growled at: -3; pleasant coolness: +10; being a bit late getting home: -5; taking a noisy and dusty road: -2; seeing the nice meadows: +4; getting home in good time: +2.
We can construct a decision tree.



Decision trees can be used by applying *backwards induction*. The idea is that in order to determine what to do at a (the decision at the first fork), one needs to decide what one would do at b and c. In other words, the tree is analyzed form the right (from the outcomes) to the left (to the earlier decisions). So, at b, one would take the right path because it leads to outcome O2 with utility is +5 while the utility of the left path leading to O1 is +4. We can represent this choice by pruning the left path, that is, by placing an 'x' on it. By the same token, at c one would choose to go right, and therefore we may place an 'x' over the left option. We are now left with a simplified tree at a: going left will have utility +5, while going right will have utility +12. Hence, we should go right twice.

Adding probabilities

The previous example did not involve probabilities. However, introducing them is not much of a problem, as the following example shows.

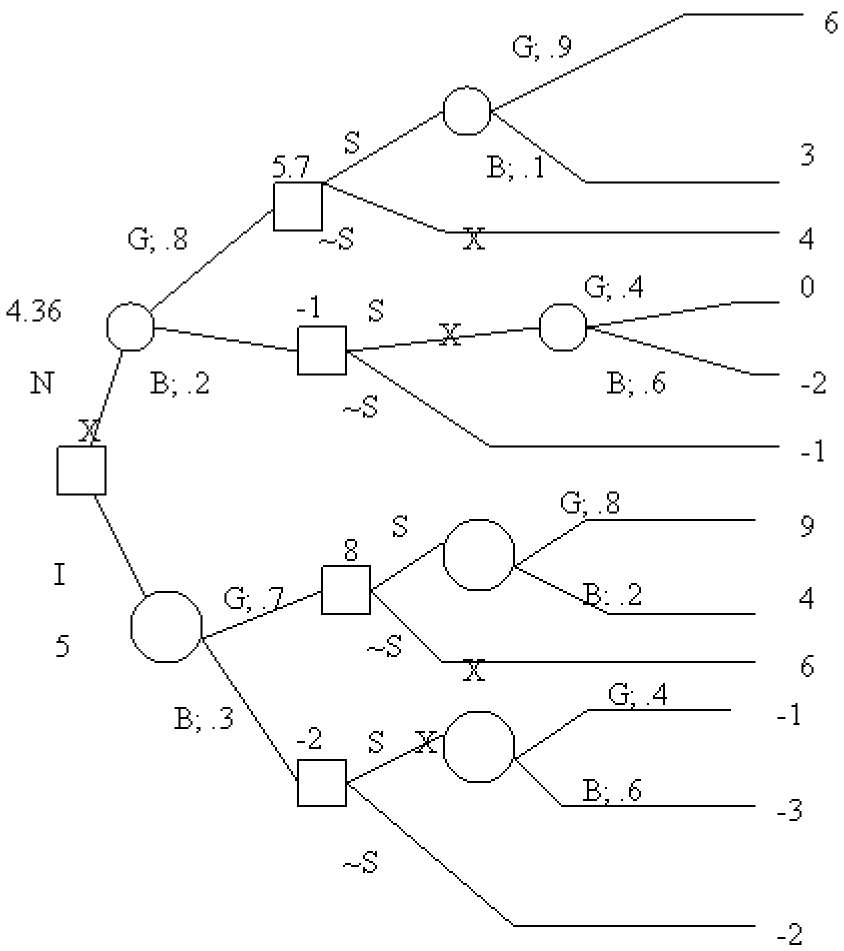
You are about to produce a new garment C and must determine whether to merchandise it only nationally (N) or internationally as well (I).

If you choose N and sales are good (G), then you'll make 4, and if they are bad (B) you'll lose 1. (All payoffs are in millions). You believe that the probability of good national sales is .8 and that of bad national sales is .2. You must also decide whether to produce and sell a matching scarf S. If S's sales are good, you'll make an additional 2, and if they are bad you'll lose an additional 1. You think that if C sells well the probability that S's sales are good is .9 and the probability that S sells badly is .1. By contrast, if C sells badly, the probability that S's sales are good is .4 and the probability that S sells badly is .6.

If you choose I and sales are good you'll make 6, and if they are bad you'll lose 2. You believe that the probability of good sales is .7 and that of bad salses is .3. As in the other case, you must decide about S. If S's sales are good, you'll make an additional 3, and if they are bad you'll lose an additional 2. You think that if C sells well the probability that S's sales are good is .8 and the probability that S sells badly is .2. By contrast, if C sells badly, the probability that S's sales are good is .4 and the probability that S sells badly is .6. Because of the seasonal nature of your business, you need to decide now, a few months in advance, what to do.

The tree below digrams the decision. Obviously, the first decision is whether to choose N or I. Let's follow the subtree tree arising from N, as that stemming from I is analogous. Upon deciding N, there is an act of nature, hence the circle with G (good sales) and B (bad sales). After G, we write .8, which is the probability of good sales if you choose N. Your second decision regards S, which is represented by the square with the two branches S (produce and sell S) and ~S (don't produce S). If you choose S, then the sales can be good or bad (an act of nature) represented by the circle and the relevant probabilities; if you choose ~S, then you are left with the gains or losses from selling C nationally. The payoff are easy to calculate. Conside the uppermost branch, representing your choice of N, good sales of C, your choice of S and good sales of S. Since C sells well, you make 4 and since S sells well, you make 2, so that the total payoff for this branch is 4 + 2 = 6. The second topmost branch has a payoff of 3 because although C sells well (you gain 4), S does not (you lose 1).

The tree is analyzed by backward induction. Starting at the top, the last decision you make is whether to choose S or ~S. The expected payoff of choosing S is 5.7, which is 6 times .9 plus 3 times .1, exactly as if we were considering a bet in which you get 6 with probability .9 and 3 with probability .1. Since 5.7 > 4, you clip ~S, which simply means that the best decision given N and good sales of C is to choose S. Since 5.7 is the best expected payoff available, we write 5.7 next to the relevant square. By contrast, given N and bad sales of C, the best option is ~S, with payoff of -1, as the option S has an expected payoff of -1.2. Hence, S is clipped. We can now calculate the expected payoff of choosing N, namely 5.7 times .8 plus -1 times .2, which is 4.36. We write this figure next to the choice N. The lower half of the tree is analyzed analogously, resulting in an expectd payoff of 5 for I. Since 5 > 4.36, N is clipped and I is the best choice, that with the highest expected payoff.



Revising decision trees in the light of new information

Consider the following scenario. You have to choose which of two widgets, A and B, to market. You also believe that the probability of high demand for A is 80% and the probability of high demand for B is 30%. (High and low demands are the only two alternatives). However, while you get only $3 for every sold A, you get $5 for every sold B. Suppose that if an item is in high demand you shall sell 10,000 of them and with low demand only 4000. Item A cost $1 to produce, while item B costs $2. Moreover, time constraints due to the holiday season compel you to produce 10,000 items of whichever of the two widgets you decide to market. What should you do?

We already know how to construct the decision tree listing the outcomes and the payoffs in thousands. For example, since 10,000 items must be produced no matter what the sales will be, the production cost of A is 10. If A sells well, all 10,000 will be sold, with a gain of 20. Hence, the payoff will be 20-10=10. If demand is low, only 4,000 will be sold, with a gain of 8. Hence, the payoff will be 8-10=-2.

From now on it is just a matter of employing backward induction to determine which widget to market. The expected payoff of marketing A if the panel has a positive reaction is +$7.9; if the panel has a negative reaction the expected payoff drops to +$5.2. With respect to B, if the panel has a positive reaction, the expected payoff is $10.4; if the panel has a negative reaction, it becomes -$.64. The expected value of marketing A is 6.8; that of marketing B is 3.2. Hence, one should market A. Since the expected payoff difference between MA and MB is 3.6, one should pay less than 3.6 for the panel; in other words, given the data of the problem, the extra information is not worth more than 3.6 (siue.edu, 2022).

One can easily see that decision trees could get complex and difficult to construct. For example, getting the information may take time when time is of the essence (perhaps, A and B are holiday season widgets). Moreover, in our examples we came up with the relevant probabilities by fiat. In real situations, of course, coming up with the correct probabilities can be very difficult, especially when appealing to history or statistical surveys does not help. However, such cases are best left to a decision theory course.

The Decision Analysis process transforms a broadly stated decision opportunity into a traceable, defendable and actionable plan. It encompasses one or more discrete analyses at one or more lower (e.g., system element) levels and aggregates them into a higher-level view (e.g., system "scorecard" presentation) relevant to the decision maker and other stakeholders. Decision Analysis can be the central process for formulating, managing and executing an effective and efficient program at any point in the life cycle.

Solution	Evaluation Method	Cost	Performance	Schedule	Total Score
Solution 1	Simulation	4	2	1	29
Solution 2	Discussion	2	3	3	25
Solution 3	Prototype	3	1	4	23
Solution 4	Discussion	1	4	2	23

Sample Evaluation Table

Decision Analysis and associated trade studies should be integrated with, and mutually supportive of, aspects of several SE processes in the early stages of the program, in particular (dau.edu, 2022).

Technical Planning (see Systems SE (SE) Guidebook, Section 4.1.1)
Technical Assessment (see SE Guidebook, Section 4.1.3.)

Stakeholder Requirements Definition (see SE Guidebook, Section 4.2.1)
Requirements Analysis (see SE Guidebook, Section 4.2.2)
Architecture Design (see SE Guidebook, Section 4.2.3)
Results

4.5 Steps in Decision Analysis

Activities and Products
Decision Analysis teams generally include a lead analyst with a suite of reasoning tools, subject matter experts with access to appropriate models and analytical tools and a representative set of end users and other stakeholders. A robust Decision Analysis process acknowledges that the decision maker has full responsibility, authority and accountability for the decision at hand.

Activities

Decision Analysis typically includes the following steps:
Identifying the problem or issue

Reviewing requirements and assumptions to establish the overall decision context

Framing/structuring the decision in terms of supporting program/project objectives

Identifying methods and tools to be used in the analyses (see SE Guidebook, Section 2.2 Tools, Techniques and Lessons Learned)

Developing decision criteria (objectives and measures), criteria weight and associated rationale

Identifying, recording and tracking assumptions

Identifying and defining alternatives to be evaluated (for high-level analyses, these are generally directed, although additional ones may arise during the course of the analysis)

Analyzing and assessing alternatives against criteria
Synthesizing results
Analyzing sensitivities
Developing decision briefing with action/implementation plan(s)
Making appropriate recommendation(s) to decision maker as expected/requested
Products and Tasks

Sound recommendations and action plans are the principal output of a well-framed and well-executed Decision Analysis process. The ability to drill down quickly from overall trade-space visualizations to detailed analyses that support the synthesized views is particularly useful to decision makers in understanding the basis of observations and conclusions.

SELF-ASSESSMENT EXERCISE 2

1. Explain the Decision trees and backward induction
2. Itemize the Steps in Decision Analysis

Product	Tasks
12-1-1: Prepare decision analysis	Identify stakeholder and technical requirements, as well as assumptions to establish the overall decision context. Frame the decision in terms of supporting program / project objectives. Identify methods and tools to be used in the decision analysis. For major defense acquisition programs (MDAPS) and major automated information system (MAIS) programs, describe how the tools support the program's SE approach in the program's systems engineering plan and incorporate in the documentation of the decision analysis recommendation. Develop decision criteria. Identify and define alternatives to be evaluated. Analyze and assess alternatives against decision criteria. Synthesize results. Document analysis and recommend action/implementation to decision maker.

Source: AWQI eWorkbook

4.6 Summary

The unit explained that, A well-executed decision analysis or trade-off analysis helps the Program Manager (PM) and the Systems Engineer understand the impact of various uncertainties, identify one or more course(s) of action that balance competing objectives and objectively communicate the results to decision makers. As such, it provides the basis for selecting a viable and effective alternative from among many under consideration.

Decision Analysis applies to technical decisions at all levels, from evaluating top-level architectural concepts to sizing major system elements to selecting small design details. The breadth and depth of the analysis should be scaled to both the scope of the decision and the needs and expectations of the decision maker(s).

4.7 References/Further Readings/Web Resources

References/Further Reading

siue.edu (2022). Concept and Types of Decision Retrieved from https://www.siue.edu/~evailat/decision.htm dau.edu (2022). Decision Analysis. Retrieved from https://www.dau.edu/tools/se-brainbook/Pages/Management%20Processes/Decision-Analysis.aspx

4.8 Possible Answers to SAEs

Answers to SAEs 1

1 Decision analysis is a formalized approach to making optimal choices under conditions of uncertainty. It allows the user to enter costs, probabilities, and health-related quality of life values among other inputs of interest, and then calculates probabilistically weighted means of these outcome measures. In public health, these outcome measures usually include costs

2. Basic Elements decision theory

There are 4 basic elements in decision theory: *acts, events, outcomes*, and *payoffs*. Acts are the actions being considered by the agent -in the example elow, taking the raincoat or not; events are occurrences taking place outside the control of the agent (rain or lack thereof); outcomes are the result of the occurrence (or lack of it) of acts and events (staying dry or not; being burdened by the raincoat or not); payoffs are the values the decision maker is placing on the occurrences (for example, how much being free of the nuisance of carrying an raincoat is worth to one).

3. Decision Criteria

How one uses a decision tree or a decision matrix depends on the decision criteria one adopts. Consider the following *payoff matrix* where acts are rows, events columns, and the resulting squares contain the payoffs (outcomes are not represented to avoid clutter). So, suppose that we are considering which widget out of 3 to produce and our goal is making money

Answers to SAEs 2

1. Decision trees and backward induction

Typically, more than one decision is involved in decision making, in which case it is best to use a tree instead of a matrix. For example, consider the following situation, in which no probabilities are involved.

You have arrived at a fork in the road on you way home.
If you go left, you'll have to go through a very sunny patch in the mid of a very hot day. However, this will also allow you to admire a beautiful bloom of wildflowers growing by the side of the path. You shall then arrive at another fork. If you take another left, you will have to go by a neighbor's house with a very unpleasant guard dog that will growl at you from the other side of the fence. By contrast, if you go right at the second fork, you'll go by a very noisy and dusty part of the road. Whichever of the two you take, you shall get home quickly

2. Steps in Decision Analysis

Identifying the problem or issue

Reviewing requirements and assumptions to establish the overall decision context

Framing/structuring the decision in terms of supporting program/project objectives

Identifying methods and tools to be used in the analyses (see SE Guidebook, Section 2.2 Tools, Techniques and Lessons Learned)

Developing decision criteria (objectives and measures), criteria weight and associated rationale

Identifying, recording and tracking assumptions

Identifying and defining alternatives to be evaluated (for high-level analyses, these are generally directed, although additional ones may arise during the course of the analysis)

Analyzing and assessing alternatives against criteria
Synthesizing results
Analyzing sensitivities

Developing decision briefing with action/implementation plan(s)
Making appropriate recommendation(s) to decision maker as expected/requested
Products and Tasks
Sound recommendations and action plans are the principal output of a well-framed and well-executed Decision Analysis process

UNIT 5 CRITICAL PATH OF A PROJECT (CPA) AND PROGRAM EVALUATION REVIEW TECHNIQUE (PERT)

Unit Structure

1.1 Introduction

In our last class we discussed the network analysis. This unit will continue with the discussion on the Critical Path of a Project (CPA) and Program Evaluation Review Technique (PERT) and demonstrate the network analysis of a project**.**
CPM History

The critical path method was developed in the late 1950s by Morgan R. Walker and James E. Kelley. The origins of the critical path method are closely related with the Program Evaluation and Review Technique (PERT), a similar method which is commonly used in conjunction with CPM.

1.2 Learning Outcomes

At the end of this unit, should be able to:

- Explain what the Critical Path of a Project is about
- Explain what the Critical Path Method (CPM) is about
- Outline the reason why CPM is important in Project Management?
- State Key Elements of CPM
- Itemize the steps in mapping CPM
- Trance the history of Program Evaluation Review Technique (PERT)

- Explain what is Program Evaluation Review Technique (PERT) is about
- Outline the steps in mapping out a complex project

1.3 Critical Path of a Project (CPA) and Program Evaluation Review Technique (PERT)

1.3.1 What Is the Critical Path of a Project?

In project management, the critical path is the longest sequence of tasks that must be completed to complete a project. The tasks on the critical path are called critical activities because if they're delayed, the whole project completion will be delayed.

1.3.2 What Is the Critical Path Method (CPM)?

The critical path method (CPM) is a technique that's used by project managers to create a project schedule and estimate the total duration of a project.
The CPM method, also known as critical path analysis (CPA), consists in using a network diagram to visually represent the sequences of tasks needed to complete a project. Once these task sequences or paths are defined, their duration is calculated to identify the critical path, which determines the total duration of the project.

1.4 Why Is CPM Important in Project Management?

Important of Critical Path for project managers
Finding the critical path is very important for project managers because it allows them to:

- Accurately estimate the total project duration
- Identify task dependencies, resource constraints and project risks
- Prioritize tasks and create realistic project schedules
- To find the critical path, project managers use the critical path method (CPM) algorithm to define the least amount of time necessary to complete each task with the least amount of slack.
- Once done by hand, nowadays the critical path can be calculated automatically with project scheduling software equipped with Gantt charts, which makes the whole CPM method much easier.
- Project Manager can calculate the critical path for you on our award-winning Gantt charts—learn more.
- Now that we know what's the critical path of a project, we can learn about the critical path method (CPM), an important project management technique that's based on this concept.

Projects are made up of tasks that have to adhere to a schedule in order to meet a timeline. It sounds simple, but without mapping the work, your project scope can quickly get out of hand and you'll find your project off track.

Using the critical path method is important when managing a project because it identifies all the tasks needed to complete the project, then determines the tasks that must be done on time, those that can be delayed if needed and how much float or slack you have.

When done properly, critical path analysis can help you:

Identify task dependencies, resource constraints and project risks
Accurately estimate the duration of each task

Prioritize tasks based on their float or slack time, which helps with project scheduling and resource allocation
Identify critical tasks that have no slack and make sure those are completed on time

Monitor your project progress and measure schedule variance
Use schedule compression techniques like crash duration or fast tracking.

SELF-ASSESSMENT EXERCISE 1

1. What Is the Critical Path of a Project?
2. What Is the Critical Path Method (CPM)?
3. Why Is CPM Important in Project Management?

1.4.1 CPM Key Elements

Before we learn the steps to calculate the critical path, we'll need to understand some key CPM concepts.

Earliest start time (ES): This is simply the earliest time that a task can be started in your project. You cannot determine this without first knowing if there are any task dependencies

Latest start time (LS): This is the very last minute in which you can start a task before it threatens to delay your project schedule

Earliest finish time (EF): The earliest an activity can be completed, based on its duration and its earliest start time

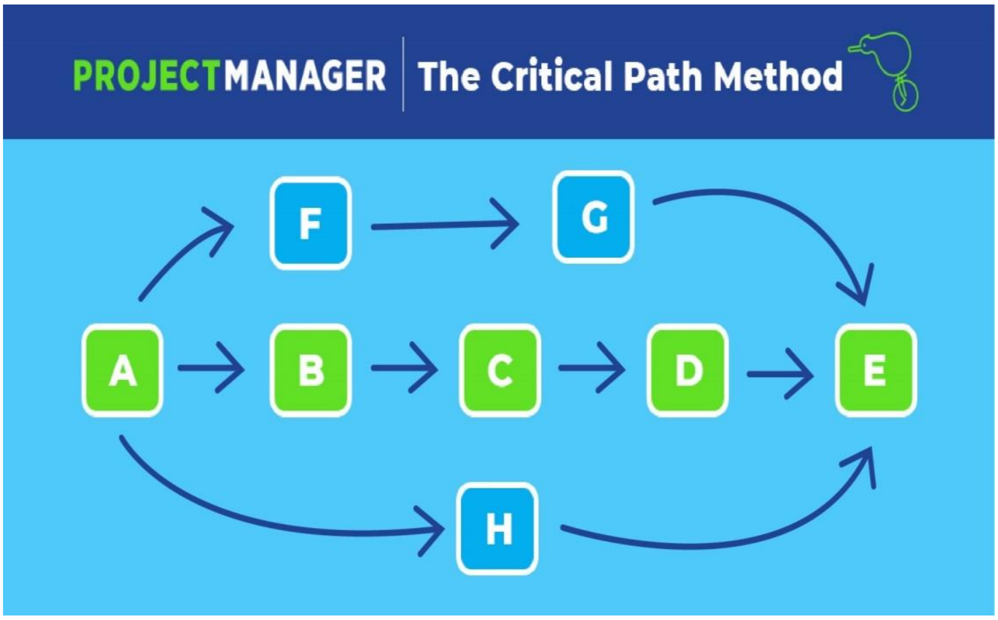
Latest finish time (LF): The latest an activity can be completed, based on its duration and its latest start time

Float: Also known as slack, float is a term that describes how long you can delay a task before it impacts its task sequence and the project schedule. The tasks on the critical path have zero float, because they can't be delayed

Let's take a look at some critical path examples to better understand these critical path analysis elements.

Critical Path Examples

Here's an example of a CPM diagram. Although it's high-level, it can help you visualize the meaning of a critical path for a project schedule. For now, we'll use this critical path diagram to explain the elements that make up the CPM method.

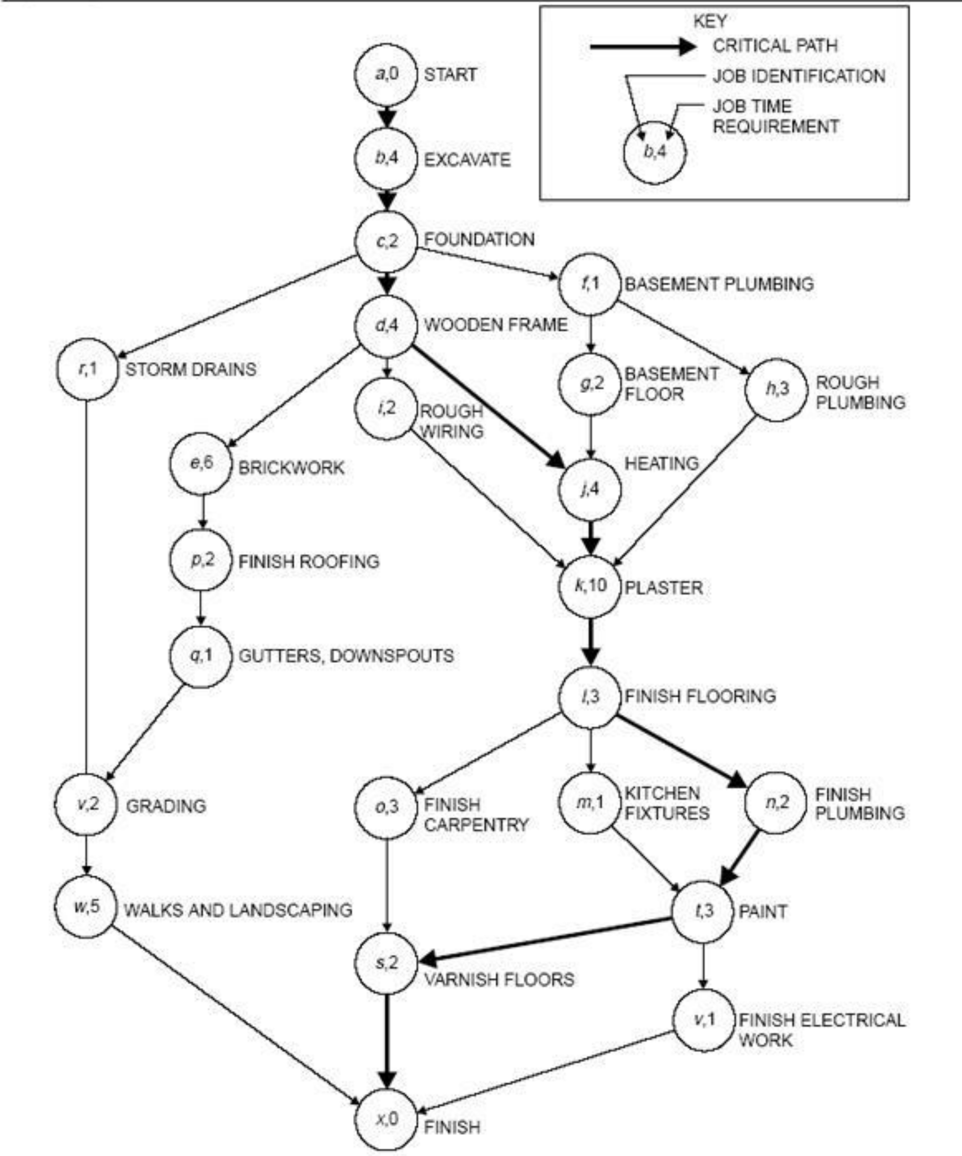


Adopted from https://www.projectmanager.com/guides/critical-path-method.

As you can see in this critical path diagram, project activities are represented by letters and the critical path is highlighted in green. Tasks F, G and H are non-critical activities with float or slack. We can also identify task dependencies between the critical path activities, and also between activities (A, F and G) or (A, H and E), which are parallel tasks.

Here's another critical path example from Harvard Business Review, which shows a critical path schedule for the construction of a house. Each circle in the CPM diagram represents a project activity, as well as it's duration, while the bolded arrows link the critical path activities. As projects become more complex, you'll find more parallel tasks, like in this example.

EXHIBIT II
Project Graph



Source: Harvard Business Review

How to Find the Critical Path of a Project in 8 Steps
Now that you know the key concepts of the critical path method, here's how to calculate

1.4.2 Steps Mapping CPM complex Projects

the critical path in 8 steps.
1.Collect Project Activities
Use a work breakdown structure to collect all the project activities that lead to the final deliverable.

2. Identify Task Dependencies

Figure out which tasks are dependent on other tasks before they can begin. Use your judgement and your team members' feedback. Failing to define task dependencies correctly makes the critical path method useless.

3. Create a Critical Path Diagram
A critical path analysis chart, or network diagram, depicts the order of activities.

4. Estimate Timeline
To use the critical path method, you'll need to estimate the duration of each task. Use data from past projects and other sources of information such as subject matter experts.

5. Use the Critical Path Algorithm

The critical path algorithm has two parts; a forward pass and a backwards pass.

Forward Pass
Use the network diagram and the estimated duration of each activity to determine their Earliest Start (ES) and Earliest Finish (EF). The ES of an activity is equal to the EF of its predecessor, and its EF is determined by the formula EF = ES + t (t is the activity duration). The EF of the last activity identifies the expected time required to complete the entire project (projectmanager.com, 2022).

Backward Pass
Begins by assigning the last activity's Earliest Finish as its Latest Finish. Then the formula to find the LS is LS = LF – t (t is the activity duration). For the previous activities, the LF is the smallest of the start times for the activity that immediately follows.

6. Identify the Float or Slack of Each Activity
Use this formula to determine the float or slack of each task. Float = LS – ES
7. Identify the Critical Path
The activities with 0 float make up the critical path. All of these critical path activities are dependent tasks except for the first task in your CPM schedule. All project tasks with positive slack are parallel tasks to the critical path activities.

8. Revise during Execution
Continue to update the critical path network diagram as you go through the execution phase.

These critical path analysis steps determine what tasks are critical and which can float, meaning they can be delayed without negatively impacting the project schedule. Now you have the information you need to plan the critical path schedule more accurately and have more of a guarantee you'll meet your project deadline.

You also need to consider other changes or constraints that might change the project schedule. The more you can account for these unexpected events or risks, the more accurate your critical path schedule will be. If time is added to the project because of these constraints, that is called a critical path drag, which is how much longer a project will take because of the task and constraint (projectmanager.com, 2022).

1.5 History of Program Evaluation Review Technique (PERT)

PERT was developed by the U.S. Navy in the 1950s to help coordinate the thousands of contractors it had working on myriad projects.
While PERT was originally a manual process, today there are computerized PERT systems that enable project charts to be created quickly.
The only real weakness of the PERT process is that the time required for completion of each task is very subjective and sometimes no better than a wild guess. Frequent progress updates help refine the project timeline once it gets underway.

1.5.1 What is Program Evaluation Review Technique (PERT)?

The Program Evaluation Review Technique, or PERT, is a visual tool used in project planning. Using the technique helps project planners identify start and end dates, as well as interim required tasks and timelines. The information is displayed as a network in chart form.

PERT helps project planners identify:
Start and end dates

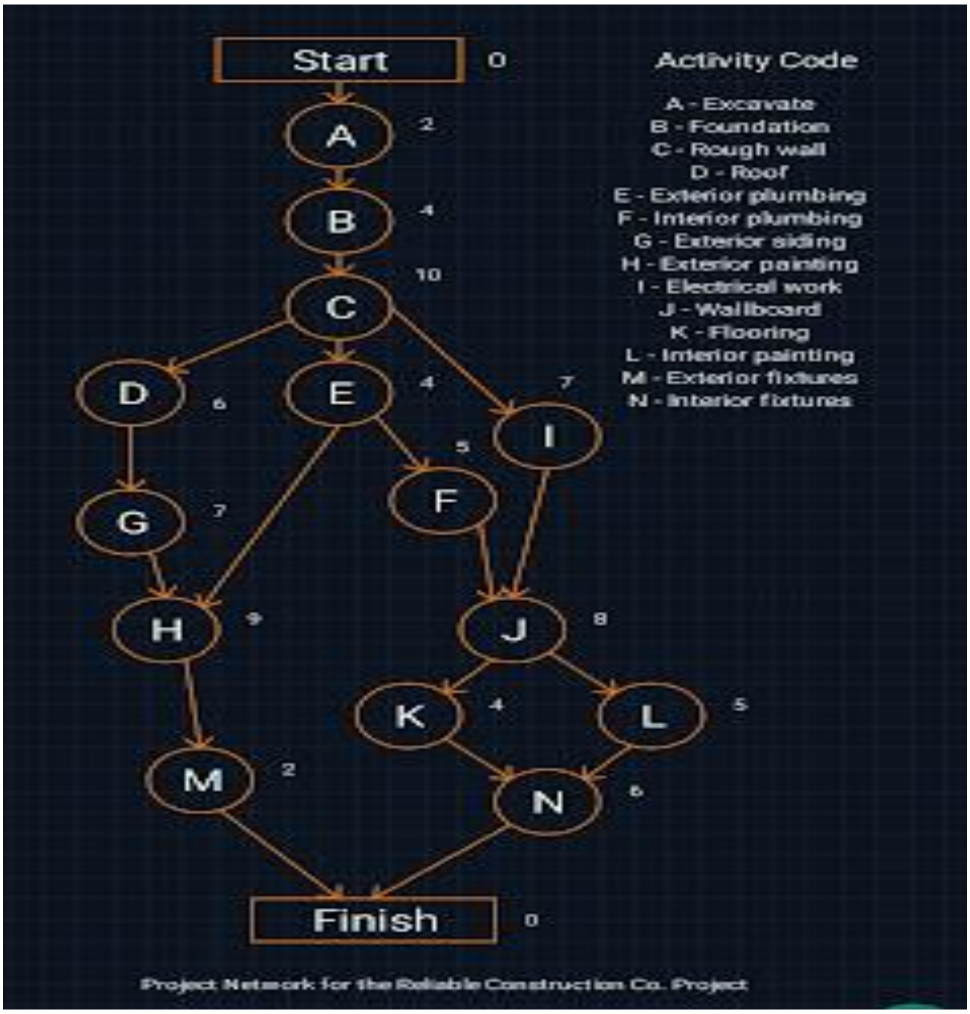
Anticipated total required completion time
All activities, referred to as events on the chart, that impact the completion time
The required sequence of events
The probability of completion by a certain date

1.5.2 Steps in mapping out a complex project

PERT has a set series of steps in mapping out a complex project, which include:
1. List all the tasks and milestones (a.k.a. events) required for completion of the project

- Determine the required sequence of tasks
- Design a chart to visually display all the steps
- Estimate the time required for each task
- Identify the critical path – the longest series of tasks in the project
- Adjust the chart to reflect progress made once the project starts

A PERT chart uses numbered circles or rectangles to represent milestones and straight lines with arrows at the end to represent tasks to be completed. The direction of the arrows, and the numbers, indicate the required sequence. Typically, the numbers increase by 10 at each milestone, so that new tasks can be added along the way without requiring the whole chart to be redrawn and numbered.



Project Network for the Reliable Construction Co. Project

SELF-ASSESSMENT EXERCISE 2

1. Outline the Key Elements of CPM
2. State the Steps Mapping CPM complex Projects
3. What is Program Evaluation Review Technique (PERT)?
4. State the Steps in mapping out a complex project

1.6 Summary

In project management, the critical path is the longest sequence of tasks that must be completed to complete a project. The tasks on the critical path are called critical activities because if they're delayed, the whole project completion will be delayed

The critical path method (CPM) is a technique that's used by project managers to create a project schedule and estimate the total duration of a project.

The CPM method, also known as critical path analysis (CPA), consists in using a network diagram to visually represent the sequences of tasks needed to complete a project

Finding the critical path is very important for project managers because it allows them to:

- Accurately estimate the total project duration
- Identify task dependencies, resource constraints and project risks
- Prioritize tasks and create realistic project schedules
- To find the critical path, project managers use the critical path method (CPM) algorithm to define the least amount of time necessary to complete each task with the least amount of slack.
- Once done by hand, nowadays the critical path can be calculated automatically with project scheduling software equipped with Gantt charts, which makes the whole CPM method much easier.

We learned the steps to calculate the critical path; we also understand some key CPM concepts.

Earliest start time (ES): This is simply the earliest time that a task can be started in your project. You cannot determine this without first knowing if there are any task dependencies

Latest start time (LS): This is the very last minute in which you can start a task before it threatens to delay your project schedule

Earliest finish time (EF): The earliest an activity can be completed, based on its duration and its earliest start time

Latest finish time (LF): The latest an activity can be completed, based on its duration and its latest start time

Float: Also known as slack, float is a term that describes how long you can delay a task before it impacts its task sequence and the project schedule. The tasks on the critical path have zero float, because they can’t be delayed

Let’s take a look at some critical path examples to better understand these critical path analysis elements

The critical paths in 8 steps are:

1. Collect Project Activities
Use a work breakdown structure to collect all the project activities that lead to the final deliverable.

2. Identify Task Dependencies

Figure out which tasks are dependent on other tasks before they can begin. Use your judgement and your team members' feedback. Failing to define task dependencies correctly makes the critical path method useless.

3. Create a Critical Path Diagram
A critical path analysis chart, or network diagram, depicts the order of activities.

4. Estimate Timeline
To use the critical path method, you'll need to estimate the duration of each task. Use data from past projects and other sources of information such as subject matter experts.

5. Use the Critical Path Algorithm
The critical path algorithm has two parts; a forward pass and a backwards pas
The Program Evaluation Review Technique, or PERT, is a visual tool used in project planning. Using the technique helps project planners identify start and end dates, as well as interim required tasks and timelines. The information is displayed as a network in chart form.

PERT helps project planners identify:

Start and end dates
Anticipated total required completion time
All activities, referred to as events on the chart, that impact the completion time
The required sequence of events
The probability of completion by a certain date
PERT has a set series of steps in mapping out a complex project, which include:

- List all the tasks and milestones (a.k.a. events) required for completion of the project
- Determine the required sequence of tasks
- Design a chart to visually display all the steps
- Estimate the time required for each task
- Identify the critical path – the longest series of tasks in the project
- Adjust the chart to reflect progress made once the project starts

1.7 References/Further Readings/Web Resources

projectmanager.com (2022). Critical Path Method. Retrieved from https://www.projectmanager.com/guides/critical-path-method#:~:text=The%20critical%20path%20method%20(CPM,which%20are%20known%20as%20paths.

1.8 Possible Answers to SAEs

Answers to SAEs 1

1. In project management, the critical path is the longest sequence of tasks that must be completed to complete a project. The tasks on the critical path are called critical activities because if they're delayed, the whole project completion will be delayed

2. The critical path method (CPM) is a technique that's used by project managers to create a project schedule and estimate the total duration of a project. The CPM method, also known as critical path analysis (CPA), consists in using a network diagram to visually represent the sequences of tasks needed to complete a project

3. Important of Critical Path for project managers
Finding the critical path is very important for project managers because it allows them to:

1. Accurately estimate the total project duration
2. Identify task dependencies, resource constraints and project risks
3. Prioritize tasks and create realistic project schedules
4. To find the critical path, project managers use the critical path method (CPM) algorithm to define the least amount of time necessary to complete each task with the least amount of slack.
5. Once done by hand, nowadays the critical path can be calculated automatically with project scheduling software equipped with Gantt charts, which makes the whole CPM method much easier.

Answers to SAEs 1

1. Key Elements of CPM
Before we learn the steps to calculate the critical path, we'll need to understand some key CPM concepts.

Earliest start time (ES): This is simply the earliest time that a task can be started in your project. You cannot determine this without first knowing if there are any task dependencies

Latest start time (LS): This is the very last minute in which you can start a task before it threatens to delay your project schedule

Earliest finish time (EF): The earliest an activity can be completed, based on its duration and its earliest start time

Latest finish time (LF): The latest an activity can be completed, based on its duration and its latest start time

Float: Also known as slack, float is a term that describes how long you can delay a task before it impacts its task sequence and the project schedule. The tasks on the critical path have zero float, because they can't be delayed

Let's take a look at some critical path examples to better understand these critical path analysis elements

2. Steps Mapping CPM complex Projects
the critical path in 8 steps.

1.Collect Project Activities
Use a work breakdown structure to collect all the project activities that lead to the final deliverable.

2. Identify Task Dependencies
Figure out which tasks are dependent on other tasks before they can begin. Use your judgement and your team members' feedback. Failing to define task dependencies correctly makes the critical path method useless.

3. Create a Critical Path Diagram
A critical path analysis chart, or network diagram, depicts the order of activities.

4. Estimate Timeline
To use the critical path method, you'll need to estimate the duration of each task. Use data from past projects and other sources of information such as subject matter experts.

5. Use the Critical Path Algorithm
The critical path algorithm has two parts; a forward pass and a backwards pas

3. The Program Evaluation Review Technique, or PERT, is a visual tool used in project planning. Using the technique helps project planners identify start and end dates, as well as interim required tasks and timelines. The information is displayed as a network in chart form.
PERT helps project planners identify:
Start and end dates
Anticipated total required completion time
All activities, referred to as events on the chart, that impact the completion time
The required sequence of events
The probability of completion by a certain date

4. Steps in mapping out a complex project

PERT has a set series of steps in mapping out a complex project, which include:

- List all the tasks and milestones (a.k.a. events) required for completion of the project

- Determine the required sequence of tasks
- Design a chart to visually display all the steps
- Estimate the time required for each task

- Identify the critical path – the longest series of tasks in the project
- Adjust the chart to reflect progress made once the project starts

MODULE 4

Unit 1: Application of Statistical Tools in Public Administration

Unit Structure

1.1 Introduction

This unit will be discussing the descriptive Statistical tools in Public Administration. Statistical tools in Public Administration are fundamental in summarizing and describing the main features of a dataset. They provide a fundamental in summarizing and describing the main features of a dataset. They provide a way to present large amounts of data in a simplified manner, making it easier to understand and communicate key aspects of the data.

1.2 Learning Outcome

At the end of the unit, you should be able to:

- Explain the concept of Descriptive Statistical tools in Public Administration
- Discuss the Origin of Descriptive statistics

- Describe the primary purpose of descriptive
- Examine the descriptive Statistical tools in Public Administration
- Outline the Mean
- Highlight the Median
- Briefly explain the Mode
- Outline the Variance
- Explain the origin of Variance in Public Administration
- Describe the purpose of Variance in Public Administration
- Provide an example of Variance Used in Public Administration
- Explain the application of Variance in Public Administration
- Discuss the Standard Deviation
- Analyze the application of standard deviation

1.3 Descriptive Statistical tools in Public Administration

1.3.1 Origin of Descriptive statistics

Descriptive statistics have their roots in the early development of statistical theory and practice. The term "descriptive statistics" emerged in the late 19th to early 20th century as statistical methods became more formalized. Pioneering statisticians like Karl Pearson and Francis Galton contributed to the development of these methods. Descriptive statistics are part of the broader field of statistics, which originated from efforts to understand and describe variations in data. By summarizing data through various measures, descriptive statistics help in organizing and interpreting the data without drawing inferences or making predictions (Field, 2018).

1.3.2 The primary purpose of descriptive

The primary purpose of descriptive statistics is to summarize and describe the main features of a dataset. This process makes large amounts of data more comprehensible by providing simple numerical and graphical summaries. These summaries help in understanding the basic characteristics of the data, such as central tendencies, dispersion, and distribution shapes. Key purposes include:
Summarization: Condensing data into meaningful summaries.
Description: Providing a clear picture of data characteristics.
Comparison: Facilitating comparisons between different datasets or groups.

1.4 Descriptive Statistical tools in Public Administration

Descriptive statistics employ various tools to summarize and describe data:

1.4.1 Mean: The arithmetic average of a dataset. It is calculated by summing all values and dividing by the number of values. The mean provides a measure of central tendency.

Formula: Mean=Mean=N∑xi - Example: If a survey reports average household income, the mean income represents the typical income level.

1.4.2 Median: The middle value in a dataset when it is ordered from lowest to highest. If the number of values is even, the median is the average of the two middle numbers. The median is useful for understanding the central point of a dataset, especially when the data is skewed.

Example: Median household income can provide a better understanding of typical income in the presence of outliers.

1.4.3 Mode: The value that occurs most frequently in a dataset. It can be particularly useful in categorical data.
Example: In a survey on preferred types of public transport, the mode indicates the most popular choice among respondents.

SELF-ASSESSMENT EXERCISE 1

1.	Discuss the origin of Descriptive statistics
2.	Explain the primary purpose of descriptive

1.5 Variance

1.5.1. Origin of Variance in Public Administration

Variance, a statistical measure of the dispersion or spread of a dataset, finds its roots in classical statistics. It is defined as the average of the squared deviations from the mean of a dataset. In public administration, variance originates from the need to understand and manage the variability in public sector performance, resource allocation, and outcomes. Its application in public administration has evolved as organizations seek to improve efficiency, accountability, and decision-making through quantitative analysis.

1.5.2. Purpose of Variance in Public Administration

In public administration, variance serves several purposes:

Performance Measurement: Variance helps in assessing the performance of public sector programs and projects by comparing actual outcomes with expected outcomes or standards. It provides insights into whether deviations from planned targets are due to external factors or inefficiencies within the system (Rosenbloom & Kravchuk, 2020).

Resource Allocation: By analyzing variance, public administrators can identify areas where resource allocation deviates from the plan. This information is crucial for budget adjustments and optimizing resource distribution (Kettunen, 2019).

Policy Evaluation: Variance analysis aids in evaluating the effectiveness of public policies by highlighting discrepancies between intended and actual results. This allows policymakers to adjust strategies and improve policy outcomes (Berman, 2018).

1.5.3. Examples of Variance Used in Public Administration

Budget Variance: One common example is the analysis of budget variance, where the actual spending is compared to the budgeted amount. Variance in this context helps in identifying areas of overspending or underspending, which can then be addressed to better align with budgetary constraints (Wildavsky, 2019).

Performance Variance: In performance management, variance can be used to compare actual performance metrics against targets. For instance, if a public health program aims to reduce disease incidence by 10%, but the actual reduction is only 5%, variance analysis can help in diagnosing the reasons for this shortfall and guiding corrective actions (Andrews, 2020).

Program Effectiveness: Variance in program outcomes can indicate the effectiveness of different interventions. For example, if two different educational programs are implemented, variance in student performance outcomes can reveal which program is more effective (Rainey, 2018).

1.5.4 Application of Variance in Public Administration

Budgeting and Financial Control: Variance analysis is widely used in budgeting to track discrepancies between budgeted and actual expenditures. This helps in maintaining financial control and ensuring that public funds are used efficiently (Mikesell, 2020).

Performance Management: In performance management systems, variance analysis helps in setting realistic performance targets and assessing whether organizations or departments are meeting their goals.

It also aids in identifying areas needing improvement (Hatry, 2019).

Strategic Planning: Variance analysis supports strategic planning by providing data-driven insights into how well an organization is achieving its strategic objectives. This enables better alignment of strategies with actual performance (Bryson, 2018).

Variance, a statistical tool originating from classical statistics, plays a critical role in public administration by providing insights into performance, resource allocation, and policy effectiveness. It helps public administrators monitor and manage discrepancies between planned and actual outcomes, ultimately guiding more informed decision-making and improving organizational efficiency.

The average of the squared differences from the mean. It provides a measure of data spread but is expressed in squared units of the original data.

Formula: $=N\sum(xi-Mean)^2$

1.6 Standard Deviation

Standard deviation, often referred to in the context of "standard deviance," is a statistical measure that quantifies the amount of variation or dispersion in a set of values. It originated from the field of statistics, where it is used to assess the extent to which individual data points differ from the mean (average) of the dataset. The concept was formalized in the early 20th century and is attributed to the work of statisticians such as Karl Pearson and Ronald A. Fisher.

In public administration, standard deviation is employed to summarize and describe the dispersion of data within a dataset. It provides insight into how spread out the values is around the mean, which helps in understanding the variability of key performance indicators, resource allocation, or policy impacts. This measure is crucial for evaluating the consistency and reliability of administrative processes and outcomes.

Examples of Standard Deviance Used in Public Administration

1. Budget Variance Analysis: Standard deviation is used to analyze budget variances. For instance, if a public agency tracks its monthly expenditures against the budget, the standard deviation of these

expenditures helps in understanding the consistency of spending and identifying anomalies or irregularities.

2. Public Health Data: In public health administration, standard deviation can be used to analyze variations in health outcomes across different regions. For example, the standard deviation of disease incidence rates can indicate the variability in health conditions among different areas, which helps in targeted intervention planning.

3. Employee Performance Metrics: Public sector organizations may use standard deviation to evaluate the variability in employee performance metrics. For instance, in assessing the performance of civil servants, the standard deviation of performance scores can reveal the consistency of performance across different departments or units.

1.6.1 Application of standard deviation

In practical application, standard deviation is calculated using the formula:
σ=N1∑i=1N(xi−μ)^2

Where:

σ = standard deviation
N = number of observations
xi = each individual observation
μ = mean of the observations

Standard Deviation is a measure of the amount of variation or dispersion in a dataset. It indicates how much individual data points deviate from the mean.

Formula: Standard Deviation =N∑(xi−Mean)^2 Example: Standard deviation of public service satisfaction scores helps assess the consistency of responses.

Example 1: Variance in expenditure data helps in understanding the range of spending behavior.

For example 2, if a public agency measures the time taken to process applications, the standard deviation of these processing times will indicate how consistent the processing times are. A low standard deviation would suggest that processing times are fairly uniform, while a high standard deviation would indicate significant variability.

Standard deviation is a vital statistical tool in public administration, offering insights into the variability and consistency of data. By understanding how data points deviate from the mean, public administrators can make informed decisions regarding resource allocation, performance evaluation, and policy implementation. Its application helps in identifying patterns, assessing the effectiveness of programs, and ensuring that public services are delivered consistently.

Descriptive statistics are widely used in public administration to present data in a manageable form and support decision-making processes:

- Summarizing Census Data: Descriptive statistics help summarize demographic information such as age, income, and education levels from census data.
- Example: Reporting the average age and income of a population to understand demographic trends.
- Survey Results: Used to present findings from surveys conducted on public services, policies, or community needs.
- Example: Summarizing survey responses on public health services to identify average satisfaction levels and common issues.
- Policy Analysis: Helps in assessing the effectiveness of policies by summarizing key metrics.
- Example: Describing average improvements in public health indicators before and after a new health policy implementation.
- Resource Allocation: Assists in understanding resource distribution and needs by summarizing data on resource use and demand.
- Example: Summarizing average expenditure on education to guide budget allocation decisions.

SELF-ASSESSMENT EXERCISE 2

1. Define variance
2. Explain the purpose of Variance in Public Administration
3. Discuss the Standard Deviation

1.7 Summary

In this unit, the concept of Descriptive Statistical tools in Public Administration was comprehensively covered. The unit discussed the origin and significance of Descriptive statistics focusing on its primary purpose in the field of Public Administration. Notable Statistical tools like Mean, Median, Mode, Variance, and Standard Deviation were highlighted, with a specific emphasis on their relevance in analyzing data within the realm of Public Administration.

The origin of Variance in Public Administration was explored in depth to provide a clear understanding of its roots and evolution as a statistical measure. The discussion also elaborated on the purpose of Variance in this context, emphasizing its critical role in assessing and interpreting data variability within administrative settings. To enhance comprehension, practical examples showcasing the application of Variance in various scenarios in Public Administration were presented, illustrating how this statistical tool can be utilized effectively to derive insights and make informed decisions.

Furthermore, insights were shared on the significance of Variance and how it can be a valuable asset in optimizing processes and resources in Public Administration. By examining real-world instances where Variance is commonly used, the unit aimed to demonstrate the versatility and applicability of this statistical concept in addressing challenges and achieving objectives within administrative environments.

The unit also shed light on the concept of Standard Deviation and its practical applications in the field of Public Administration. By elucidating how Standard Deviation is calculated and interpreted, the discussion aimed to equip learners with the necessary knowledge and skills to analyze data variability with precision and accuracy. Practical examples and case studies were provided to illustrate the real-world implications of using Standard Deviation as a statistical tool in Public Administration, emphasizing its role in decision-making and performance evaluation.

Overall, this unit provided a comprehensive overview of Descriptive Statistical tools, including Variance and Standard Deviation, within the context of Public Administration. By exploring the origins, purposes, and applications of these statistical measures, learners were encouraged to develop a solid foundation in using data analysis techniques to enhance decision-making and improve outcomes in administrative settings..

1.8 References/Further Reading/Web Resources

Andrews, R. (2020). Public Sector Performance: Management, Measurement, and Evaluation. Routledge.

Berman, E. M. (2018). Public Administration: An Action Orientation. Cengage Learning.

Bryson, J. M. (2018). Strategic Planning for Public and Nonprofit Organizations. Jossey-Bass.

Hatry, H. P. (2019). Performance Measurement: Getting Results. Urban Institute Press.

Kettunen, P. (2019). Public Sector Financial Management: Theory and Practice. Springer.

Mikesell, J. L. (2020). Fiscal Administration: Analysis and Applications for the Public Sector. Wadsworth Cengage Learning.

Rainey, H. G. (2018). Understanding and Managing Public Organizations. Jossey-Bass.

Rosenbloom, D. H., & Kravchuk, R. S. (2020). Public Administration: Understanding Management, Politics, and Law in the Public Sector. McGraw-Hill Education.

Wildavsky, A. (2019). The Politics of the Budgetary Process. Little, Brown and Company.

Pearson, K. (1893). The Grammar of Science. London: Macmillan.

Fisher, R. A. (1921). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

Everitt, B. S., & Skrondal, A. (2010). The Cambridge Dictionary of Statistics (4th ed.). Cambridge University Press.

Smith, J. (2021). "Using Standard Deviation for Performance Evaluation in Public Administration." Journal of Public Administration Research and Theory, 31(2), 456-473.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method (4th ed.). Wiley.

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics (5th ed.). Sage Publications.

Gravetter, F. J., & Wallnau, L. B. (2017). Statistics for the Behavioral Sciences (10th ed.). Cengage Learning.

Kirk, R. E. (2013). Experimental Design: Procedures for the Behavioral Sciences (4th ed.). Sage Publications.

Weisberg, H. F., & Krosnick, J. A. (2012). Statistical Methods for the Social Sciences (4th ed.). Pearson. This book provides comprehensive coverage of statistical methods, including descriptive statistics, used in social sciences and public administration.

Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics (4th ed.). Sage. Field's book is a practical guide on using statistical tools, including descriptive statistics, for analyzing data in various fields, including public administration.

Triola, M. F. (2018). Elementary Statistics (13th ed.). Pearson. This textbook covers fundamental concepts of descriptive statistics and their application in real-world scenarios.

1.9 Possible Answers SAEs

Answer to SAE 1

Origin of Descriptive statistics

Descriptive statistics have their roots in the early development of statistical theory and practice. The term "descriptive statistics" emerged in the late 19th to early 20th century as statistical methods became more formalized. Pioneering statisticians like Karl Pearson and Francis Galton contributed to the development of these methods

The primary purpose of descriptive

The primary purpose of descriptive statistics is to summarize and describe the main features of a dataset. This process makes large amounts of data more comprehensible by providing simple numerical and graphical summaries. These summaries help in understanding the basic characteristics of the data, such as central tendencies, dispersion, and distribution shapes. Key purposes include:
Summarization: Condensing data into meaningful summaries.
Description: Providing a clear picture of data characteristics.
Comparison: Facilitating comparisons between different datasets or groups

Answer to SAE 2

Define variance

Variance, a statistical measure of the dispersion or spread of a dataset, finds its roots in classical statistics. It is defined as the average of the squared deviations from the mean of a dataset. In public administration, variance originates from the need to understand and manage the variability

Purpose of Variance in Public Administration

In public administration, variance serves several purposes:
Performance Measurement: Variance helps in assessing the performance of public sector programs and projects by comparing actual outcomes with expected outcomes or standards. It provides insights into whether deviations from planned targets are due to external factors or inefficiencies within the system (Rosenbloom & Kravchuk, 2020).

Resource Allocation: By analyzing variance, public administrators can identify areas where resource allocation deviates from the plan. This information is crucial for budget adjustments and optimizing resource distribution (Kettunen, 2019).

Policy Evaluation: Variance analysis aids in evaluating the effectiveness of public policies by highlighting discrepancies between intended and actual results. This allows policymakers to adjust strategies and improve policy outcomes (Berman, 2018)

Standard Deviation

Standard deviation, often referred to in the context of "standard deviance," is a statistical measure that quantifies the amount of variation or dispersion in a set of values. It originated from the field of statistics, where it is used to assess the extent to which individual data points differ from the mean (average) of the dataset

UNIT 2: TIME SERIES ANALYSIS IN PUBLIC ADMINISTRATION

Unit Structure

2.1 Introduction
2.2 Learning Outcome
2.3 Time Series Analysis in Public Administration
 2.3.1 Define Time Series Analysis (TSA)
 2.3.2 Purpose of Time Series Analysis in Public Administration
 2.4 Tools of Time Series Analysis in Public Administration
 2.5 Application of Time Series Analysis in Public Administration
2.6 Summary
1.7 References/Further Reading/Web Resources
1.8 Possible Answers SAEs

2.1 Introduction

In our previous unit, we discussed the concept of Descriptive Statistical tools in Public Administration was comprehensively covered. The unit discussed the origin and significance of Descriptive statistics focusing on its primary purpose in the field of Public Administration. Notable Statistical tools like Mean, Median, Mode, Variance, and Standard Deviation were highlighted, with a specific emphasis on their relevance in analyzing data within the realm of Public Administration. In this unit, we will be discussing; Time Series Analysis in Public Administration, Purpose of Time Series Analysis in Public Administration; Tools of Time Series Analysis in Public Administration and the Application of Time Series Analysis in Public Administration

2.2 Learning Outcome

At the end of the unit, you should be able to:

- Discuss the Time Series Analysis in Public Administration
- Define Time Series Analysis (TSA)
- Describe the Purpose of Time Series Analysis in Public Administration
- Highlight the Tools of Time Series Analysis in Public Administration
- Explain the Application of Time Series Analysis in Public Administration

2.3 Time Series Analysis in Public Administration

2.3.1 Define Time Series Analysis (TSA)

Time Series Analysis (TSA) is a statistical technique used to analyze data points collected or recorded at specific time intervals. Time Series Analysis (TSA) stands as a fundamental statistical method essential for examining and interpreting data sequences obtained over distinct time points. It is a powerful tool aiding in the comprehension of trends, patterns, and relationships in time-stamped data, facilitating informed decision-making across various domains. This technique serves as a cornerstone in forecasting future behaviors based on historical information, enabling researchers, businesses, and policymakers to make well-informed projections and strategies. By harnessing the analytical capabilities of TSA, users can delve into the dynamics of time-dependent data, unearthing invaluable insights that drive meaningful conclusions and actionable insights. With its emphasis on exploring the sequential nature of data sets, TSA helps reveal hidden correlations and dependencies that might otherwise go unnoticed, paving the way for a comprehensive understanding of the underlying mechanisms governing the temporal evolution of phenomena. Its application spans a wide range of fields, including finance, economics, weather forecasting, and epidemiology, where the temporal aspect of data plays a crucial role in shaping outcomes and determining future trends. Through the systematic application of statistical techniques within the realm of TSA, practitioners can extract meaningful signals from noisy time series data, separating genuine trends from random fluctuations to extract meaningful patterns and signals.

In essence, Time Series Analysis (TSA) serves as an indispensable analytical approach that empowers individuals and organizations to unlock the latent potential embodied within time-stamped datasets. By leveraging its robust statistical framework, analysts can discern the underlying structure inherent in temporal data, unraveling the complex interplay of variables and trends that drive temporal dynamics. Through the lens of TSA, data points become more than mere observations; they transform into windows through which one can peer into the past, present, and future, gaining critical insights that drive strategic decision-making and informed actions. Embracing the complexities of time series data, TSA equips users with the tools needed to navigate the temporal dimension of information, identifying patterns, anomalies, and trends that hold the key to unlocking value and knowledge. As a versatile analytical tool, TSA transcends disciplinary boundaries, finding applications in diverse fields ranging from marketing to healthcare, where understanding temporal patterns can lead to improvements in forecasting accuracy, resource allocation, and risk mitigation strategies.

Its origins can be traced back to the early 20th century when statisticians like Sir Francis Galton and Karl Pearson laid the groundwork for statistical methods and their applications in various fields. Over time, TSA has evolved to become a vital tool in public administration, aiding in the analysis of data related to government operations, economic indicators, and social phenomena.

2.3.2 Purpose of Time Series Analysis in Public Administration

The primary purpose of TSA is to identify patterns, trends, and relationships in data collected over time. In public administration, this can include analyzing budget expenditures, monitoring population growth, and forecasting future trends. TSA helps policymakers and administrators make informed decisions based on historical data and future projections.

Self-Assessment Exercise 1

Define Time Series Analysis (TSA) Explain the Purpose of Time Series Analysis in Public Administration

2.4 Tools of Time Series Analysis in Public Administration

1. Moving Averages:

Definition: A moving average smooths out fluctuations in time series data to identify trends. It calculates the average of data points within a specific window and updates the average as new data becomes available.
Example: To analyze annual government expenditure, a 3-year moving average might be used to smooth out short-term fluctuations and highlight long-term trends.

2. Exponential Smoothing:
Definition: Exponential smoothing assigns exponentially decreasing weights to past observations, giving more importance to recent data. This method is useful for short-term forecasting.

Example: Forecasting monthly public health expenditure might use exponential smoothing to account for recent trends more heavily.

3. ARIMA (AutoRegressive Integrated Moving Average):
Definition: ARIMA models are used for time series forecasting by combining autoregressive (AR) models, differencing (I) to make the data stationary, and moving average (MA) models.

Example: Forecasting annual GDP growth rates can be effectively modeled using ARIMA to account for past values and errors.

2.5 Application of Time Series Analysis in Public Administration

Time Series Analysis is instrumental in public administration for various applications, such as:

1. Forecasting Budget Expenditures: By analyzing historical budget data, administrators can predict future expenditures and allocate resources more effectively.

2. Population Growth Analysis: TSA can be used to forecast changes in population size, aiding in planning for infrastructure and social services.

3. Economic Indicators: Monitoring and forecasting economic indicators like unemployment rates or inflation can help in formulating economic policies.

SELF-ASSESSMENT EXERCISE 2

1. Highlight the Tools of Time Series Analysis in Public Administration
2. Application of Time Series Analysis in Public Administration

2.6 Summary

Time Series Analysis provides valuable insights into data collected over time, offering tools and techniques to identify trends, make forecasts, and inform decision-making in public administration. By utilizing moving averages, exponential smoothing, and ARIMA models, public administrators can enhance their ability to predict future trends and plan accordingly. These methods help in managing budgets, planning for demographic changes, and analyzing economic conditions, ultimately leading to more informed and effective governance. In conclusion, Time Series Analysis (TSA) emerges as a pivotal methodology that underpins informed decision-making, strategic planning, and predictive modeling by harnessing the nuances of temporal data. Its significance lies in its ability to distill complex temporal datasets into actionable insights, enabling stakeholders to anticipate trends, detect anomalies, and optimize strategies based on historical patterns.

1.7 References/Further Reading/Web Resources

Box, G.E.P., Jenkins, G.M., & Reinsel, G.C. (2015). Time series analysis: forecasting and control. Wiley.

Chatfield, C. (2003). The analysis of time series: an introduction. CRC Press.

Enders, W. (2014). Applied econometric time series. Wiley.

Hyndman, R.J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts. Retrieved from https://otexts.com/fpp3/

1.8 Possible Answers SAEs

Define Time Series Analysis (TSA)

Time Series Analysis (TSA) is a statistical technique used to analyze data points collected or recorded at specific time intervals. Time Series Analysis (TSA) stands as a fundamental statistical method essential for examining and interpreting data sequences obtained over distinct time points.

Purpose of Time Series Analysis in Public Administration

The primary purpose of TSA is to identify patterns, trends, and relationships in data collected over time. In public administration, this can include analyzing budget expenditures, monitoring population growth, and forecasting future trends. TSA helps policymakers and administrators make informed decisions based on historical data and future projections

Possible Answers SAEs 2

Tools of Time Series Analysis in Public Administration

1. Moving Averages:

Definition: A moving average smooths out fluctuations in time series data to identify trends. It calculates the average of data points within a specific window and updates the average as new data becomes available.
Example: To analyze annual government expenditure, a 3-year moving average might be used to smooth out short-term fluctuations and highlight long-term trends.

2. Exponential Smoothing:

Definition: Exponential smoothing assigns exponentially decreasing weights to past observations, giving more importance to recent data. This method is useful for short-term forecasting.

Example: Forecasting monthly public health expenditure might use exponential smoothing to account for recent trends more heavily.

3. ARIMA (AutoRegressive Integrated Moving Average):

Definition: ARIMA models are used for time series forecasting by combining autoregressive (AR) models, differencing (I) to make the data stationary, and moving average (MA) models.

Example: Forecasting annual GDP growth rates can be effectively modeled using ARIMA to account for past values and errors.

Application of Time Series Analysis in Public Administration

Time Series Analysis is instrumental in public administration for various applications, such as:

1. Forecasting Budget Expenditures: By analyzing historical budget data, administrators can predict future expenditures and allocate resources more effectively.

2. Population Growth Analysis: TSA can be used to forecast changes in population size, aiding in planning for infrastructure and social services.

3. Economic Indicators: Monitoring and forecasting economic indicators like unemployment rates or inflation can help in formulating economic policies

UNIT 3: COST-BENEFIT ANALYSIS IN PUBLIC ADMINISTRATION

Unit Structure

3.1 Introduction

In our previous unit, we discussed the Time Series Analysis provides valuable insights into data collected over time, offering tools and techniques to identify trends, make forecasts, and inform decision-making in public administration. In this unit, we will discuss Cost-Benefit Analysis (CBA) which has its roots in economics and public policy, evolving significantly since its inception.

3.2 Learning Outcome

At the end of the unit, you should be able to:

- Explain the origin of Cost-Benefit Analysis in Public Administration
- Define the concept of Cost-benefit analysis
- Outline the Purpose of Cost-Benefit Analysis
- Highlight the tools used in Cost-Benefit Analysis
- Provide the examples of Cost-Benefit Analysis in Public Administration
- Outline the Application of Cost-Benefit Analysis

3.3. Origin of Cost-Benefit Analysis in Public Administration

Cost-Benefit Analysis (CBA) has its roots in economics and public policy, evolving significantly since its inception. The concept emerged from the early 20th century, with foundational contributions from economists like Jules Dupuit and Arthur Cecil Pigou. Dupuit's work in 1844 on the valuation of public goods laid the groundwork for CBA, emphasizing the need to compare the benefits and costs of public projects. Pigou further developed the concept in the early 20th century with his work on welfare economics, which integrated CBA into public policy decision-making (Pigou, 1920).

In public administration, CBA became more formalized during the 1950s and 1960s, particularly with the work of economists such as Samuelson and others who advanced the techniques for evaluating public projects (Samuelson, 1954). The approach was adopted to assess government investments and policies, aiming to provide a structured method for determining whether the benefits of a project or policy outweigh its costs.

3.3.1 Define the concept of Cost-benefit analysis

Cost-benefit analysis, a critical tool in decision-making processes, involves evaluating whether the advantages gained from a particular project or policy exceeds the expenses incurred in implementing it. This analytical approach provides a systematic framework to compare and weigh the potential benefits against the associated costs in order to make informed choices. By examining not only the financial aspects but also considering the broader implications and impacts, decision-makers can assess the overall value and feasibility of a proposed initiative. In essence, cost-benefit analysis serves as a method to quantify and rationalize the trade-offs between different courses of action, thereby assisting in allocating resources efficiently and effectively optimizing outcomes.

SELF-ASSESSMENT EXERCISE 1

1. Discuss the Origin of Cost-Benefit Analysis in Public Administration
2. Define the concept of Cost-benefit analysis

3.4 Purpose of Cost-Benefit Analysis

The primary purpose of CBA in public administration is to evaluate and compare the economic impacts of different projects or policies. It helps decision-makers understand whether the benefits of a proposed intervention outweigh its costs, thereby facilitating more informed and effective public sector decisions. Key features of CBA include:

1. Quantification of Benefits and Costs: CBA seeks to measure all relevant benefits and costs in monetary terms to provide a clear comparison.

2. Time Value of Money: It accounts for the time value of money, recognizing that the value of benefits and costs can change over time.

3. Decision Support: It supports decision-making by providing a systematic framework for assessing the economic efficiency of projects or policies.

3.5. Tools Used in Cost-Benefit Analysis

Several tools and techniques are employed in CBA to analyze and present data:

1. Net Present Value (NPV): NPV calculates the difference between the present value of benefits and the present value of costs over time. It is a critical metric in assessing the profitability of a project. A positive NPV indicates that the benefits exceed the costs, while a negative NPV suggests the opposite (Boardman et al., 2018).

2. Internal Rate of Return (IRR): IRR is the discount rate that makes the NPV of a project zero. It represents the rate of return at which the project breaks even. IRR is useful for comparing the profitability of different projects or investments (Brealey et al., 2020).

3. Benefit-Cost Ratio (BCR): BCR compares the total benefits of a project to its total costs. A BCR greater than 1 indicates that the benefits exceed the costs, making the project viable (Kopp & Smith, 1993).

3.5.1. Examples of Cost-Benefit Analysis in Public Administration

1. Infrastructure Projects: CBA is frequently used to evaluate public infrastructure projects such as highways, bridges, and public transit systems. For example, a CBA might assess the economic benefits of reduced travel time and increased safety against the costs of construction and maintenance (Litman, 2021).

2. Healthcare Programs: Public health initiatives, such as vaccination programs or disease prevention strategies, often utilize CBA to compare

the benefits of improved health outcomes and reduced healthcare costs with the costs of implementing the program (Drummond et al., 2015).

3. Environmental Policies: Environmental regulations and conservation efforts are assessed using CBA to weigh the ecological benefits and long-term sustainability against the economic costs. For instance, the analysis of policies aimed at reducing carbon emissions involves evaluating the benefits of mitigating climate change against the costs of implementing emission controls (Nordhaus, 2007).

3.6. Application of Cost-Benefit Analysis

CBA is used to present data in a manageable and interpretable form. It helps summarize complex datasets, such as census data or survey results, by providing a clear, monetary comparison of benefits and costs. This facilitates better communication of data-driven insights to policymakers and stakeholders, ensuring that decisions are based on a thorough understanding of economic impacts (Mishan & Quah, 2021).

SELF-ASSESSMENT EXERCISE 2

1. Highlight the Tools Used in Cost-Benefit Analysis
2. Outline the Application of Cost-Benefit Analysis

3.7 Summary

The unit provided an in-depth explanation of the concept of Cost-Benefit Analysis, which is a systematic approach used to weigh the benefits of a decision against the costs involved. This analysis is crucial in various fields as it helps in making informed choices by comparing the potential gains and losses associated with a particular course of action.

In understanding the purpose of Cost-Benefit Analysis, it is important to recognize its role in assisting decision-makers in identifying and evaluating the economic impact of their choices. By quantifying both the positive and negative aspects, organizations can prioritize projects or policies that yield the best outcomes while considering the resources required.

Several tools are utilized in Cost-Benefit Analysis to facilitate a comprehensive assessment of the costs and benefits associated with a decision. Some common tools include Net Present Value (NPV), Internal Rate of Return (IRR), and Benefit-Cost Ratio (BCR), which help in calculating financial metrics to determine the feasibility and profitability of a project.

In the context of Public Administration, Cost-Benefit Analysis plays a significant role in evaluating government projects, policies, and programs. For example, when assessing the implementation of a new infrastructure project, such as building highways or improving public transportation systems, policymakers use Cost-Benefit Analysis to determine whether the potential social and economic benefits outweigh the initial investment and ongoing costs.

The application of Cost-Benefit Analysis extends beyond financial considerations, as it also encompasses social, environmental, and ethical factors. By conducting a thorough analysis that incorporates these diverse perspectives, decision-makers can make more informed choices that align with broader societal goals and values.

In summary, Cost-Benefit Analysis serves as a valuable decision-making tool that enables organizations and governments to assess the full picture of benefits and costs associated with various options. By following a structured methodology and utilizing appropriate tools, stakeholders can weigh the pros and cons effectively, ultimately leading to more efficient resource allocation and strategic decision-making processes.

1.8 References/Further Reading/Web Resources

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2018). Cost-Benefit Analysis: Concepts and Practice. Cambridge University Press.

Brealey, R. A., Myers, S. C., & Allen, F. (2020). Principles of Corporate Finance. McGraw-Hill Education.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press.

Kopp, R. J., & Smith, V. K. (1993). Valuing the Quality of Life: The Contribution of Economics to the Study of Health and Safety. University of Chicago Press.

Litman, T. (2021). Evaluating Transportation Economic Development Benefits. Victoria Transport Policy Institute.

Mishan, E. J., & Quah, E. (2021). Cost-Benefit Analysis. Routledge.

Nordhaus, W. D. (2007). The Challenge of Global Warming: Economic Models and Environmental Policy. Yale University.

Pigou, A. C. (1920). The Economics of Welfare. Macmillan.

Samuelson, P. A. (1954). The Pure Theory of Public Expenditure. Review of Economics and Statistics, 36(4), 387-389.

1.9 Possible Answers SAEs

Answer to SAEs1

Q1. Origin of Cost-Benefit Analysis in Public Administration
Cost-Benefit Analysis (CBA) has its roots in economics and public policy, evolving significantly since its inception. The concept emerged from the early 20th century, with foundational contributions from economists like Jules Dupuit and Arthur Cecil Pigou. Dupuit's work in 1844 on the valuation of public goods laid the groundwork for CBA, emphasizing the need to compare the benefits and costs of public projects.

Q2 Concept of Cost-benefit analysis
Cost-benefit analysis, a critical tool in decision-making processes, involves evaluating whether the advantages gained from a particular project or policy exceeds the expenses incurred in implementing it. This analytical approach provides a systematic framework to compare and weigh the potential benefits against the associated costs in order to make informed choices

Answer to SAEs 2

Q1 Purpose of Cost-Benefit Analysis
The primary purpose of CBA in public administration is to evaluate and compare the economic impacts of different projects or policies. It helps decision-makers understand whether the benefits of a proposed intervention outweigh its costs, thereby facilitating more informed and effective public sector decisions. Key features of CBA include:

1. Quantification of Benefits and Costs: CBA seeks to measure all relevant benefits and costs in monetary terms to provide a clear comparison.

2. Time Value of Money: It accounts for the time value of money, recognizing that the value of benefits and costs can change over time

Q2 Tools Used in Cost-Benefit Analysis
Several tools and techniques are employed in CBA to analyze and present data:

1. Net Present Value (NPV): NPV calculates the difference between the present value of benefits and the present value of costs over time. It is a critical metric in assessing the profitability of a project. A positive NPV

indicates that the benefits exceed the costs, while a negative NPV suggests the opposite (Boardman et al., 2018).

2. Internal Rate of Return (IRR): IRR is the discount rate that makes the NPV of a project zero. It represents the rate of return at which the project breaks even. IRR is useful for comparing the profitability of different projects or investments (Brealey et al., 2020).

3. Benefit-Cost Ratio (BCR): BCR compares the total benefits of a project to its total costs. A BCR greater than 1 indicates that the benefits exceed the costs, making the project viable (Kopp & Smith, 1993).

UNIT 4 DECISION TREES ANALYSIS IN PUBLIC ADMINISTRATION

Unit Structure

4.1 Introduction

In the previous unit, we provided an in-depth explanation of the concept of Cost-Benefit Analysis, which is a systematic approach used to weigh the benefits of a decision against the costs involved. The unit outline several tools are utilized in Cost-Benefit Analysis to facilitate a comprehensive assessment of the costs and benefits associated with a decision. Some common tools include Net Present Value (NPV), Internal Rate of Return (IRR), and Benefit-Cost Ratio (BCR), which help in calculating financial metrics to determine the feasibility and profitability of a project. In this unit, we will discuss the concept of decision trees analysis; purpose decision trees analysis and tools of decision trees analysis in Public Administration.

4.2 Learning Outcome

At the end of the unit, you should be able to:

- Discuss the origin of Decision Trees Analysis
- Explain the Purpose Decision Trees Analysis in Public Administration
- Describe the Tools of Decision Trees Analysis in Public Administration
- Provide an example of Decision Trees Analysis in Public Administration

4.3 Decision Trees Analysis in Public Administration

4.3.1 Origin of Decision Trees Analysis

Decision Trees analysis originated in the field of machine learning and statistics, with roots tracing back to early developments in decision theory and algorithms in the 1960s. The concept of decision trees was formalized by researchers like Ross Quinlan in the 1980s with the development of algorithms such as ID3 (Iterative Dichotomiser 3) and C4.5. These methods provided a structured way to visualize and make decisions based on data. Decision Trees have since become a significant tool in various fields, including public administration, due to their simplicity and interpretability in complex decision-making scenarios (Quinlan, 1986; Breiman et al., 1986).

4.5 Purpose Decision Trees Analysis in Public Administration

The primary purpose of Decision Trees in public administration is to assist in decision-making by providing a clear, visual representation of different decision paths and their potential outcomes. They help summarize and describe complex datasets by breaking them down into manageable segments and illustrating the impact of various factors on decision outcomes. This enables policymakers to evaluate different scenarios and make informed choices based on the projected benefits and risks associated with each option (Liaw & Wiener, 2002).

SELF-ASSESSMENT EXERCISE 1

Q1. Discuss the origin of Decision Trees Analysis

Q2. Explain the Purpose Decision Trees Analysis in Public Administration

4.6 Tools of Decision Trees Analysis in Public Administration

1. Classification Trees: These are used for categorizing data into distinct classes or categories. For instance, in public administration, classification trees can be used to predict which individuals or groups are likely to fall into specific categories, such as eligibility for social services or potential risks in policy implementation (Breiman et al., 1986).

2. Regression Trees: Unlike classification trees, regression trees are used for predicting continuous outcomes. They can be employed to estimate the impact of policy changes on variables such as public health indicators or economic performance (Hastie, Tibshirani, & Friedman, 2009).

3. Others: Other types of decision trees include CART (Classification and Regression Trees), which can handle both classification and regression tasks, and CHAID (Chi-squared Automatic Interaction Detector), which is particularly useful for identifying interaction effects between variables (Kass, 1980).

4.7 Examples of Decision Trees Analysis in Public Administration

1. Policy Impact Assessment: Decision Trees are used to evaluate the potential impacts of new policies or programs. For example, a government agency might use a decision tree to assess the effects of a proposed health intervention on different population groups, considering factors like cost, expected health improvements, and resource allocation (Liaw & Wiener, 2002).

2. Resource Allocation: Decision Trees can assist in determining the optimal allocation of resources. For instance, they can help in deciding how to distribute funding among various public projects based on predicted outcomes and benefits (Breiman et al., 1986).

3. Risk Management: In public administration, Decision Trees are employed to identify and manage risks associated with various projects or policies. They provide a systematic approach to evaluating potential risks and their impacts, helping administrators make decisions that minimize adverse outcomes (Hastie et al., 2009).

SELF-ASSESSMENT EXERCISE 2

Describe the Tools of Decision Trees Analysis in Public Administration Provide an examples of Decision Trees Analysis in Public Administration

1.8 Summary

Decision Trees analysis is a powerful tool in public administration for enhancing decision-making processes. Originating from early developments in decision theory and machine learning, Decision Trees offer a clear and interpretable method for summarizing and analyzing complex datasets. They encompass various types, including classification and regression trees, each serving different purposes. In practice, Decision Trees are used for policy impact assessment, resource allocation, and risk management, providing valuable insights and supporting informed decision-making.

1.9 References/Further Reading/Web Resources

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1986). Classification and Regression Trees. Belmont, CA: Wadsworth.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. Applied Statistics, 29(2), 119-127.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R News, 2(3), 18-22.

Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81-106.

1.10 Possible Answers SAEs

Answer to SAE 1

Q1 Origin of Decision Trees Analysis

Decision Trees analysis originated in the field of machine learning and statistics, with roots tracing back to early developments in decision theory and algorithms in the 1960s. The concept of decision trees was formalized by researchers like Ross Quinlan in the 1980s with the development of algorithms such as ID3 (Iterative Dichotomiser 3) and C4.5. These methods provided a structured way to visualize and make decisions based on data. Decision Trees have since become a significant tool in various fields, including public administration, due to their simplicity and interpretability in complex decision-making scenarios (Quinlan, 1986; Breiman et al., 1986).

Q2. Purpose Decision Trees Analysis in Public Administration

The primary purpose of Decision Trees in public administration is to assist in decision-making by providing a clear, visual representation of different decision paths and their potential outcomes. They help summarize and describe complex datasets by breaking them down into manageable segments and illustrating the impact of various factors on decision outcomes

Answer to SAE 2

Q1 Tools of Decision Trees Analysis in Public Administration

1. Classification Trees: These are used for categorizing data into distinct classes or categories. For instance, in public administration, classification trees can be used to predict which individuals or groups are likely to fall into specific categories, such as eligibility for social services or potential risks in policy implementation (Breiman et al., 1986).

2. Regression Trees: Unlike classification trees, regression trees are used for predicting continuous outcomes. They can be employed to estimate the impact of policy changes on variables such as public health indicators or economic performance (Hastie, Tibshirani, & Friedman, 2009).

3. Others: Other types of decision trees include CART (Classification and Regression Trees), which can handle both classification and regression tasks, and CHAID (Chi-squared Automatic Interaction Detector), which is particularly useful for identifying interaction effects between variables (Kass, 1980)

Q2 Examples of Decision Trees Analysis in Public Administration

1. Policy Impact Assessment: Decision Trees are used to evaluate the potential impacts of new policies or programs. For example, a government agency might use a decision tree to assess the effects of a proposed health intervention on different population groups, considering factors like cost, expected health improvements, and resource allocation (Liaw & Wiener, 2002).

2. Resource Allocation: Decision Trees can assist in determining the optimal allocation of resources. For instance, they can help in deciding how to distribute funding among various public projects based on predicted outcomes and benefits (Breiman et al., 1986).

Unit 5 Regression Statistical Analysis in Public Administration

Unit Structure

5.1 Introduction

In the previous unit, we examined the concept of Decision Trees analysis which we explained that it is the powerful tool in public administration for enhancing decision-making processes. Originating from early developments in decision theory and machine learning, Decision Trees offer a clear and interpretable method for summarizing and analyzing complex datasets. In this unit, we will be discussing, Regression Statistical Analysis in Public Administration, purpose of regression statistical analysis in Public Administration, types of regression analysis.

5.2 Learning Outcome

At the end of the unit, you should be able to:

- Discuss the origin of Regression Statistical Analysis in Public Administration
- Outline the Purpose of Regression Statistical Analysis in Public Administration
- Highlight the Types of Regression Analysis
- Provide an application of Regression Statistical Analysis in Public Administration

5.3 Regression Statistical Analysis in Public Administration

5.3.1 Origin of Regression Statistical Analysis in Public Administration

Regression analysis has its roots in early statistical methods developed by Sir Francis Galton and Karl Pearson in the late 19th century. It was initially used to study the relationships between physical traits and their inheritance. Over time, regression analysis evolved into a key tool for understanding relationships between variables across various fields, including public administration.

5.3.2 Purpose of Regression Statistical Analysis in Public Administration

In public administration, regression analysis is employed to examine relationships between variables and to predict outcomes. The primary purposes include:

1. Understanding Relationships: Analyzing how different factors influence one another.
2. Predicting Outcomes: Forecasting future trends or impacts based on current data.
3. Evaluating Policies: Assessing the effectiveness of public policies and programs.

SELF-ASSESSMENT EXERCISE 1

Q1. Discuss the origin of Regression Statistical Analysis in Public Administration Q2. Outline the Purpose of Regression Statistical Analysis in Public Administration

5.4 Types of Regression Analysis

1. Simple Linear Regression:

Description: Analyzes the relationship between two variables: one independent (predictor) and one dependent (outcome). It aims to fit a straight line (linear relationship) through the data points.

Example: Investigating the relationship between public spending on education and literacy rates.

2. Multiple Regressions:
Description: Extends simple linear regression by including two or more independent variables. This helps in understanding how multiple factors collectively influence the dependent variable.
Example: Evaluating how factors like budget allocation, training programs, and community involvement affect public health outcomes.

3. Logistic Regression:
Description: Used when the dependent variable is categorical (e.g., yes/no, success/failure). It estimates the probability of a certain event occurring based on the independent variables.

Example: Assessing the likelihood of a community's support for a new public policy based on demographic and socio-economic variables.

5.5 Application of Regression Statistical Analysis in Public Administration

1. Policy Evaluation:
Regression analysis helps in determining the effectiveness of policies. For instance, analyzing how different factors such as economic conditions and public spending affect unemployment rates.

2. Forecasting:
Public administrators use regression models to predict future trends. For example, forecasting crime rates based on historical data and socio-economic indicators.

3. Understanding Public Outcomes:
Helps in comprehending how various factors impact public outcomes. For example, studying how environmental policies affect public health.

Self-Assessment Exercise 2

Q1. Highlight the Types of Regression Analysis Q2. Provide an application of Regression Statistical Analysis in Public Administration

5.6 Summary

Regression statistical analysis is a fundamental tool in public administration for examining relationships between variables and predicting outcomes. Its application in policy evaluation, forecasting, and understanding public outcomes provides valuable insights for decision-making and planning. By using simple linear regression, multiple regression, and logistic regression, public administrators can analyze various factors, assess the impact of policies, and make data-driven decisions.

1.7 References/Further Reading/Web Resources

Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics (5th ed.). McGraw-Hill Education. This book provides a comprehensive introduction to regression analysis and its applications.

Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics (4th ed.). Sage Publications. Field's book offers practical insights into statistical analysis, including regression techniques.

Wooldridge, J. M. (2016). Introductory Econometrics: A Modern Approach (6th ed.). Cengage Learning. Wooldridge's text is valuable for understanding the application of econometric models, including regression analysis in various fields.

1.10 Possible Answers SAEs

Answer to SAEs 1

Q1. Origin of Regression Statistical Analysis in Public Administration

Regression analysis has its roots in early statistical methods developed by Sir Francis Galton and Karl Pearson in the late 19th century. It was initially used to study the relationships between physical traits and their inheritance. Over time, regression analysis evolved into a key tool for understanding relationships between variables across various fields, including public administration.

Q2. Purpose of Regression Statistical Analysis in Public Administration

In public administration, regression analysis is employed to examine relationships between variables and to predict outcomes. The primary purposes include:

1. Understanding Relationships: Analyzing how different factors influence one another.
2. Predicting Outcomes: Forecasting future trends or impacts based on current data.
3. Evaluating Policies: Assessing the effectiveness of public policies and programs.

Answer to SAEs 2

Q1. Types of Regression Analysis

1. Simple Linear Regression:
Description: Analyzes the relationship between two variables: one independent (predictor) and one dependent (outcome). It aims to fit a straight line (linear relationship) through the data points.
Example: Investigating the relationship between public spending on education and literacy rates.

2. Multiple Regressions:
Description: Extends simple linear regression by including two or more independent variables. This helps in understanding how multiple factors collectively influence the dependent variable.

Example: Evaluating how factors like budget allocation, training programs, and community involvement affect public health outcomes.

3. Logistic Regression:
Description: Used when the dependent variable is categorical (e.g., yes/no, success/failure). It estimates the probability of a certain event occurring based on the independent variables.

Example: Assessing the likelihood of a community's support for a new public policy based on demographic and socio-economic variables.

Q2. Application of Regression Statistical Analysis in Public Administration

1. Policy Evaluation:
Regression analysis helps in determining the effectiveness of policies. For instance, analyzing how different factors such as economic conditions and public spending affect unemployment rates.

2. Forecasting:
Public administrators use regression models to predict future trends. For example, forecasting crime rates based on historical data and socio-economic indicators.

3. Understanding Public Outcomes:
Helps in comprehending how various factors impact public outcomes. For example, studying how environmental policies affect public health.

MODULE 5

UNIT 1: STATISTICAL TOOLS II

Unit Structure

1.1 Introduction

In this unit, we develop the basic principles of probabilistic analysis, with special emphasis on the applications of set operations and events.

1.2 Learning Outcomes

At the end of this unit, you should be able to:

- Define concept of sets as they are used in mathematical operations
- Explain the theory of sets and how it is used in probability analysis.
- Explain set enumerations and how they can be applied in solving business-related problems

1.3 Sets Theory

The theory of sets serves as a preliminary concept necessary for the understanding of the theory of probabilities. A mathematical set is a collection of distinct objects, often referred to as elements or members.

Examples

(a) The employees of a company working in the Public Relations Department could be represented as:
PR = {Joseph, Adamu, Adebola, Nkom, Margerate}

(b) The location of shops for a big automobile parts dealer could be represented as:
S = {Abuja, Enugu, Lagos, Aba, Onitsha, Kano, Ikot Ekpene}
1.3.1 Subsets

A subset of a set, say A, is a set which contains some of the elements of set A. For instance:
If set A = {a, e, i, o, u}, then:
X = {a, e, i} is a subset of A
Y = {e, i} is a subset of A
Z = {i, o, u} is a subset of A
1.3.2 The Number of a Set

The *number* of a set A, written as n[A], is defined as the number of elements in set A. For example:

If A = {a, e, i, o, u}, then n[A] = 5 (that is, 5 elements in set A).
1.3.3 Set Equality

Two sets are said to be equal only when they have identical elements.
For example:

If A = {1, 2, 3} and B = {1, 2, 3}, then Set A = Set B.
1.4 Universal Set

A universal set, denoted by U is a set containing different subsets of its elements. For example, a combination of different behaviours in a given population can be considered as universal, while a selected sample of such behaviours are referred to as the subsets. A set of all English alphabets make up the universal set, while a set containing the vowels would be referred to as the subset.

Symbol of Universal Set

The universal set is usually represented by the symbol E or U. It consists of all the elements of its subsets, including its own elements https://www.cuemath.com/algebra/universal-set/.

Example of Universal Set

Let's consider an example with three sets, A, B, and C. Here, A = {2, 4, 6}, B = {1, 3, 7, 9, 11}, and C = {4, 8, 11}. We need to find the universal set for all three sets A, B, and C. All the elements of the given sets are contained in the universal set. Thus, the universal set U of A, B, and C is given by U = {1, 2, 3, 4, 6, 7, 8, 9, 11}
We can see that all the elements of the three sets are present in the universal set without any repetition. Thus, we can say that all the elements in the universal set are unique. The sets A, B, and C are contained in the universal set, then these sets are also called subsets of the Universal set.

$A \subset U$ (A is the subset of U)
$B \subset U$ (B is the subset of U)
$C \subset U$ (C is the subset of U)

1.5 Complement of Universal Set

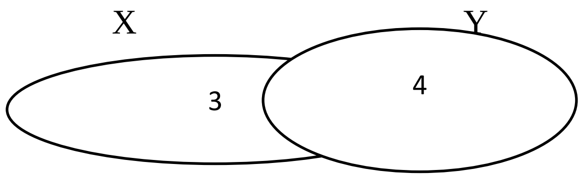
For a subset A of the universal set (U), its complement is represented as A' which includes the elements of the universal set but not the elements of set A. The Universal set consists of a set of all elements of all its related subsets, whereas the empty set contains no elements of the subsets. Thus, the complement of the universal set is an empty set, denoted by '{}' or the symbol 'Φ'.

1.5.1 Complement of a Set

The complement of a set A (denoted by A') contained in a given universal set, U, is the set of elements in the universal set that are not contained in set A. For example: If set A represents the set of all skilled workers in a given universal set, U, then the complement of set A, A', is the set of unskilled workers who are members of the universal set.

1.6 Venn Diagrams

Venn diagrams are simple pictorial representations of a set. They are useful for demonstrating relationships between sets.
This can be represented by a Venn diagram as follows:



The elements {3, 4} are contained in the circle common to sets X and Y.

SELF-ASSESSMENT EXERCISE 1

A mathematical set is a collection of distinct objects, often referred to as----------- State the two basic operations on sets

1.7 Set Enumeration

Set enumeration considers sets in terms of number of elements contained within the various areas defined by their union or intersection. Identifying the number of elements in these areas is known as set enumeration. As an illustration, consider the following enumeration problem.

Suppose an accounting firm currently employs 16 staff members. Given that three staff members have no formal qualifications, and of the seven staff members, who are graduates, 5 are also qualified as chartered accountants, it is possible to evaluate: (a) the number of staff who arenon-graduates, chartered accountants and (b) the number of graduates who are not qualified chartered accountants.

The two values can be calculated as follows:

Since three staff members have no formal qualifications, there should be 16– 3 = 13 staff members with at least one of the two qualifications that is, a graduate or a chartered accountant.

But there are 7 staff members who are graduates, which implies that 16 –7–3= 6 must be non-graduate, qualified chartered accountants, which gives the answer to possibility (a) above.

In addition, since 5 staff members are qualified chartered accountants and graduates, there would be 7 – 5 = 2 staff who are graduates only. This gives answer to possibility (b) above.
In an extended problem, the above approach is not structured enough for the solution. In the following discussions, we present a more structured procedure which solves the above problems and forms a basis for more logical approach for solving enumerations problems in general.

1.7.1 Circle Venn Diagram

Circle Venn Diagram Examples:

For the purposes of an administrative research, a survey of 1000 women is conducted in a town. The results show that 52 % liked watching comedies, 45% liked watching fantasy movies and 60% liked watching romantic movies. In addition, 25% liked watching comedy and fantasy both, 28% liked watching romantic and fantasy both and 30% liked watching comedy and romantic movies both. 6% liked watching none of these movie genres (https://www.intellspot.com/venn-diagram-examples/).

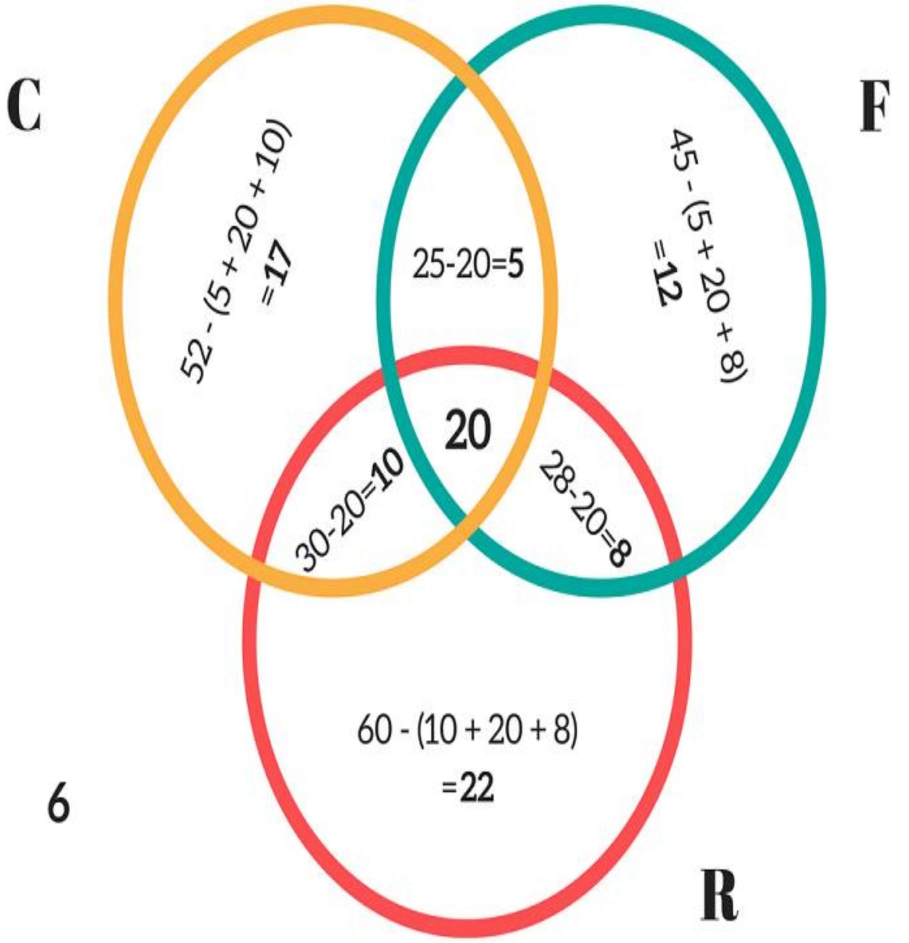
Here are our questions we should find the answer:
How many women like watching all the three movie genres?
Find the number of women who like watching only one of the three genres.

Find the number of women who like watching at least two of the given genres.

Let's represent the data above in a more digestible way using the Venn diagram formula elements:

n(C) = percentage of women who like watching comedy = 52%
n(F) = percentage of women who like watching fantasy = 45%
n(R)= percentage of women who like watching romantic movies= 60%
$n(C \cap F) = 25\%$; $n(F \cap R) = 28\%$; $n(C \cap R) = 30\%$
Since 6% like watching none of the given genres so, $n(C \cup F \cup R) = 94\%$.
Now, we are going to apply the Venn diagram formula for 3 circles.
$94\% = 52\% + 45\% + 60\% – 25\% – 28\% – 30\% + n(C \cap F \cap R)$
Solving this simple math equation, lead us to:
$n(C \cap F \cap R) = 20\%$

It is a great time to make our Venn diagram related to the above situation (problem):



(https://www.intellspot.com/venn-diagram-examples/).

See, the Venn diagram makes our situation much more clear! From the Venn diagram example, we can answer our questions with ease.

The number of women who like watching all the three genres = 20% of 1000 = 200.

Number of women who like watching only one of the three genres = (17% + 12% + 22%) of 1000 = 510

The number of women who like watching at least two of the given genres = (number of women who like watching only two of the genres) +(number of women who like watching all the three genres) = (10 + 5 + 8 + 20)% i.e. 43% of 1000 = 430. As we mentioned above 2 and 3 circle diagrams are much more common for problem-solving in many areas such as business, statistics, data science and etc. However, 4 circles Venn diagram also have its place.

1.7.2 Summary of the General Enumeration Problems

Note in particular that there are:

(a) 4 distinct areas for two attribute sets, and
(b) 8 distinct areas for three attribute sets.
The general procedure for solving enumeration problems follows the

following steps:

Step 1: Identify the attribute sets
Step 2: Draw an outline Venn diagram
Step 3: Use the information given to fill in as much of the diagramas possible

Step 4: Evaluate the number of elements in unknown areas.
Consider the following example:

A survey was carried out by a researcher, one of the aims being to discover the extent to which computers are being used by firms in a given area. 32 firms had both stock control and payroll computerized, 65 firms had just one of these two functions computerised, and 90 firms had a computerized payroll. If 22 firms had neither of these functions computerized, how many firms were included in the survey?

Step 1: The two attributes involved are computerised payroll, with set (say P), and computerised stock control, with set (say S).
Step 2: Using standard notations, Let p = number of firms with a computerised payroll only; s = number of firms with a computerised stock control only; ps = number of firms with both payroll and stock control computerised; and
x = number of firms with neither functions computerised, construct a Venn diagram describing the situation.
Step 3: The following equations can be set up from the given information.
ps = 32
p + s = 65
ps + p = 90
x = 22

Substituting for ps = 32 in equation 3, we get p = 58
Substituting for p = 58 in equation, we get s = 7.

It follows that the number of firms included in the survey equals:

$p + s + ps + x = 58 + 7 + 32 + 22 = 119$

SELF-ASSESSMENT EXERCISE 2

Set enumeration considers sets in terms of number of elements contained within the various areas defined by their union or intersection. Identifying the number of elements in these areas is known as -----------
State the general procedure for solving enumeration problems:

1.8 Summary

The theory of set is currently used in several business operations. It is used in much software programming, in the study of consumer behaviour, and in solving complex business problems. The unit has put together some preliminary concepts in set notations. You must have been exposed to the simple operations on sets.

A mathematical set is a collection of distinct objects, often referred to as Elements or members

There are two basic operations on sets, including:

1. Set union
2. Set intersection

Set enumeration considers sets in terms of number of elements contained within the various areas defined by their union or intersection. Identifying the number of elements in these areas is known as set enumeration.

State the general procedure for solving enumeration problems:

Step 1: Identify the attribute sets
Step 2: Draw an outline Venn diagram
Step 3: Use the information given to fill in as much of the diagram as possible
Step 4: Evaluate the number of elements in unknown areas.

1.8 Reference

A. Francis (1998). *Business Mathematics and Statistics, 5th edition.* Great Britain: Ashford Colour Press.

Intellspot.com (2022) Set theory. Retrieved from (https://www.intellspot.com/venn-diagram-examples/). Accessed on 14/08/2022.

1.9 Possible Answer to SAEs

SELF-ASSESSMENT EXERCISE

A mathematical set is a collection of distinct objects, often referred to as Elements or members

There are two basic operations on sets, including:

i. Set union

ii. Set intersection

SELF-ASSESSMENT EXERCISE 2

1. Set enumeration considers sets in terms of number of elements contained within the various areas defined by their union or intersection. Identifying the number of elements in these areas is known as set enumeration.

State the general procedure for solving enumeration problems:

Step 1: Identify the attribute sets

Step 2: Draw an outline Venn diagram

Step 3: Use the information given to fill in as much of the diagram as possible

Step 4: Evaluate the number of elements in unknown areas.

UNIT 2: STATISTICAL TOOLS III

Unit Structure

2.1 Introduction

In this unit, we pay a special attention to the concept of probability, probability laws, computation of probabilities, and their applications to business decisions. At the end of this lecture, students will be expected to be able to make effective decisions under uncertainties.

The basic elements of probability theory are the outcomes of the process or phenomenon under study.

2.2 Learning Outcomes

At the end of this unit, you should be able to:

- Define Probability
- Explain the theory of probability
- Define probability
- State the laws of probability
- Calculate probabilities
- Apply probabilities in making decisions involving uncertainties

2.3 what is Probability

1.3.1 Definitions of Probability

Probability is a concept that used as “chance,” “likelihood” “possibility” and “proportion as part of everyday speech and business situation. For example, most of the following, which might be heard in any business situation, are in fact statements of probability.

Each possible type of occurrence is referred to as an event. The collection of all the possible events is called the sample space.

A compound or joint event is an event that has two or more characteristics. For example, the event of a student who is “an economics major and B or above average” is a joint or compound event since the student must be an economics major and have a B or above average.

The event “black ace” is also a compound event since the card must be both black and ace in order to qualify as a black ace.
Probability is a concept that most people understand naturally, since such words as “chance,” “likelihood” “possibility” and “proportion are used as part of everyday speech. For example, most of the following, which might be heard in any business situation, are in fact statements of probability.

a)“There is a 30% chance that this job will not be finished in time”.
b)“There is every likelihood that the business will make a profit next year”.
c)“Nine times out of ten, he arrives late for his appointments”.

In statistical sense, probability simply puts a well-defined structure around the concept of everyday probability, enabling a logical approach to problem solving to be followed.
There are basically two separate ways of calculating probability.

1. Calculation based on theoretical probability. This is the name given to probability that is calculated without an experiment that is, using only information that is known about the physical situation.

2. Calculation based on empirical probability. This is probability calculated using the results of an experiment that has been

performed a number of times. Empirical probability is often referred to as *relative frequency* or *Subjective probability*.

2.3.2 Definition of Theoretical Probability

Let E represent an event of an experiment that has an equally likely outcome set, U, then the theoretical probability event E occurring when the experiment is written as Pr (E) and given by:
Where n (E) = the number of outcomes in event set E n (U) = total possible number of outcomes in outcome set, U.

$$\Pr(E) = \frac{\text{number of different ways that the event can occur}}{\text{number of different possible outcomes}} = \frac{n(E)}{n(U)}$$

If, for example, an ordinary six-sided die is to be rolled, the equally likely outcome set, U, is {1,2,3,4,5,6} and the event "even number" has event set {2,4,6}. It follows that the theoretical probability of obtaining an even number can be calculated as:

$$\Pr(\text{even numbers}) = \frac{n(\text{even numbers})}{n(U)} = \frac{3}{6} = 0.50$$

SELF-ASSESSMENT EXERCISE 1

Define the term Probability State the two ways of calculating probability Explain the theoretical Probability

2.4 Definition of Empirical (Relative Frequency) Probability

If E is some event of an experiment that has been performed a number of times, yielding a frequency distribution of events or outcomes, then the empirical probability of event E occurring when the experiment is performed one more time is given by: Pr(E)□□ number of times the event occurred number of times the experiment was performed f(E)
<u>f(E)</u>
f Where f(E) f = the frequency of event E.
f = total frequency of the experiment.

Put differently, the empirical probability of an event E occurring is simply the proportion of times that event E actually occurred when the experiment was performed.

For example, if, out of 60 orders received so far this financial year, 12 were not completely satisfied, the proportion, 12/60 = 0.2 is the empirical probability that the next order received will not be completely satisfied.

2.5 Laws of Probability

1) Addition Law for mutually exclusive events
2) Addition Law for events that are not mutually exclusive
3) Multiplication Law for Independent events
4) Multiplication Law for Dependent events.

2.5.1 Addition Law for Mutually Exclusive Events

Two events are said to be mutually exclusive events if they cannot occur at the same time. The addition law states that if events A and B are mutually exclusive events, then: Pr (A or B) = Pr (A) + Pr (B)

2.5.2 Multiplication Law for Independent Events

This law states that if A and B are independent events, then:
Pr (A and B) = Pr (A). Pr (B)

As an example, suppose, in any given week, the probability of an assembly line failing is 0.03 and the probability of a raw material shortage is 0.1.
If these two events are independent of each other, then the probability of an assembly line failing and a raw material shortage is given by:
Pr (Assembly line failing and Material shortage) = (0.03)(0.1) = 0.003

2.5.3 Multiplication Law for Dependent Events

This Law states that if A and B are dependent events, then:
Pr (A and B) = Pr(A).Pr(B/A)
Note that Pr (B/A) in interpreted as probability of B given that event A has occurred.

Example

A display of 15 T-shirts in a Sports shop contains three different sizes: small, medium and large. Of the 15 T-shirts: 3 are small 6 are medium 6 are large.

If two T-shirts are randomly selected from the T-shirts, what is the probability of selecting both a small T-shirt and a large T-shirt, the first not being replaced before the second is selected?

Solution

Since the first selected T-shirt is not replaced before the second T-shirt is selected, the two events are said to be dependent events. It follows that:

Pr (Small T-shirt and Large T-shirt)
= Pr(Small).Pr(Large/Small)
= (3/15)(4/14)
= (0.2)(0.429)
= 0.086

2.6 Conditional Probability

Assuming two events, A and B, the probability of event A, given that event B has occurred is referred to as the conditional probability of event A.

In symbolic term:
Pr(*A B*) Pr(*A*)□□ Pr(*B A*)/Pr(*B*) Pr (A□□ B)□
Pr(B)Pr(A)□□ Pr(B)
Pr(B)□ Pr(*A*)
Where Pr (A/B) = conditional probability of event A
Pr (A□B) = joint probability of events A and B
Pr (B) = marginal probability of event B
In general, Pr(A *B*) Joint Probability of events A and B Marginal Probability of event B

2.7 The Bayes Theorem

Bayes theorem is a formula which can be thought of as "reversing" conditional probability. That is, it finds a conditional probability, A/B given, among other things, its inverse, B/A. According to the theorem, given events A and B, Pr(*A B*)□□□Pr(*A*)□□ Pr(*B A*) Pr(*B*)
As an example in the use of Bayes theorem, if the probability of meeting a business contract date is 0.8, the probability of good weather is 0.5 and the probability of meeting the date given good weather is 0.9, we can calculate the probability that there was good weather given that the contract date was met.

2.8 Probability and Expected Values

The expected value of a set of values, with associated probabilities, is the arithmetic mean of the set of values. If some variable, X, has its values specified with associated probabilities, P, then:
Expected value of X = E (X) = PX

Age	Male (M)	Female (F)	Marginal Probability
Below 30 (B)	0.2857	0.3333	0.62
30 and Above (A)	0.2857	0.0952	0.38
Marginal Probability	0.57	0.43	1.00

Example

An ice-cream salesman divides his days into 'Sunny' 'Medium' or 'Cold'. He estimates that the probability of a sunny day is 0.2 and that 30% of his days are cold. He has also calculated that his average revenue on the three types of days is N220, N130, and N40 respectively.
If his average total cost per day is N80, calculate his expected profit per day.

Solution

We first calculate the different values of profit that are possible since we are required to calculate expected profit per day, as well as their respective probabilities.

Given that Pr (sunny day) = 0.2; Pr (cold day) = 0.3. Since in theory, Pr (sunny day) + Pr (cold day) +Pr (medium day) = 1. It follows that:

Pr (medium day) = 1 - 0.2 - 0.3 = 0.5
The total costs are the same for any day (N80), so that the profits that the salesman makes on each day of the three types of day are:

Sunny day:
N (220-80) = N140
Medium day: N (130-80) = N50
Cold day: N (40-80) = -N40 (loss).

SELF-ASSESSMENT EXERCISE 2

State the Laws of Probability
State the Addition Law for Mutually Exclusive Events
Explain the Multiplication Law for Independent Events
Since the first selected T-shirt is not replaced before the second T-shirt is selected, the two events are said to be dependent events. It follows that: Pr (Small T-shirt and Large T-shirt). Calculate.

2.9 Summary

Probability is a concept that most people understand naturally, since such words as "chance," "likelihood," "possibility" and "proportion are used as part of everyday speech. It is a term used in making decisions involving uncertainty. Though the concept is often viewed as very abstract and difficult to relate to real world activities, it remains the best tool for solving uncertainties problems.

To remove some of the abstract nature of probabilities, this unit has provided you with the simplest approach to understanding and calculating, as well as applying the probability concept. It defines probability in two basic forms:

the theoretical definition; and
the empirical definition

The issues discussed in this unit can be summarised in the following way: there are basically two separate ways of calculating probability which are as stated below:

theoretical probability: this is calculated without an experiment,
ii that is, using only information that is known about the physical situation.

Calculation based on empirical probability. This is probability calculated using the results of an experiment that has been performed a number of times. Empirical probability is often referred to as Relative frequency or Subjective probability. There are four basic laws of probability:

1. Addition law for mutually exclusive events
2. Addition law for events that are not mutually exclusive
3. Multiplication law for independent events

4. Multiplication law for dependent events

A joint probability implies the probability of joint events. Joint probabilities can be conveniently analyzed with the aid of joint in which all possible events for a variable are recorded in a row and those of other variables are recorded in a column, with the values listed in corresponding cells.

The Marginal Probability of an event is its simple probability of occurrence, given the sample space.

Assuming two events, A and B, the probability of event A, given that event B has occurred is referred to as the conditional probability of event A.

The expected value of a set of values, with associated probabilities, is the arithmetic mean of the set of values. If some variable, X, has its values specified with associated probabilities, P, then:
Expected value of X = E (X) =□ PX

2.10 References/Further Readings/Web Resources

A. Francis (1998). Business Mathematics and Statistics, 5th edition. Great Britain: Ashford Colour Press.

2.11 Possible Answers to SAEs

Possible Answers to SAEs 1

1. Probability is a concept that used as “chance,” “likelihood” “possibility” and “proportion as part of everyday speech and business situation. For example, most of the following, which might be heard in any business situation, are in fact statements of probability.

Each possible type of occurrence is referred to as an event. The collection of all the possible events is called the sample space.

A compound or joint event is an event that has two or more characteristics. For example, the event of a student who is “an economics major and B or above average” is a joint or compound event since the student must be an economics major and have a B or above average.

The event “black ace” is also a compound event since the card must be both black and ace in order to qualify as a black ace.

a)“There is a 30% chance that this job will not be finished in time”.
b)“There is every likelihood that the business will make a profit next year”.
c)“Nine times out of ten, he arrives late for his appointments”.

In statistical sense, probability simply puts a well-defined structure around the concept of everyday probability, enabling a logical approach to problem solving to be followed.

2. There are basically two separate ways of calculating probability.

1. Calculation based on theoretical probability. This is the name given to probability that is calculated without an experiment that is, using only information that is known about the physical situation.

2. Calculation based on empirical probability. This is probability calculated using the results of an experiment that has been performed a number of times. Empirical probability is often referred to as *relative frequency* or *Subjective probability.*

2. Definition of Theoretical Probability

Let E represent an event of an experiment that has an equally likely outcome set, U, then the theoretical probability event E occurring when the experiment is written as Pr (E) and given by:
Where n (E) = the number of outcomes in event set E n (U) = total possible number of outcomes in outcome set, U.

$$Pr(E) = \frac{\text{number of different ways that the event can occur}}{\text{number of different possible outcomes}} = \frac{n(E)}{n(U)}$$

If, for example, an ordinary six-sided die is to be rolled, the equally likely outcome set, U, is {1,2,3,4,5,6} and the event "even number" has event set {2,4,6}. It follows that the theoretical probability of obtaining an even number can be calculated as:

$$Pr(\text{even numbers}) = \frac{n(\text{even numbers})}{n(U)} = \frac{3}{6} = 0.50$$

Self-Assessment Exercise 2

1 Laws of Probability
1) Addition Law for mutually exclusive events
2) Addition Law for events that are not mutually exclusive
3) Multiplication Law for Independent events
4) Multiplication Law for Dependent events.

2. Addition Law for Mutually Exclusive Events
Two events are said to be mutually exclusive events if they cannot occur at the same time. The addition law states that if events A and B are mutually exclusive events, then: Pr (A or B) = Pr (A) + Pr (B)
3. Multiplication Law for Independent Events

This law states that if A and B are independent events, then:
Pr (A and B) = Pr (A). Pr (B)

As an example, suppose, in any given week, the probability of an assembly line failing is 0.03 and the probability of a raw material shortage is 0.1.

If these two events are independent of each other, then the probability of an assembly line failing and a raw material shortage is given by:

Pr (Assembly line failing and Material shortage) = (0.03)(0.1) = 0.003

4. A display of 15 T-shirts in a Sports shop contains three different sizes: small, medium and large. Of the 15 T-shirts: 3 are small 6 are medium 6 are large.

If two T-shirts are randomly selected from the T-shirts, what is the probability of selecting both a small T-shirt and a large T-shirt, the first not being replaced before the second is selected?

Solution
Since the first selected T-shirt is not replaced before the second T-shirt is selected, the two events are said to be dependent events. It follows that:

Pr (Small T-shirt and Large T-shirt)
= Pr(Small).Pr(Large/Small)
= (3/15)(4/14)
= (0.2)(0.429) = 0.086

UNIT 3 BASIC ADVANCE MATHEMATICS

Unit Structure

3.1 Introduction

Algebra is the branch of mathematics that helps in the representation of problems or situations in the form of mathematical expressions. It involves variables like x, y, z, and mathematical operations like addition, subtraction, multiplication, and division to form a meaningful mathematical expression. All the branches of mathematics such as trigonometry, calculus, coordinate geometry, involve the use of algebra. Algebra deals with symbols and these symbols are related to each other with the help of operators. It is not just a mathematical concept, but a skill that all of us use in our daily life without even realizing it. Understanding algebra as a concept is more important than solving equations and finding the right answer, as it is useful in all the other topics of mathematics that you are going to learn in the future or you have already learned in past.

3.2 Learning Outcomes

At the end of this unit, you should be able to:

- Define the Basic Algebra and the basic rules.
- Calculate the Linear equations
- Calculate the Quadratic formula (Factorization)
- Calculate the Quadratic formula (Formula Method)
- Apply the Quadratic formula to decision making

3.3 Basic Algebra

3.3.1 Linear equations

Linear equations are classified as first-degree equations.

Basic Rules and Properties of Algebra

The basic rules or properties of algebra for variables, algebraic expressions, or real numbers a, b and c are as given below,
Commutative Property of Addition: a + b = b + a
Commutative Property of Multiplication: a × b = b × a
Associative Property of Addition: a + (b + c) = (a + b) + c
Associative Property of Multiplication: a × (b × c) = (a × b) × c
Distributive Property: a × (b + c) = (a × b) + (a × c), or, a × (b - c) = (a × b) - (a × c)

Reciprocal: Reciprocal of a = 1/a
Additive Identity Property: a + 0 = 0 + a = a
Multiplicative Identity Property: a × 1 = 1 × a = a
Additive Inverse: a + (-a) = 0
For video demonstration, follow this link: https://www.khanacademy.org/math/algebra/x2f8bb11595b61c86:foundation-algebra/x2f8bb11595b61c86:intro-variables/v/why-aren-t-we-using-the-multiplication-sign?modal=1

Example 1
One simple example of an expression in algebra is 2x + 4 = 8.
Solve the Unknown variables in a linear equation using simple algebraic operations,.
5X – 6 = 3X
2(p + 4) = 7p+2
Solution

5X – 6 = 3X
5X – 3X = 6

(5-3)X = 6
2X = 6
2X/2 = 6/2 = 2

2(p + 4) = 7p+2
2p + 8 = 7p + 2

2p - 7p = 2-8
(2 - 7)p = -6
-5p/5 = -6/5 6/5 =1/2

Example 2
$$\frac{(7X + 3) - (9X + 8)}{2} = \frac{6}{4}$$
Solution

$$\frac{(7X + 3) - (9X + 8)}{2} = \frac{6}{4}$$

2(7X + 3) – 1(9X + 8) = 4(6)

(14X + 6) – (9X + 8) = 24

Clear the Bracket

14X + 6 – 9X + 8 = 24

14X – 9X = 24 – 6 – 8

$$\frac{5X}{5} = \frac{10}{5} \qquad X = 2$$

SELF-ASSESSMENT EXERCISE 1

Calculate x - 5 = 2. Five less than a number equals to two. What is the number?

Expand (2x + 3y)2 using the algebraic identities.

The present age of a person is double the age of his son. Ten years ago, his age was four times the age of his son. Use the concept of algebra and find the present age of the son.

3.4 Quadratic formula (Factorization)

Factorization involves the determination of the factors that form the given quadratic equation. This will then make the solutions for the unknown easy to come by.

Example

Explain the Factorization methods of quadratic equation

Solve for x in the quadratic equation: X2 + X – 12 = 0

Solution

X2 + X – 12 = 0

(X – 3) and (X + 4) are the factors, so that

(X – 3)(X+ 4) = 0

3.5 Quadratic formula (Formula Method)

The quadratic formula helps us solve any quadratic equation. First, we bring the equation to the form $ax^2+bx+c=0$, where a, b, and c are coefficients. Then, we plug these coefficients in the formula: $(-b\pm\sqrt{(b^2-4ac)})/(2a)$. See examples of using the formula to solve a variety of equations.

Example

Solve the equation using formula method

4x2 -17x+15 = 0

Solution:

X = -b+_ b2-4ac
2a

Using the formula, a = 4; b = -17; c = 15. By substitution to the formula, we get:
X = -(-17)+_ (-17)2-4(4)(15)
2(4)
17+ 289-240/8
17+ 49/8
X = 17+7 or X = 17-7
X = 3 or X = 1.25

3.6 Application of Quadratic formula

Example
XYZ Company produces product A for which cost (including labour and material) is N6/unit. Fixed cost is N80, 000. Each unit is sold for N10.

Determine the number of units which must be sold for the company to earn a profit of N60, 000.
Solution:

By definition, Profit = Revenue-Cost
That is = R – C

Let x represent the level of output, so that:
Total Cost = C = Fixed Cost (FC) + Variable Cost (VC),

That is, TC = FC + VC
Therefore: FC = N80,000, VC = 6X, P=10. Expected Profit = 60,000

Where: The variable cost (VC) per unit produced is N6, so that for x units,
Determine the C
Determine the R
C= FC + VC; R= Price x Quantity P(X); VC = 6X
C = FC + VC = 80,000 + 6X
Revenue (R) = (unit price) (quantity sold)
Thus, R = PX
From the problem, P = 10, so that,
R = PX = 10X
The expected profit ( ) is:
= N60, 000
To find the value of X, R= PX – (FC + VC *X)

$60,000 = 10x - (80,000 + 6X)$
Solving for X, we get:
$10X - (80,000 + 6X) = 60,000$
$10X - 80,000 - 6X = 60,000$
$10X - 6X = 60,000 + 80,000$
$(10 - 6)X = 140,000$
$4X = 140,000$
$4X/4 = 140,000/4$
$X = 35,000$
Therefore, 35,000 units must be sold to earn a profit of N60, 000.

SELF-ASSESSMENT EXERCISE 2

Explain the Factorization methods of quadratic equation
Solve for x in the quadratic equation: $X2 + X - 12 = 0$
Solve the equation using formula method. $4x2 -17x+15 = 0$

https://www.cuemath.com/algebra/

1.7 Summary

This unit discussed the basic Define the Basic Algebra and the basic rules. The unit displayed Linear equations, Quadratic formula (Factorization), Calculate the Quadratic formula (Formula Method) and Application of Quadratic formula to decision making.

1.8 Possible Answers to SAEs

Answers to SAEs 1

Calculate $x - 5 = 2$. Five less than a number equals to two. What is the number?

Solution:
Using the concept of Algebra, we will assume the number to be a variable. Let the number be x. As per the question, we can write $x - 5 = 2$. On solving this, we get $x = 7$. Therefore, the required number is 7.

Expand (2x + 3y)2 using the algebraic identities.

Solution:
Here we shall use an identity in algebra, $(a + b)^2 = a^2 + 2ab + b^2$
(2x + 3y)2 = (2x)2 + 2(2x)(3y) + (3y)2
$= 4x^2 + 12xy + 9y^2$
Therefore, the answer is $(2x + 3y)^2 = 4x^2 + 12xy + 9y^2$

The present age of a person is double the age of his son. Ten years ago, his age was four times the age of his son. Use the concept of algebra and find the present age of the son.

Solution:
Let us consider the present age of the son as 'x' years. It is given that the age of the person is double the age of his son, so the age of the person is '2x' years. Now considering the situation 10 years ago, the age of the son was (x - 10) years and the age of the person was (2x - 10) years. The question says that 10 years ago the age of the person was 4 times the age of his son. Therefore, this can be

expressed as,
$2x - 10 = 4(x - 10)$
$2x - 10 = 4x - 40$
$2x - 4x = -40 + 10$
$-2x = -30$
$2x = 30$
$x = 30/2$
$x = 15$

Therefore, the present age of the son is 15 years.
Answers to SAEs 2

Explain the Factorization methods of quadratic equation
This involves the determination of the factors that form the given quadratic equation. This will then make the solutions for the unknown easy to come by.

Solve for x in the quadratic equation: X2 + X – 12 = 0

Solution
X2 + X – 12 = 0
(X – 3) and (X + 4) are the factors, so that
(X - 3)(X+ 4) = 0
Solve the equation using formula method
4x2 -17x+15 = 0
Solution:

X = -b+_ b2-4ac
2a

Using the formula, a = 4; b = -17; c = 15. By substitution to the formula, we get:

X = -(-17)+_ (-17)2-4(4)(15)
2(4)

17+ 289-240/8
17+ 49/8
X = 17+7 or X = 17-7
X = 3 or X = 1.25